MACROMOLECULAR INTERACTOME OF *TETRAHYMENA* CHD FAMILY CHROMATIN REMODELERS

Eva Francesca DeRango-Adem

A thesis submitted to the Faculty of Graduate Studies in partial fulfillment of the requirements for the Master of Science

Graduate Program in Biology

York University

Toronto, Ontario

October 2017

Abstract

Tetrahymena thermophila undergoes RNA-guided, DNA elimination during development. Bioinformatic and functional analyses on tChd family chromatin remodelers suggest the proteins are candidate DNA and RNA binding proteins. *In vivo* and *in vitro* functional analyses on tChd3 and tChd7 identified diverse macromolecular interactions throughout the lifecycle. Chd3 bound a novel protein named Miz1, which is a candidate E3 SUMO ligase. Chd3 and Miz1 co-localize to vegetatively repressed *PDD1*, and may regulate transcription via epigenetic SUMOylation of chromatin.

Chd7 binds (un)modified H3K36, and highly transcribed genes during vegetative growth. This binding may regulate transcriptional turnover by Chd7-mediated reading of the acetyl/methyl status at H3K36. *In vitro* analyses suggest Chd7 binds poly(G) RNAs. G-rich sequences are transcribed during RNA-guided DNA elimination events, and Chd7 may be involved during this part of *Tetrahymena* development. These analyses provide a glimpse into how RNAs and epigenetics orchestrate chromatin dynamics, in complex eukaryotic regulatory mechanisms.

Acknowledgments

First, I want to express my gratitude to Prof. Ronald Pearlman. It has been a privilege to work in his lab for the past four years – first as a Research Practicum student, then as an Undergraduate Honours Thesis student, and finally as a Master of Science student. His expertise, guidance, understanding, and kindness were essential to my success, and the completion of this thesis. I am also glad to have found a fellow chocolate cake enthusiast, who was always willing to try my baked good s.

Second, I must thank Prof. Mark Bayfield. His direct involvement in my project, and his willingness to help, were instrumental during the latter stages of my thesis.

Third, I would like to thank Dr. Jyoti Garg. Her unrelenting assistance – and most importantly her friendship - were invaluable throughout my time at the Pearlman Lab. I have gotten to know her and her beautiful family. I am so thankful to have made a new, lifelong friend. I look forward to sharing even more tasty recipes in the years to come!

Fourth, I would also like to thank Prof. Jeffrey Fillingham, Dr. Jean-Phillipe Lambert, Dr. Debashish Ray, Alejandro Saettone, Anita Samardzic, Roberto Toste, Stefano Marrella, Michelle Prioriello, and Ritu Sharma. These people have provided advice, collaboration, encouragement, and laughter.

Finally, I would like to thank my parents Rosanna DeRango and Teferi Adem – who never questioned why I needed to be at school until 8:00 p.m. on Friday nights to collect my mating cells.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

Ac	Acetylated
Amp	Ampicillin
AP-MS	Affinity Purification Mass Spectrometry
AP-WB	Affinity Purification Western Blot
APS	Ammonium Persulfate
ATP	Adenosine Triphosphate
BME	Beta-mercaptoethanol
BLAST	Basic Local Alignment Search Tool
BLOTTO	Bovine-Lacto Transfer Technique Optimizer
bp	Base Pairs
CBC	Chromodomain-Bromodomain-Chromodomain
CBS	Chromosome Breakage Sequences
Chd	Chromodomain Helicase DNA-Binding Protein
ChIP	Chromatin Immunoprecipitation
Da	Dalton
DAPI	4',6-Diamidino-2-Phenylindole
Dcl1	Dicer-like Protein 1
ddH_2O	Double Distilled Water
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
DRBP	DNA and RNA-binding Protein
dsRNA	Double-Stranded RNA
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
EtBr	Ethidum Bromide
FITC	Fluorescein isothiocyanate
FZZ	3x FLAG-ZZ epitope tag
g	Grams
h	Hours

H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H3K4	Histone H3 lysine 4
H3K9	Histone H3 lysine 9
H3K27	Histone H3 lysine 27
H3K36	Histone H3 lysine 36
H4	Histone 4
HTP	High Throughput
HRP	Horseradish Peroxidase
IES	Internally Eliminated Sequence
IF	Indirect Immunofluorescence
IgG	Immunoglobulin G
IP	Immunoprecipitation
k	Kilo
Kan	Kanamycin
Kb	Kilobase Pair
kDa	Kilodalton
KO	Knock-Out
L	Litres
LC	Liquid Chromatography
Lnc	Long Non-Coding
LTQ	Linear Trap Quadrupole
m	Milli
M	Molar
MAC	Macronucleus
Me	Methylated (Me3 trimethyl, Me2 dimethy, Me1 monomethyl)
MIC	Micronucleus
MS	Mass Spectrometry
N	Nano
NGS	Next Generation Sequencing

NP-40	Nonidet P-40
OD	Optical Density
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
Pdd	Programmed DNA Deletion Protein
pH	Potential Hydrogen
PIP	Protein Immunoprecipitation
PMSF	Phenylmethanesulfonylfluoride
PNK	Polynucleotide Kinase
PPD	Paz/Piwi Domain Protein
PPP	PHD-PHD-PHD
PTM	Post-Translational Modification
qPCR	Quantitative PCR
rDNA	Ribosomal Deoxyribonucleic Acid
Rho	Rhodamine
RBP	RNA-binding protein
RIP	RNA Immunoprecipitation
RISC	RNA-Induced Silencing Complex
RNA	Ribonucleic Acid
RNAi	RNA Inference
RNAPII	RNA Polymerase II
rpm	Rotations Per Minute
SAINT	Significance Analysis of INTeractome
scRNA	Scan RNA
SDS	Sodium Dodecyl Sulphate
siRNA	Small Interfering RNA
SPP	Sucrose Proteose Peptone
SUMO	Small Ubiquitin-like Modifier
TCA	Trichloroacetic Acid

TEB	Tris-EDTA-Borate
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TF	Transcription Factor
TGD	*Tetrahymena* Genome Database
TRIS	Trisma Base (Trishydroxymethylamino methane)
TWEEN 20	Polyoxyethelenesorbitan Monolaurate 20
µ	Micro
UTR	Untranslated Region
UV	Ultraviolet
V	Volts
Veg	Vegetative Growth
WB	Western Blot
WCE	Whole Cell Extract
WT	Wild Type
YPD	Yeast Extract Peptone Dextrose
YT	Yeast Extract Tryptone

1.0 Introduction

1.10 Chromatin: A dynamic macromolecular complex

1.110 DNA and histones: a nucleoprotein complex

Eukaryotic cells can package six feet of genomic DNA within a six micron nucleus. This is made possible through orders of magnitude of compaction. 146bp of DNA wrap around an octamer of core histone proteins in a nucleoprotein complex called the nucleosome (**Figure 1**). There are four core histone proteins, and the histone octamer is comprised of one H3-H4 tetramer, and two H2A-H2B dimers (Luger et al., 1997). A structural histone fold motif enables histone:histone interactions, and the globular domains of histones enable DNA interactions (Arents & Moudrianakis, 1995). Nucleosomes are connected and form regularly spaced thread-like structures analogous to "beads on a string", via free DNA and the histone linker protein H1 (Robinson & Rhodes, 2006). Multiple nucleosomes can interact, forming a 30nm chromatin fibre. Ultimately these chromatin fibres loop, and can compact further into chromosomes during chromosomal segregation events such as the mitotic and meiotic cell cycles (Fischle, Wang, & Allis, 2003). A new study coupling electron microscopy to DNA-binding dyes has enabled live-cell visualization of chromatin dynamics within the nucleus (Ou et al., 2017). The classical, ridged magnitudes of compaction described above - the current paradigm for chromatin dynamics - were challenged. Instead, the researchers concluded that chromatin is a highly disordered complex, and composed of flexible chains that can bend at different lengths (Ou et al., 2017). Further, the diameter of chromatin ranges between 5- to 24-nm (Ou et al., 2017).

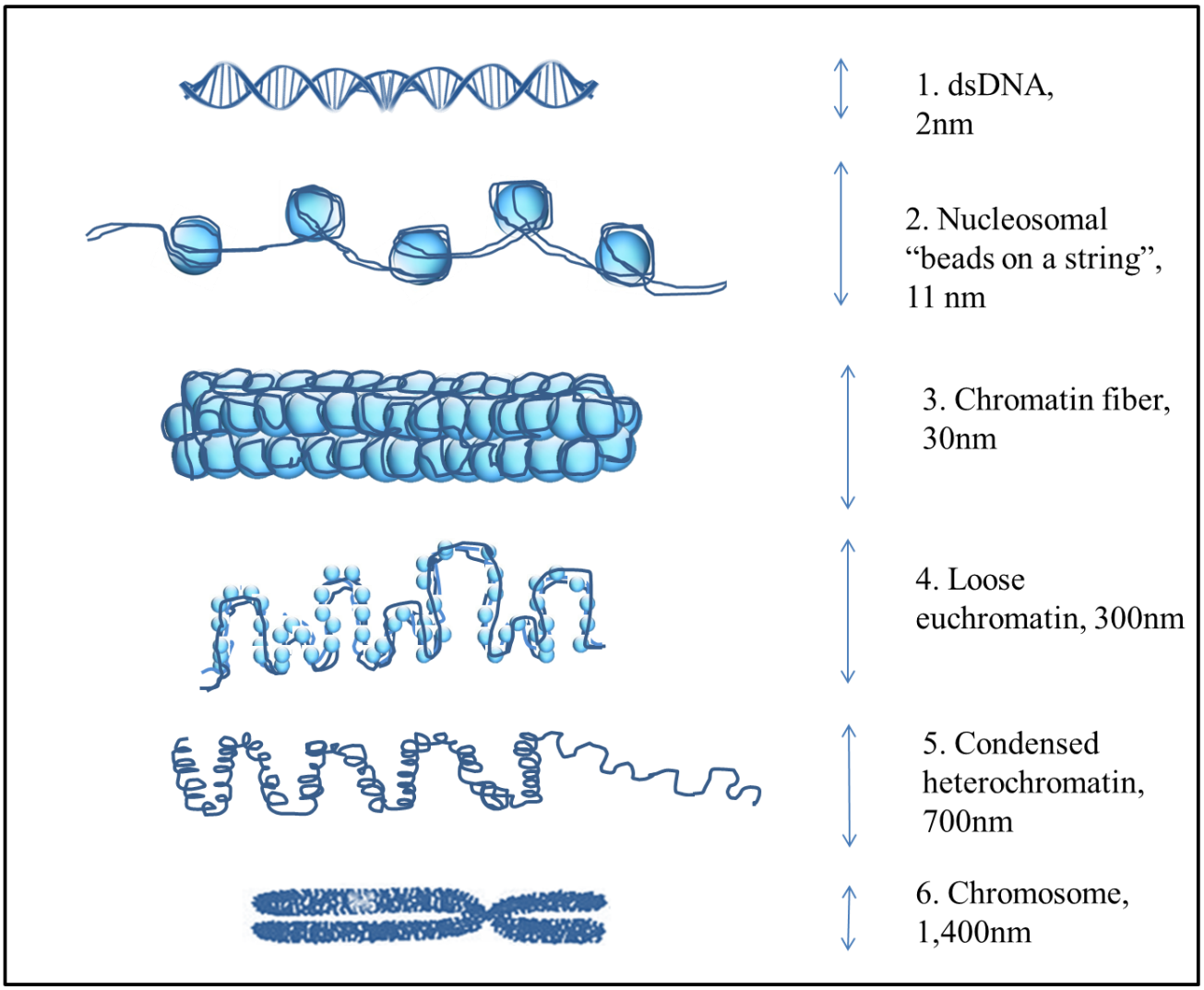


Figure 1. *Magnitudes of DNA compaction.* In eukaryotes, DNA undergoes magnitudes of compaction. First, the dsDNA is wrapped around a histone octamer in a complex known as the nucleosome. Multiple nucleosomes are linked by short stretches of free genomic DNA, which emulate "beads on a string". This string of chromatin can then coil into itself (in either a solenoid or zig-zag topology), generating a 30nm chromatin fiber. Further compaction occurs when the fiber loops. This looped chromatin is somewhat accessible to the transcriptional machinery, and is known as euchromatin (discussed in depth in a later section). If these loops are coiled further, they form a condensed and transcriptionally silent from of chromatin known as heterochromatin. Finally, during interphase of mitotic and meiotic events, the chromatin compacts further into highly condensed chromatids. Image is adapted from Sha and Boyer (2008).

1.120 Chromatin remodeling

Chromatin is a highly dynamic complex that can be characterized as either "loose", transcriptionally active euchromatin, or tightly compacted and transcriptionally silent heterochromatin. Chromatin structure can be altered by a diverse class of proteins known collectively as chromatin remodelers. The alteration of chromatin structure and DNA compaction are essential for transcriptional control, by regulating the availability of DNA to cellular machinery (e.g. transcription factors, polymerases) during transcription, replication and DNA damage repair (Aalfs & Kingston, 2000). Chromatin remodeling occurs in three ways (**Figure 2**). First, some chromatin remodelers have the ability to replace core histone proteins with histone variants. Second, histone proteins can be post-translationally, or epigenetically, modified by chromatin remodelers by the addition or removal of chemical subgroups (Luger et al., 1997). Lastly, histones can undergo physical manipulation of nucleosome structure via altering nucleosomal positioning. The latter two forms of chromatin remodeling – epigenetic modification of histones and physical manipulation of nucleosomes – will be emphasized.

1.20 Chromatin remodeling by the epigenetic modification of histones

1.210 The epigenetic landscape: The Histone Code

Proteins have the capacity to be chemically and post-translationally modified. On histones, these modifications can alter gene expression without altering genomic information, and are known as epigenetic. Histones have flexible N- and C- termini tails which protrude from the nucleosomal core (Luger et al., 1997). The N-terminal tail of core histone proteins are especially amenable to chemical modifications, and can be the substrate for post translational modifications (PTMs) by

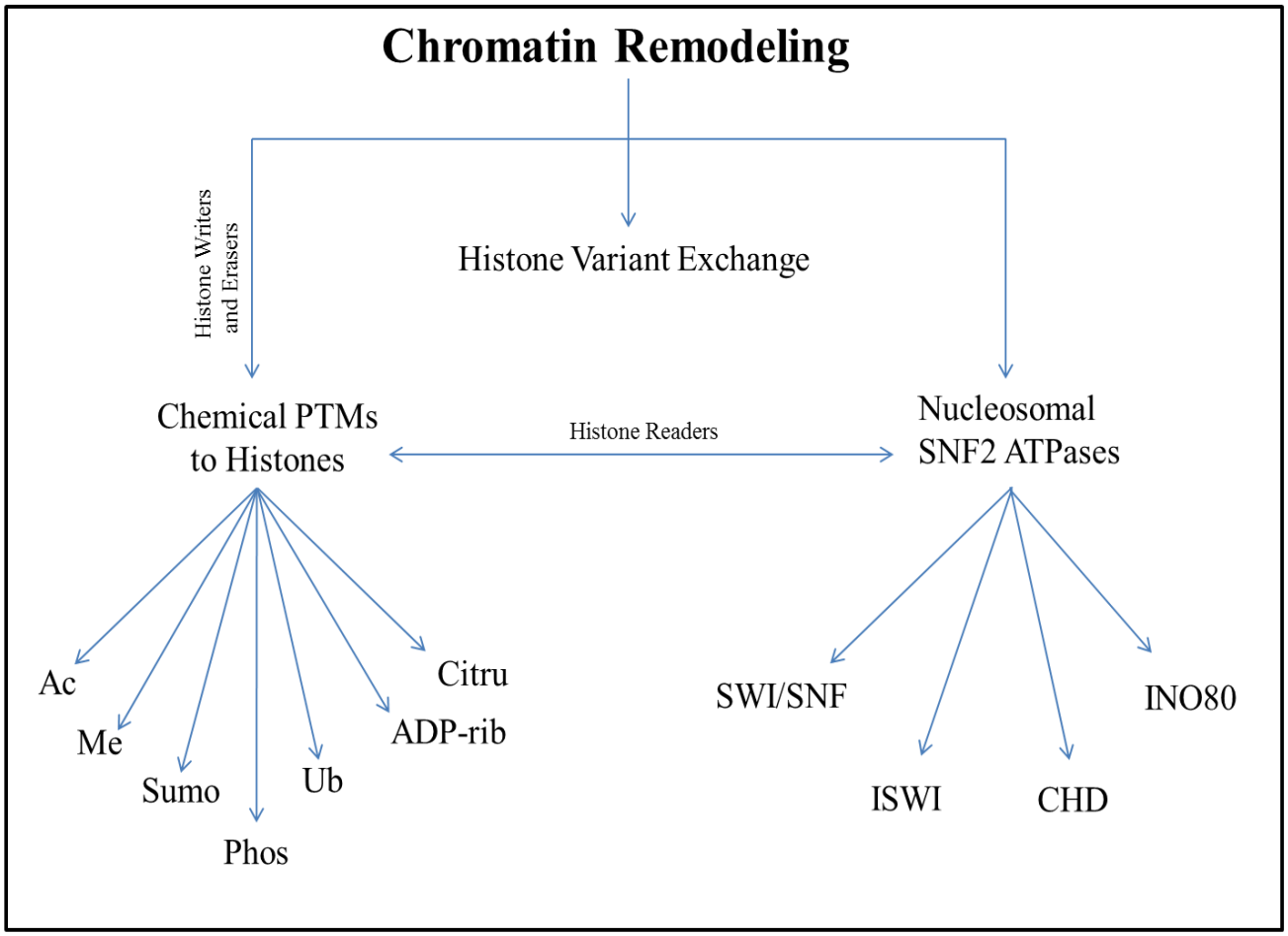


Figure 2. *Forms of chromatin remodeling.* Chromatin remodeling can occur in one of three ways. First, the addition or removal of post-translational modifications (PTMs) to histone proteins, in a process known as epigenetic modification, can mediate transcriptional regulation. Common histone PTMs include acetylation (Ac), methylation (Me), sumoylation (Sumo), phosphorylation (Phos), ubiquitiylation (Ub), ADP-ribosylation (ADP-rib), and citrullation (Citru). Second, the addition of histone variants is a form of chromatin remodeling. Certain enzymes (eg. SWR family of proteins) can exchange core histone proteins with histone variants. Finally, the physical manipulation of nucleosomes by Snf2 ATPases is a form of chromatin remodeling mediated by Swi/Snf2, ISWI, Chd or Ino80 enzymes, which have the capacity to add, slide or eject nucleosomes as a form of transcriptional regulation.

ATP-dependent chromatin remodelers (Luger et al., 1997). Histone epigenetic modifications include acetylation, methylation, SUMOylation, ubiquitylation, phosphorylation, biotinylation, citrullination, and ADP-ribosylation (**Figure 3**). There are hundreds of histone-related enzymes which add, remove, and interpret epigenetic modifications; these proteins are called histone writers, erasers and readers, respectively. The epigenetic landscape of histones - collectively known as the "Histone Code" - includes both individual and multiple synergistic epigenetic marks, which affect cellular function via altered pre- and post-transcriptional regulation or splicing events (Strahl & Allis, 2000). The acetylation and methylation of histone residues by chromatin-associated enzymes are well characterized, and will be discussed below.

1.220 Acetylation of histone lysine residues

All core histone proteins can be acetylated. Acetylation of lysine residues on histones H3 and H4 are exclusively marks of active transcription. This histone modification directly changes DNA:histone dynamics. Specifically, acetylation of lysine alters gene expression via changes to electrostatic interactions. The addition of negatively charged acetyl groups neutralizes the positive charge on lysine-rich histone proteins (Struhl, 1998). This effect in turn loosens the histone's interaction with negatively-charged DNA, and promotes gene transcription by polymerases (Struhl, 1998). Histone H3 acetylation occurs on K9, K14, K18, K23, K36, and K56, while histone H4 can be acetylated on K5, K8, K12 and K16 (Shahbazian & Grunstein, 2007). The acetylation of lysine residues is catalyzed by **h**istone **a**cetyl**t**ransferases, or HATs. These enzymes have the capacity to add acetyl groups ($COCH_3$) from acetyl coenzyme A to a histone target. More than 20 mammalian HATs have been identified which can be classified into five families: GNAT1 (i.e. Gcn5-related N-acetyltransferases), MYST, TAFII250, metazoan-specific p300/CBP (E1A-associated protein of 300kDa/CREB-binding protein), and nuclear

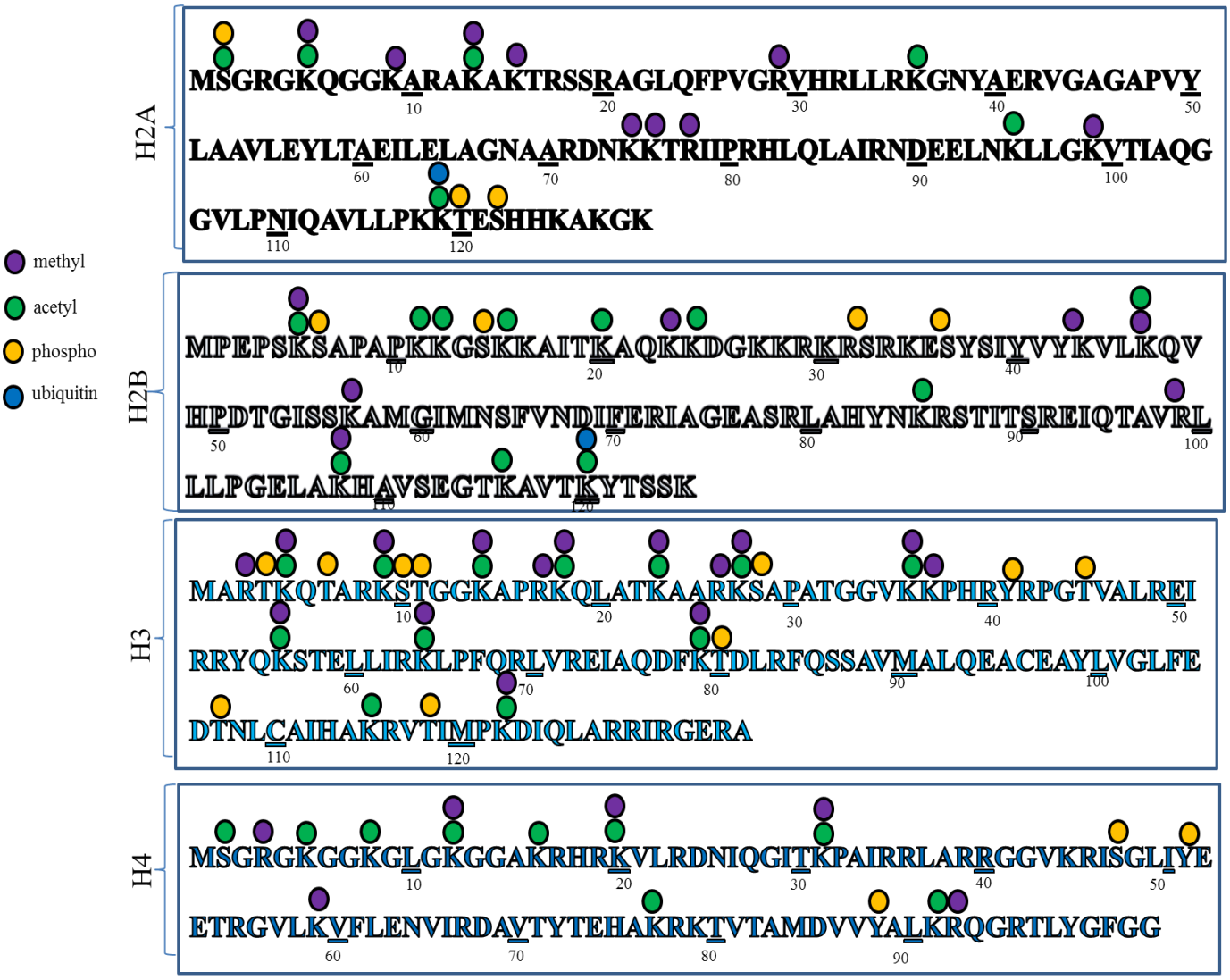


Figure 3. *The epigenetic landscape of histones H2A, H2B, H3 and H4.* Chemical modifications can occur on the N-terminal tails of histone proteins. Depending on the epigenetic mark, the chemical moiety is added to specific histone amino acid residues, by specific histone writers (e.g. acetylation occurs on lysine residues catalyzed by histone acetyltransferases). Both individual ,and multiple synergistic effects of epigenetic marks have the ability to regulate diverse chromatin-associated processes. Epigenetic marks for acetylation (green), methylation (purple), phosphorylation (yellow), and ubiquitination (blue). Proteins have the capacity to be chemically post-translationally modified. This image is adapted from www.ActiveMotif.com.

receptor coactivators (e.g. ACTR) (Marmorstein & Roth, 2001). In *Saccharomyces cerevisiae*, HATs can be generally placed into two classes based on intracellular location patterns. The HAT-A class are nuclear, and catalyze acetylation of core histones during transcriptional processes (Brownell & Allis, 1996). HAT-B enzymes are cytoplasmic, and acetylate H4 or H3 before chromatin assembly, such that pre-acetylated histones are shuttled to the nucleus for deposition (Brownell & Allis, 1996).

The deacetylation of histone lysine residues occurs by **h**istone **d**e**ac**etyltransferases, or HDACs. These enzymes hydrolytically remove the acetyl groups that are added by HATs. There are four main classes of HDACs (reviewed by Haberland, Montgomery, & Olson, 2009). The first class of HDACs includes HDAC1, HDAC2, HDAC3, and HDAC8 – homologous to *S. cerevisiae* yRpd3 (de Ruijter et al., 2003). The second class of HDACs includes HDAC4, HDAC5, HDAC6, HDAC7, HDAC9, and HDAC10 – homologous to yHda1 (de Ruijter et al., 2003; Tauton, Hassig, & Schrediber, 1996). Class III HDACs are also known as the Sirtuins, and these enzymes utilize NAD+ as a cofactor (de Ruijter et al., 2003). SIRT1-7 are homologous to yeast ySir2 (de Ruijter et al., 2003).

1.230 Methylation of histone lysine and arginine residues

Methylation of histones occurs on both lysine and arginine residues. The enzymes responsible for their modifications include **h**istone **m**ethyl**t**ransferases (HMTs) and **h**istone **d**e**m**e**t**hylases (HDMTs). Common histone H3 methylation sites include R2, K4, K9, K27, K36, and K79 (Y. Zhang & Reinberg, 2001). On histone H4, methylation frequently occurs at R3 and K20 (Y. Zhang & Reinberg, 2001). Unlike acetylation, histone residues can be mono-methylated (me1), di-methylated (me2), and tri-methylated (me3), when HMT enzymes transfer these methyl groups from S-adenosyl-L-methionine to the target residue.

Interestingly, there are numerous HMTs which catalyze the addition of methyl groups to the same histone residue. One such example is the methylated state of H3K4. Ash1, ALL-1, ALR, MLL, SET1, SET7/9, SMYD3 and Trx all have the capacity to methylate histone H3 at lysine 4 in mammals (Gu et al., 2013). Similarly, the lysine methyltransferases Dim-5, ESET, EZH2, Eu-HMTase, G9a, SUV39-h1, SUV39-h2, and SETDB1 all have the capacity to methylate histone H3 at lysine 9 (Wu et al., 2010). Both G9a and Polycomb Repressive Proteins (e.g. EZH2) can also methylate histone H3 lysine 27 (H3K27) (Yoo & Hennighausen, 2011). Unlike acetylation, which promotes transcription via electrostatic alterations to DNA:histone interactions, methylation can act as docking sites for other histone modifying enzymes. In turn, the addition of these methyl groups to residues can recruit other proteins to either activate or repress gene transcription. For example, the methylation on specific histone lysine groups – such as H3K9 or H3K27 – are associated with the heterchromatinization, and subsequent repression of DNA transcription. Alternatively, the methylation of H3K4, H3K36, and H3K79 are all associated with active transcription.

The histone code can be read by a variety of chromatin-associated proteins, many of which contain specific protein domains that enable epigenetic interactions. For example, acetylated histones can be read by bromodomain-containing proteins, such as human hBrg1 or yeast ySth1 (Zeng & Zhou, 2002). Methylated lysine residues can be read by chromomdomain-containing proteins, such as *Drosophila* HP1 or *Tetrahymena* Pdd1 (Strahl & Allis, 2000). The structural and functional mechanisms of common histone-reader domains are discussed below.

1.30 Protein domains which function as epigenetic readers

1.310 Chromodomains

Chromodomains are eukaryotically conserved protein motifs of ~50 amino acids in length, and they are primarily found in chromatin-associated proteins. Canonical chromodomains are monomeric, and are composed of a three-stranded anti-parallel beta sheet, and a C-terminal alpha helix (**Figure 4A**) (Ball et al., 1997; Horita et al., 2001). Chromodomain- containing proteins can be divided into three groups: i) single chromodomain proteins (e.g. Polycomb proteins), ii) N-terminal chromodomain proteins with C-terminal chromoshadowdomains (e.g. HP1 proteins), and iii) tandem chromodomain proteins (e.g. Chd family proteins) (Tajul-Arifin et al., 2003).

The chromodomain was first identified in *Drosophila melanogaster,* as a novel motif within Polycomb, a chromosomal protein associated with repression of developmental homeobox genes (Paro & Hogness, 1991). The amino acid sequence of Polycomb was similar to the *Drosophila* protein, HP1(heterochromatin protein 1) (James & Elgin, 1986). Since HP1 proteins were also known to bind heterochromatin and promote gene silencing, "CHROMOdomain" proteins were characterized as **chr**omatin **o**rganization **mo**difier domains (Paro & Hogness, 1991). In the ciliate *Tetrahymena thermophila,* chromodomain proteins Pdd1 and Pdd3 co-localize to nuclear regions associated with developmentally-regulated heterochromatin formation, and ultimately DNA elimination (Madireddi et al., 1996; Nikiforov, Gorovsky, & Allis, 2000). Pdd1 is essential for these processes to occur (Coyne et al., 1999). This developmentally-regulated process of irreversible genome silencing in *Tetrahymena* will be discussed in depth.

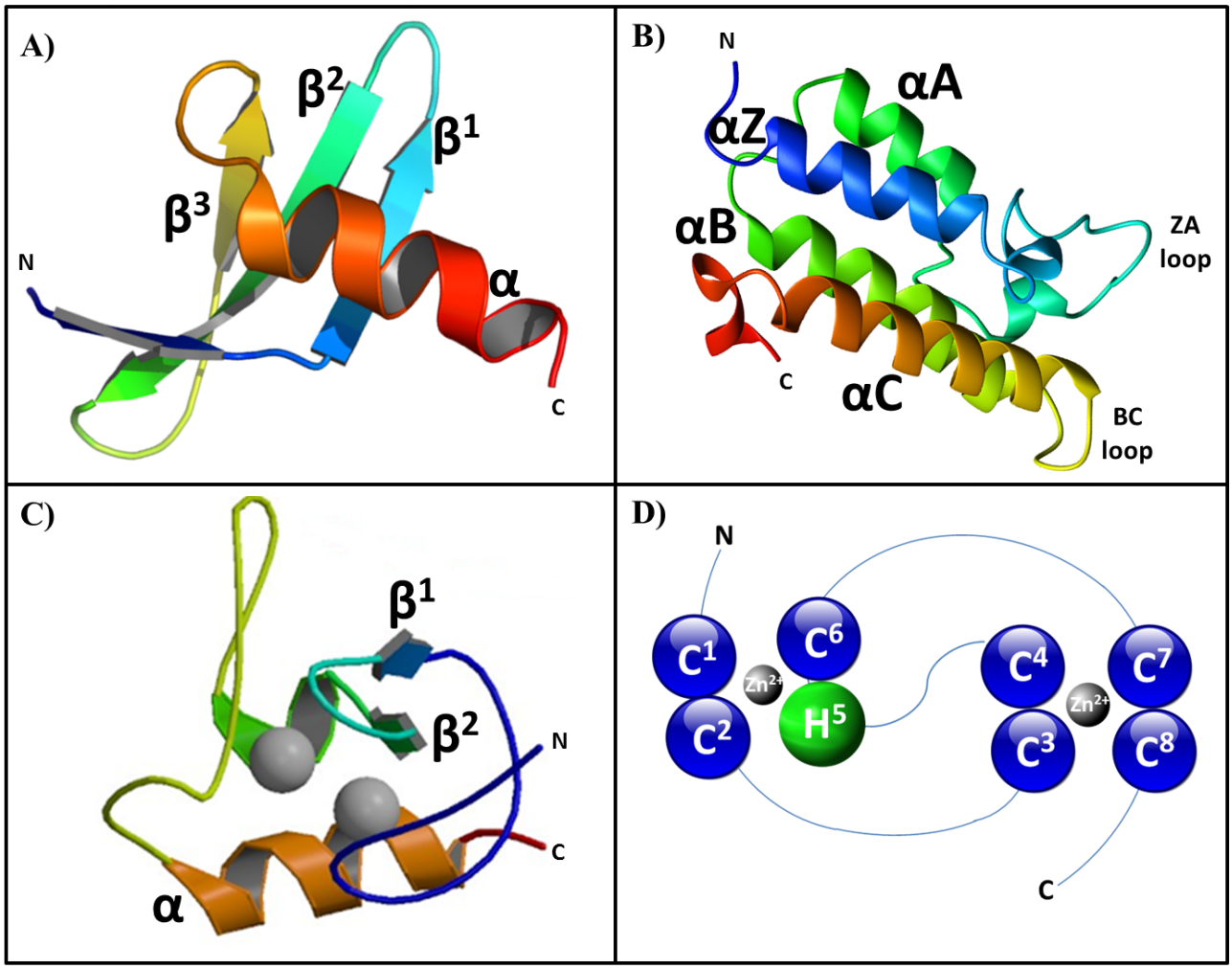


Figure 4. *Structure of epigenetic-binding domains from the RCSB Protein Data Bank*
A) The mouse HP1 protein exhibits classical chromodomain structure, with the presence of a three-stranded anti-parallel beta sheet, and a C-terminal alpha helix **B)** Structural analyses of Human Bromodomain-containing protein 4 (Brd4) – reveals the entirely helical bromodomain is comprised of four alpha helices connected by two loop regions. **C)** The PHD finger in human MLL5 reveals a small globular proteins with a two strand anti-parallel beta-sheet, and a C-terminal alpha helix. Zinc ions are illustrated as grey spheres **D)** Zinc finger proteins coordinate two zinc ions in a cross-brace topology. PHD fingers bind zinc in a C3HC4 signature. Cross-brace image adapted from Bienz 2006.

Functionally, chromodomains have diverse binding substrates, but are bona fide histone H3 methyl-lysine readers. HP1-family proteins are associated with binding repressive epigenetic marks, such as H3K9 and H3K27, via their chromodomain (Lachner et al., 2001). Missense mutations in chromodomains not only inhibit histone binding abilities, but prevent heterochromatin-associated silencing by HP1 proteins (Jacobs et al., 2001). The *Drosophila* HP1 chromodomain:histone interaction was characterized by X-ray crystallography. The H3K9 histone tail forms a beta-sheet-like structure when it inserts itself into the chromodomain, forming a beta sandwich (Jacobs & Khorasanizadeh, 2002). The epigenetically modified H3K9me methyl group inserts itself into a conserved chromodomain "aromatic cage" – comprised of tyrosine 24, tyrosine 48, and tryptophan 45 – in an induced-fit binding topology (Nielsen et al., 2002).

Soon after their discovery, the relevance of chromodomain proteins as RNA binding-modules emerged. In *Drosophila*, the histone acetyltransferse MOF forms a ribonucleoprotein complex, which enables dosage compensation of the male X chromosome (Akhtar et al., 2000). This complex, named Male Specific Lethal (MSL), targets the male X chromosome when bound to long non-coding RNAs (lncRNAs). MOF binds the lncRNA roX2 via it's chromodomain (Akhtar et al., 2000). This function is conserved among mammals, where chromodomain-containing proteins also have the capacity to bind RNAs. The human chromodomain protein CBX7 can bind an antisense long non-coding RNA called *ANRIL*, in a mechanism that likely promotes gene silencing (Yap et al., 2011).

1.320 Bromodomains

Bromodomains are named after the *Drosophila* gene *brahma*, which was implicated in transcriptional activation (Tamkun et al., 1992). Proteins containing these ~110 residue acetyl-

lysine-reader modules are present within histone acetyltransferase proteins, as well as multicomponent chromatin remodeling complexes, transcriptional co-activators, and even methyltransferases (Dhalluin et al., 1999; Mujtaba et al., 2004; Rogaia et al., 1997; Sanchez & Zhou, 2009). Bromodomain-containing HATs play a role in acetylating specific loci, altering chromatin structure, and recruiting transcription factors (reviewed by Josling et al., 2012). Bromodomain-containing proteins without HAT activity are also found in lysine-acetyltransferase complexes, as the bromodomain module is able to "read" and bind acetylated histones, and even promote acetyl-spreading to proximal nucleosomes (Schlissel, 2004).

NMR structural analysis of the bromodomain protein and transcriptional co-activator PCAF (p300/CBP-associated factor), revealed that bromodomains are comprised of four alpha helices, named Z, A, B and C (**Figure 4B**) (Dhalluin et al., 1999). There are two loops in between the helices (named the ZA and BC loops), which are packed against each other forming a hydrophobic pocket that allows acetylated lysine residues of histone proteins to bind (Dhalluin et al., 1999). This loop packing and acetyl:bromodomain interaction is conserved in diverse bromodomain-containing proteins. The histone acetyl-transferase GCN5 exemplifies this conservancy, as structural analyses in human and yeast GCN5 homologs show localization of the acetylated lysine residue within the hydrophobic cavity (Owen et al., 2000). yGcn5 is an essential subunit in the histone modifying complex SAGA (**S**pt-**A**da-**G**cn5 **a**cetyltransferase), which localizes to enhancers and regulates RNA polymerase II-directed transcription (Grant et al., 1997). Gcn5 is also part of the HAT complexes ADA and SALSA/SLIK, which bind distinct histone H3 acetylated lysine residues (Eberharter et al., 1999; Pray-grant et al., 2002).

It is essential to maintain a balance of both histone acetylation and deacetylation. Bromodomain proteins are of therapeutic interest, because histone acetylation marks active transcription (via

euchromatinization), is involved in cellular metabolism, and promotes the DNA repair response (reviewed by Muller, Filippakopoulos, & Knapp, 2011). Promiscuous acetyltransferase or lysine-reader activities in bromodomain-containing proteins have been implicated in the progression of cancer. Mutations in bromodomain proteins may lead to aberrant acetylation patterns, and ultimately altered gene transcription profiles (Ropero & Esteller, 2007). For example, the human bromodomain protein BRD4 binds to the chromatin remodeling complex named Mediator. The Mediator complex is known to bind enhancer regions, and can aberrantly bind the enhancer of the MYC oncogene, promoting the development of cancer (Lovén et al., 2014). Drugs that regulate the function of bromodomain-containing proteins such as HATs, HDACs and acetyl-readers are good targets for cancer therapies. Consequently, there is increased interest to design "Bromodomain-Inhibitors", which can irreversibly bind bromodomain-proteins of interest, and prevent their potentially cancer-promoting activities (Lovén et al., 2014).

1.330 PHD fingers

The Plant Homeodomain Finger (PHD) was first discovered in *Arabidopsis thaliana*, as an ~65 amino acid module which includes a regularly spaced and conserved stretch of Cys4-His-Cys3 (or C4HC3) (Schindler, Beckmann, & Cashmore, 1993). PHD domains are zinc fingers, as the C4HC3 motif is coordinated and stabilized by two zinc ions in a cross-brace topology (Schindler et al., 1993). Structurally, PHD fingers are globular, and form a double stranded anti-parallel beta-sheet, with a C-terminal alpha helix (**Figure 4C**) (Kwan et al., 2003). There are also two unstructured, and flexible loops, one is positioned before the beta-sheet, and the second loop connects the alpha helix to the rest of the globular protein (Kwan et al., 2003; Sanchez & Zhou, 2011). The second loop is essential for mediating protein interactions of histones, as residues within the unstructured loop hydrogen bond with the N-terminal H3 amino group (Li et al.,

2006). Consequently, PHDs are largely found in nuclear transcriptional regulators, which can have the capacity to bind long stretches of DNA, (un)modified histones, and other proteins (Ragvin et al., 2004; Schindler et al., 1993).

Many PHD-containing proteins have two or more PHD zinc fingers, which suggest these chromatin-associated proteins function cooperatively to read multiple epigenetic marks (Sanchez & Zhou, 2011). PHD fingers have the capacity to read multiple epigenetic marks at different H3 residues – including H3K4me0/2/3, H3R20/2 and H3K14Ac (reviewed by Sanchez & Zhou, 2011). Similar to chromodomain proteins and histone methyl-lysine, PHD:H3K4me3 interactions are mediated by a four residue aromatic cage, comprised of Trp and Tyr (Li et al., 2006). Instead of aromatic residues, the binding of unmodified H3K4 is mediated by acidic and hydrophobic residues (Sanchez & Zhou, 2011). Structurally, PHD fingers are similar to RING zinc finger proteins which are also coordinated by two zinc ions, but contain a C3HC4 signature in cross-brace (**Figure 4D**) (Schindler et al., 1993). Interestingly, the Kap1 PHD-RING finger protein is an E3 SUMO ligase (Ivanov et al., 2007). The protein mediates it's own SUMOylation of an adjacent bromodomain (Ivanov et al., 2007). When complexed with SETDB1 and Chd3/Mi-2, SUMOylated Kap1 functions as a transcriptional regulator by repressing the recruitment of SETDB1 to target promoters (Ivanov et al., 2007).

1.40 Physical manipulation of nucleosomes by chromatin remodelers: SNF2 ATPases

1.410 The SWI/SNF family of proteins

In addition to the chemical, post-translational modification of histone proteins, the second way chromatin remodeling occurs is by the physical manipulation of nucleosomes such as disrupting or mobilizing nucleosomes to alter nucleosomal spacing. This is mediated by ATP-dependent enzymes of the SNF2 superfamily of proteins. SNF2 ATPases are classified into four main

groups, based on additional conserved domains. The first family is the SWI/SNF2 chromatin remodelers. The SWI/SNF family of proteins have bromodomains, which are known to be acetylated lysine readers. In *S. cerevisiae*, the SWI/SNF family contains yeast SNF2 and STH1, while in mammals there is BRM and brahma-related gene 1 (BRG1) (Sudarsanam & Winston, 2000). First discovered in yeast, the genes were identified in screens for mutants that affect mating-type switching (SWI) and sucrose fermentation (Sucrose Non Fermenting - SNF) pathways (Winston & Carlson, 1992). Genetic suppression screens of SWI/SNF mutants also suggested these proteins had chromatin-associated functions, such as maintenance of chromatin organization and nucleosome-binding activity (Winston & Carlson, 1992). Yeast SWI/SNF2 proteins form a multi-protein complex of 11 other proteins – which are collectively able to bind chromatin, and alter nucleosome structure (Martens & Winston, 2002). Proteomic approaches at the Fillingham Lab at Ryerson University (in collaboration with the Gingras Lab at the Lunenfeld/Tannenbaum Research Institute), have identified the first putative SWI/SNF complex in *Tetrahymena thermophila* – which include four yeast orthologs (Swi1, Swi3, Snf5, and Snf12), one ciliate-specific protein, three *Tetrahymena*-specific proteins, and an uncharacterized PHD-finger protein (Saettone et al., submitted).

1.420 The ISWI family of proteins

The second family of ATP-dependent SNF2-helicases is the ISWI family (Imitation Switch). In yeast, there are two ISWI family proteins - Isw1 and Isw2. Unlike their bromodomain-containing SWI/SNF counterparts, these proteins are known to have SANT and SLIDE domains – two other histone-interaction modules. ISWI family proteins have been well characterized in *Drosophila*, and are part of three essential chromatin remodeling complexes which catalyze nucleosome sliding: i) ACF (ATP-utilizing chromatin assembly and remodeling factor), ii) CHRAC

(chromatin accessibility complex) and iii) NURF (nucleosome remodeling factor) (Ito, et al., 1997; Tsukiyama et al., 1995; Varga-Weisz et al., 1997). Both ACF and CHRAC promote ordered nucleosomal positioning, while the NURF complex promotes irregular nucleosome spacing, specifically at promoters to enable transcription (Ito, et al., 1997; Tsukiyama et al., 1995; Varga-Weisz et al., 1997).

1.430 The INO80 family of proteins

INO80 chromatin remodeling proteins can be split into two groups based on the catalytically active complexes the proteins are found in: the INO80 remodeling complex and the SWR1 remodeling complex. INO80 was initially discovered in a yeast screen to identify inositol biosynthesis mutants (Ebbert, Birkmann, & Schüller, 1999). A complex of 15 proteins including actin-related proteins (ARPs are also found in SWI/SNF complexes), and RvB proteins (AAA+ family proteins, or ATPase Associated with diverse cellular Activities) were identified (Ebbert et al., 1999). INO80 has DNA-dependent ATPase and DNA helicase activity, as well as involvement in the DNA-damage response (Shen, Mizuguchi, Hamiche, & Wu, 2000; DNA damage response reviewed by Bao & Shen, 2007). The SWR1 complex is involved in the exchange of histone variants within chromatin. SWR1 exchanges the core H2A histone proteins within nucleosomes to the H2A variant, H2AZ (Mizuguchi et al., 2004). The yeast SWR1 complex is comprised of 14 proteins - with actin, Arp4, Rvb1 and Rvb2 being shared with the INO80 complex (Shen et al., 2000).

The last chromatin remodelers of the SNF2 superfamily are the **C**hromodomain **H**elicase-**D**NA-binding (CHD) proteins. The Chd family includes a number of proteins that are highly conserved from yeast to humans, though the function of many of these proteins remains poorly

characterized. The structure, function and epigenetically-associated biological processes of these ATP-dependent chromatin remodelers are discussed below.

1.50 The Chd family of chromatin remodelers

Chd proteins are SNF2-helicases, and are distinguished by the presence of two tandem N-terminal chromodomains (**Figure 5A**). These proteins are ATP-dependent enzymes, and function as transcriptional regulators during diverse chromatin-related developmental processes (Ho & Crabtree, 2010; Marfella et al., 2007; Woodage et al., 1997). The number of Chd proteins within metazoan species varies, with humans having the most, at nine Chd family members. In *S. cerevisiae,* there is only one Chd protein, yChd1. The unicellular protozoan *Tetrahymena* has two Chd proteins, Chd3 and Chd7 (**Figure 5B**). Phylogenetic analyses of tChd7 indicate it is a close ortholog of both human hChd1 and hChd7, due to it's domain architecture. Unique to Alveolate species, tChd7 does not have the classical tandem Chd chromodomains present in the N-terminal region. Rather, tChd7 has a chromodomain-bromodomain-chromodomain in the N-terminal region, as well as a triplicate C-terminal PHD finger region.

The Chd family of proteins can be divided into subfamilies, based on additional domains within the protein architecture. Subfamily I includes Chd1-2, which can be distinguished by the presence of a DNA-binding domain at the C-terminus (Woodage et al., 1997). The DNA-binding domain in mammalian proteins suggest an affinity for AT-rich stretches of DNA *in vitro* (Stokes & Perry, 1995). The second subfamily of Chd proteins, which includes Chd3 and Chd4 (also known as Mi-2α and Mi-2β), do not have DNA-binding domains. Subfamily II members harbour N-terminal PHD domains (Woodage et al., 1997). The third and last subfamily of CHD

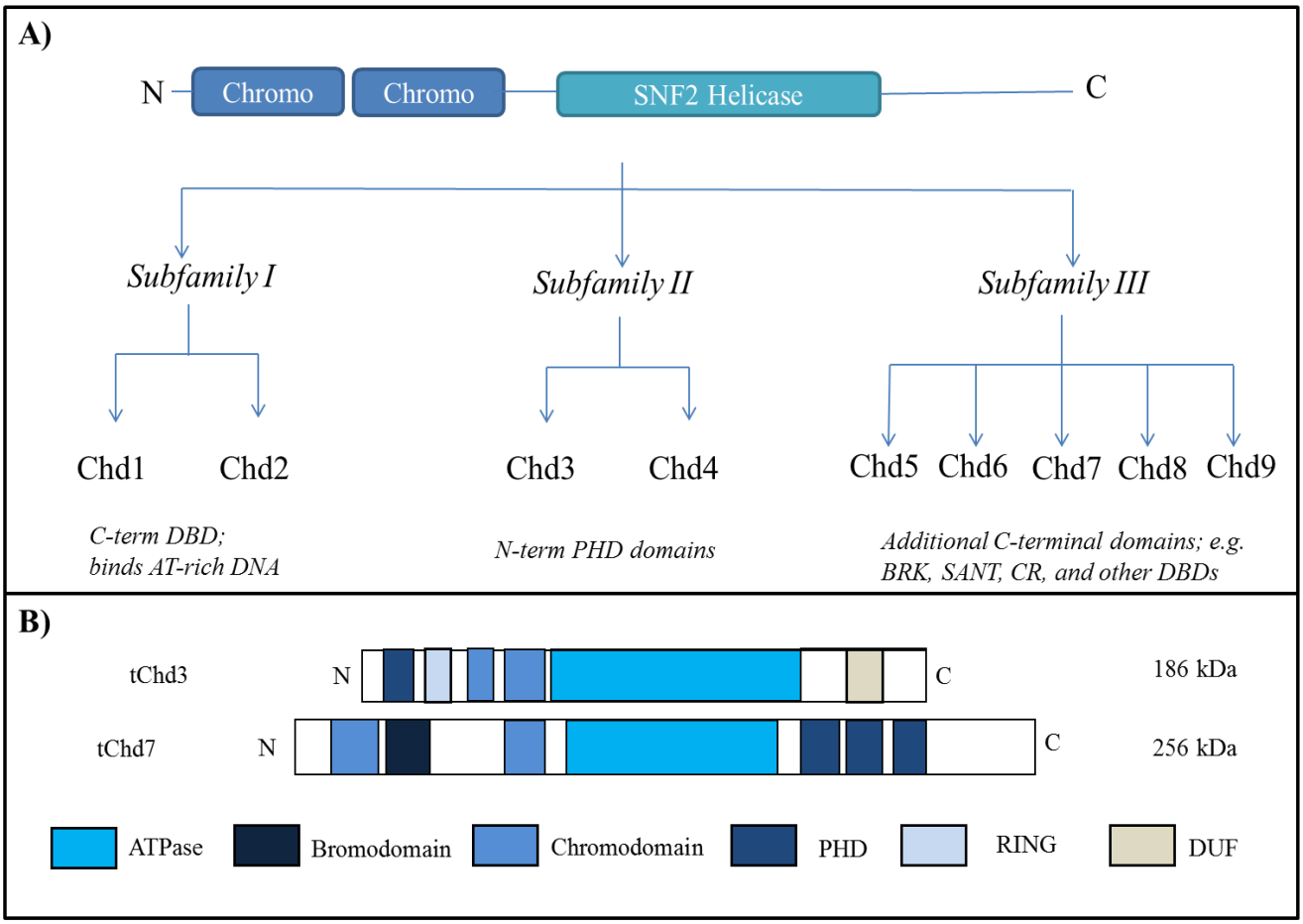


Figure 5. *The CHD family of chromatin remodelers.* **A)** Chromodomain Helicase-DNA-binding (CHD) chromatin remodelers are ATP-dependent helicases of the SNF2 superfamily. CHDs are characterized as having two tandem chromodomains in the N-terminus. CHDs are divided into three subfamilies, which are defined by additional accessory domains. **B)** *Tetrahymena thermophila* has two ~200 kDa CHD family proteins – tChd3 and tChd7. Both proteins have a large, central helicase/ATPase region. tChd3 displays the classical CHD subfamily II domain architecture, with two N-terminal chromodomains, an N-terminal PHD, and MIZ-SP/RING finger domain. An unstructured Domain of Unknown Function (DUF) is present at the C-terminal end of Chd3. *Tetrahymena* Chd7 has a very unique domain architecture, which only shows sequence similarity to other Alveolate species. Instead of N-terminal tandem chromodomains, tChd7 has a Chromo-Bromo-Chromo (CBC) N-terminal region, as well as three tandem PHD fingers (PPP) in the C-terminus.

proteins includes Chd5-9. The third family includes diverse members, but generally include BRK (Brahma and Kismet), SANT-like, Cysteine-rich, and DNA-binding domains in the C-terminal region of the protein (Marfella, Concetta G.A., 2007; Woodage et al., 1997). One Chd member from each subfamily is discussed below.

1.510 Chd1: a subfamily I member

A member of Chd Subfamily I, Chd1 was first identified as a DNA-binding protein in mouse lymphoid cells (Delmas, Stokes, & Perry, 1993). Chd1 is the most well-studied Chd family protein, and this ~200kDa protein is a critical component of transcriptional processes throughout all eukaryotic species (Murawska & Brehm, 2011). In yeast, Chd1 is able to directly bind the DNA major grove via it's SLIDE domain, activating gene transcription by promoting RNA Pol II to escape promoters by remodeling +1 nucleosomes, and binding the elongating RNA Polymerase II and elongation factors (Korber & Barbaric, 2014; Krogan et al., 2003; Pray-grant et al., 2002; Sharma, Jenkins, Héroux, & Bowman, 2011). ChIP-chip analyses on yChd1 also indicate the remodeler promotes RNA pol II-mediated transcription by increased histone H3 turnover at the 5' end of genes, and represses H3 turnover at 3' ends to restore the 3' chromatin structure in the wake of RNA pol II (Radman-Livaja et al., 2012). In the absence of yChd1, transcriptional termination does not occur, and there is increased presence of cryptic transcripts (Hennig et al., 2012; Quan & Hartzog, 2010).

Furthermore, human Chd1 is also able to bind gene-bodies, as well as transcription start sites (TSS) and enhancers-like regions in a transcription-coupled manner (Siggens et al., 2015). Ultimately, these functions manipulate chromatin structure and availability to transcriptional-related machinery by an unknown mechanism (Siggens et al., 2015). The transcription-related roles of Chd1 are likely a product of the interaction of Chd1 with multiple chromatin remodeling

complexes. Chd1 can directly bind components of the FACT (**Fa**cilitates **C**hromatin **T**ranscription) and PAF (Paf1-containing) complexes, where the Chd family protein may help these complexes mediate nucleosomal organization within ORFs of active genes, and transcriptional initiation/elongation, respectively (Krogan et al., 2002; Winkler & Luger, 2011; B. Zhu et al., 2005).

Mechanistically, perhaps the most well-characterized complexes Chd1 binds *in vivo* are SAGA and SLIK (**S**AGA-**lik**e), two transcriptionally-activating HAT complexes known to acetylate histones H3 and H2B (Pray-grant et al., 2002). Immunopurification of yChd1 and mass spectrometry co-purify native SAGA and SLIK components, including Gcn5, Ada, Taf and Spt (Pray-grant et al., 2002). Additionally, the second N-terminal chromodomain in yChd1 is able to bind the activating epigenetic mark H3K4me2 *in vitro*, and in turn promote SLIK-dependent acetylation of methylated H3K4 (Pray-grant et al., 2002). Mutations in the second yChd1 chromodomain ablates SLIK-mediated acetylation of H3K4, and loss of yChd1 results in increased global histone acetylation patterns *in vivo* (Pray-grant et al., 2002; Quan & Hartzog, 2010).

1.520 Chd3: a subfamily II member

Chd3 is a part of the second subfamily of CHD proteins. Chd3, otherwise known as Mi2-Alpha, is a core component of the Mi-2/NuRD complex (**Nu**cleosome **R**emodeling **D**eacetylase), a histone deacetylase complex associated with transcriptional repression (Xue et al., 1998). NuRD is comprised of ten key components: HDAC1, HDAC2, RbAp46 and RbAp48 (**R**etino**b**lastoma-associated proteins), MTA1, MTA2, and MTA3 (Metastasis-associated Proteins), MBD3 (methyl-CpG binding domain), Chd3 and Chd4 (Wade, Jones, Vermaak, & Wolffe, 1998; Xue et al., 1998). This complex is unique in that it couples histone deacetylation via HDAC activity to

both DNA methylation via MBD3, and ATP chromatin remodeling activity by Chd3/4 (Xue et al., 1998). Mi-2/ NuRD is hypothesized to be a major regulator of transcriptional repression, as it has the ability to directly catalyze the conversion of hyperacetylated euchromatin to transcriptionally silent and hypoacetylated heterochromatin (Denslow & Wade, 2007). It is hypothesized this occurs by the ATP-dependent nucleosome sliding ability of Chd3/4, and histone deacetylation ability of HDAC1/2 (Denslow & Wade, 2007). The deletion of chromodomains in *Drosophila* dChd3 and dChd4 inhibit their ATP-dependent nucleosome mobilization properties (Marfella et al., 2007).

Similar to other Chd family members, Chd3 is an important component of developmental processes (reviewed by Marfella et al., 2007). For example the *Arabidopsis thaliana* Chd3 homolog, Pickle, promotes root development by repressing embryonic target genes (Ogas et al., 1999). Pickle also opposes the action of Polycomb protein CLF, whereby Chd3 activates the transcription of Polycomb repressed target genes to inhibit de-differentiation (Aichinger et al., 2009). Pickle mutants are unable to germinate, and plants are unable to switch from embryonic to vegetative development (Ogas et al., 1999). In mice, Chd3 regulates the expression of transcription factors Pax6, Sox2, and Trb2, which promote neuronal layer specification (Nitarska et al., 2016). Mutations in Chd3 result in severe and specific defects to neuronal proliferation and migration within the developing cortex (Nitarska et al., 2016).

In humans, both Chd3 and Chd4 are also known to be autoantigens in the disease dermatomyositis, as 25% of sufferers present with Mi-2 antibodies (Yi Zhang, LeRoy, Seelig, Lane, & Reinberg, 1998). This inflammatory, connective tissue disease normally presents as an acute skin rash, but the systemic disorder can affect the muscles, joints, esophagus, lungs, and even heart (Dalakas & Hohlfeld, 2003). Sadly, dermatomyositis has a causative link to the onset

of cancer, with one quarter of those with dermatomytosis developing malignant cancers to different organ groups (Callen & Wortmann, 2006). The molecular mechanisms that lead to a Chd3 immune response, and why this immune response leads to the development of these diseases, remains unclear.

1.530 Chd7: a subfamily III member

A member of the third subfamily of CHD proteins, Chd7 remains one of the most elusive Chd proteins, as it's molecular mechanisms are poorly characterized. Mammalian Chd7 is implicated in rDNA biogenesis, as ChIP analyses in mouse embryonic stem cells suggest the chromatin remodeler binds hypomethylated, active rDNA, specifically promoting the transcription of the 45S pre-rRNA (Zentner et al., 2010). Mutations in *CHD7* commonly present as different disease states. Nonsense, frameshift, or missense mutations in the *CHD* gene are the major cause of the rare, autosomal and dominant genetic disorder CHARGE syndrome (**C**oloboma of the eye, **H**eart defects, Choanae **A**tresia, **R**etardation of growth, **G**enital abnormalities, and **E**ar abnormalities) (Zentner et al., 2011).

Phenotypically, almost 60% of CHARGE patients also suffer from diverse cerebellar disorders including underdeveloped olfactory bulbs, autism spectrum disorders (ASD), and other neurodevelopmental disabilities (T. Yu et al., 2013). Embryogensis is reliant on formation of the neural tube, which ultimately forms all the components of the nervous system. Defects in Chd7 are known to manifest in defects in segments of the neural tube which form the cerebellum (T. Yu et al., 2013). Interestingly, *CHD7* gene expression is enriched within the developing cerebellum, and Chd7-protein interactors within the cerebellum include Brg1, and three topoisomerases (Top1, Top2A, Top2B) (Feng et al., 2017). It is hypothesized that Chd7 maintains an open chromatin structure of differentiation-specific target genes during cerebellar

development, by recruiting other chromatin remodelers, as well as topoisomerases which help relieve helical tension with the winding and unwinding of DNA by chromatin remodelers (Feng et al., 2017).

Similar to other Chd family proteins, Chd7 has an affinity for specific epigenetic marks. In human neural crest cells, Chd7 binds PBAF, a chromatin remodeling complex that includes the transcriptional activators BRG1, BAF (**B**RG1-**a**ssociated **f**actor), ARID2 (**A**T-rich **I**nteracting **D**omain) and BRD7 (Bromodomain-containing protein 7) (Yan et al., 2005). Together, Chd7 and PBAF bind and co-localize with H3K27Ac to active promoters, as well as super enhancers (genomic regions that contain multiple enhancer units that promote the expression of transcription factors that are involved in cell identity) (Feng et al., 2017). Furthermore, Chd7 promotes the transcription of *TWIST1* by binding an upstream genetic element marked by H3K4me1. TWIST1 is a transcriptional regulator, and it's expression is essential for cell migration by promoting **e**ndothelial-to-**m**esenchymal **t**ransition (EMT) of neural crest cells (Bajpai et al., 2010).

1.540 Histone 3 Lysine 4 (H3K4) and Chd family proteins

The epigenetic state of histone 3 lysine 4 (H3K4) is associated with active transcription. H3K4 can be mono-, di- and tri-methylated, as well as acetylated. The di and tri-methylated states of H3K4 are both associated with active genes. H3K4me2 marks active genes and inactive yet poised genes; H3K4me3 marks transcription start sites (Sims & Reinberg, 2006). Unlike acetylation, which directly facilitates the euchromatinziation of genes, methylation of H3K4 is hypothesized to act as a binding site for chromatin remodeling proteins, which can enzymatically alter chromatin structure and promote transcription (Fischle et al., 2003).

The trimethylated state of H3K4 is also co-expressed with epigenetic marks associated with heterochromatin, such as methylated H3K9 or H3K27 (Vakoc et al., 2006). It is hypothesized that the presence of both activating (e.g. H3K4me3) and silencing (e.g. H3K9me3 and/or H3K27me3) marks allow genomic loci to remain poised for gene expression by enabling activating or silencing factors to bind at any time, which is especially important at developmental gene loci (Vakoc et al., 2006).

Set9-mediated methylation of H3K4 also inhibits binding of the repressive NuRD complex (Vakoc et al., 2006). Specifically, mammalian Chd1 binds H3K4me3 via it's tandem chromodomains in the 5' region of active genes (Flanagan et al., 2005). Binding of the chromatin remodeler promotes gene expression by manipulating chromatin structure, and prevents the binding of NuRD with the H3 tail, ultimately promoting transcriptional activation (Flanagan et al., 2005). Human Chd1 has two conserved aromatic residues in most N-terminal chromodomains, creating an aromatic cage (Trp322 and Trp325) with which the H3K4 methyl groups can bind (Flanagan et al., 2005). Structural analysis of mammalian Chd4 also shows a preference for both unmodified H3K4 and H3K9me1/2/3 (Flanagan et al., 2005).

Chromatin immunoprecipitation and microarray (ChIP-chip) analysis of Chd7 in human colorectal carcinoma cells, human neuroblastoma cells, and mouse ESCs suggested Chd7 also localizes with a histone H3 methylated lysine 4 (H3K4me) at gene loci (Schnetz et al., 2009). Chd7 and H3K4 methylation patterns change concomitantly, suggesting H3K4me may be developmentally-specific for Chd7 binding (Schnetz et al., 2009). However, the tChd1/tChd7:H3K4me2/3 interaction is not conserved. Yeast Chd1 lacks conserved methyl-binding aromatic residues and does not bind H3K4me3 (Sims et al., 2005). yChd1 facilitates

transcriptional elongation, by binding the PAF complex and RNA pol II (Murawska & Brehm, 2011).

1.550 H3K36 and Chd family proteins

The functional relevance of epigentically-modified H3K36 remains somewhat elusive. Similar to H3K4, this site on histone H3 can be both acetylated, and multiply methylated. The acetylated state of H3K36 is indicative of active transcription, and seems to colocalize with other H3 acetylation sites of transcription such as H3K4Ac, H3K9Ac and H3K14Ac (Larson et al., 2010). Acetylated H3K36 localizes to promoter regions of RNA pol II-transcribed genes. In yeast, the histone acetyltransferase Gcn5 is responsible for H3K36 acetylation (Larson et al., 2010). Gcn5 is a key component of transcription complex SAGA/SLIK, which includes other transcriptionally activating proteins - such as Chd1. *In vitro*, purified Gcn5 has the ability to acetylate histones at H3K36, and *in vivo* knockout studies of Gcn5 suggest loss of H3K36 acetylation (Larson et al., 2010). The acetylated state of H3K36 is inversely, but directly related to the methylation status of H3K36 (Larson et al., 2010).

Unlike the methylation of H3K9 and H3K27, methylation of lysine 36 is associated with euchromatin, and there is a progressive shift from monomethylation, to trimethylation from promoters to 3' ends of actively transcribed genes bodies (Barski et al., 2007; Butler & Dent, 2012). The methyltransferases responsible for altering the methylation state of H3K36 are the SET domain-containing proteins Set2 (yeast) and Nuclear Receptor SET domain-containing 1, NSD1 (mammalian); H3K36 is demethylated by Rph1 (Butler & Dent, 2012a). While the monomethylated state of H3K36 is poorly defined, H3K36me2 is implicated in DNA damage response, whereby the epigenetic mark is deposited at double-strand DNA breaks (Fnu et al., 2011). The mark then acts as a docking site for repair machinery such as NBS1 and Ku70 (Fnu et

al., 2011). The trimethylation status of H3K36 however, is relatively well studied and is involved in transcriptional elongation processes.

In yeast, an intricate model of epigenetic regulation of H3K36ac and H3K36me3 as an "acetyl/methyl switch" is proposed during transcriptional elongation (Butler & Dent, 2012a; Morris et al., 2007). First, the C-terminal domain of actively transcribing RNA pol II is phosphorylated at a serine residue, which acts as a docking site for the methyltransferase Set2 (Butler & Dent, 2012a). Set2-mediated trimethylation of H3K36 occurs during elongation and at gene bodies in the wake of RNA pol II (Butler & Dent, 2012a). This is followed by the induction of Eaf3, a component of the Rpd3 deacetylase complex which reads Set2-methylated H3K36me3 via it's chromodomain (Butler & Dent, 2012a). Eaf3 initiates deacetylase activity after RNA pol II has transcribed the gene body, ultimately restoring the chromatin structure and preventing pre-initiation within coding regions (Butler & Dent, 2012a). ChIP analysis of Eaf3 suggests binding to coding regions is contingent on the chromodomain *and* Set2-mediated H3K36 methylation. yChd1 is intimately associated with acetylation and methylation states of H3K36. yChd1 binds Set2 and RNAPII, where it prevents the exchange of preacetylated histones over ORFs, further suppressing the expression of transcripts from cryptic promoters (Butler & Dent, 2012a). The role of H3K36 in transcriptional elongation is conserved in mammals, however both methylated H3K9 and H3K36 are essential for preventing aberrant transcription (Bartke et al., 2010).

Interestingly, the methylation state of H3K27 and H3K36 are inversely related. The methyltransferase PRC2 can trimethylate H3K27, and methylation spreading along chromatin are essential for Polycomb silencing in mammalian cells (Yuan et al., 2011). In HeLa cells, histones unmethylated at H3K36 are methylated at H3K27, and H3K27me3 is not found

concomitantly with H3K36me2/3, suggesting H3K36 methylation inhibits the PRC2-mediated spread of H3K27 methylation (Yuan et al., 2011). As will be discussed below, the methylation state of histones during *Tetrahymena* development, specifically at H3K9 and H3K27, are of great interest, as they are essential for RNA-mediated DNA elimination events of epigenetically modified histones during development.

1.60 DNA and RNA binding proteins (DRBPs)

1.610 DRBP conservation throughout the tree of life

In a review by Bernstein & Allis (2005), the molecular intricacies of DNA, protein and RNA interactions in the context of histone PTMs among diverse organismal models were highlighted. The authors outlined an emerging view of macromolecular interactions which suggest DNA, RNA and proteins can interact as one functional unit within the cell, in epigenetically-regulated complexes called DNA and RNA-binding proteins (DRBPs). From yeast to mammalian cells, the conservation of DRBPs among eukaryotes is remarkable. Non-chromatin related DRBPs are even conserved among bacteria. This is exemplified in the prokaryote *Escherichia coli*, where Cold Shock Protein (Csp) family members are DRBPs; CspE catalyzes DNA condensation and binds nascent RNA transcripts in transcriptional complexes, both *in vivo* and *in vitro* (Hanna & Liu, 1998). Examples of chromatin-associated and epigenetically-regulated DRBPs within yeast and mammalian cells will be outlined below. First, RNAi will be briefly discussed, due to it's intimate tie to RNA guided silencing events.

1.620 RNAi: the precursor to small RNA-guided silencing events

RNA interference (RNAi) is an ancient process that is evolutionarily conserved among eukaryotes for over a billion years, and is hypothesized to have been a byproduct of eukaryotic

defense against viral RNAs and transposable elements (Obbard et al., 2009). In addition to RNAi, RNAi-like mechanisms that include epigenetically-regulated heterochromatin formation have evolved in many species, including yeast, *Tetrahymena*, and mammals. RNAi was first discovered in the early 1990s as a post-transcriptional gene knockdown mechanism in *Caenorhabditis elegans*, and current advances in molecular biology have taken advantage of this *in vivo* mechanism to target genes of interest in diverse models (Fire et al., 1991). RNAi begins with the production of long double-stranded RNAs (dsRNAs) that are similar in sequence to the post-transcriptional gene product targeted for silencing (Ketting et al., 2001). These dsRNAs are cut by the enzyme Dicer, an RNase III ribonuclease, which produces short ~25 bp small interfering RNAs (siRNAs) (Ketting et al., 2001). These siRNAs then bind in a protein complex named RISC (RNA-induced silencing complex). One protein within this complex, an Argonaute protein, contains siRNA-binding PAZ and PIWI domains (PPD), and has the ability to bind the double-stranded siRNAs (Liu et al., 2004). Once bound, Argonaute has RNaseH-like activity and can cleave one of the siRNA ds-strands (named the passenger strand), and remain bound to the other strand (named the guide strand) (Liu et al., 2004). When bound to the guide strand, the RISC complex is active, and is quite literally guided to endogenously transcribed mRNAs via hybridization, targeting the mRNA for destruction by Argonaute endonuclease activity (J. Liu et al., 2004).

1.630 The RITS Complex: siRNA-guided heterochromatin formation in yeast

In the fission yeast *Schizosaccharomyces pombe*, an RNAi-like process mediates gene silencing of repetitive DNA elements, centromeric regions, telomeric regions and mating-type loci (Obbard et al., 2009). This is mediated by the RITS (**R**NA-induced **i**nitiation of **t**ranscriptional **s**ilencing) complex. RITS couples epigenetic histone modifications and siRNA-binding proteins

to transcriptional silencing. The RNA-mediated gene silencing mechanism is contingent on the presence of the HP1 homolog Chp1, a small-RNA binding Argonaute/PIWI protein Ago1, a dicer protein Dcr1, and Tas3 (**Ta**rgeting complex **s**ubunit 3, which interacts with Ago1) (Obbard et al., 2009). During RITS-mediated silencing, siRNAs are produced from the transcriptional silencing of heterochromatinized yeast target regions (e.g. centromeric regions). First, long double stranded RNAs are processed by the RNase III–like activity of Dcr1 to generate ~22nt siRNAs, which are then recruited onto the RITS complex by binding to Ago1 (Obbard et al., 2009).

The RITS complex also binds the histone methyltransferase Clr4 (Zhang et al., 2008). The siRNAs are hypothesized to guide the H3K9 methyltransferase complex to targeted loci, nucleating heterochromatin formation and spreading via the dimethylation of H3K9 (Zhang et al., 2008). The chromodomain protein Chp1 then is able to bind H3K9me2 modified histones, which is stabilized by Tas3, to enable RNA:chromatin, sequence-specific interactions (Obbard et al., 2009). It is hypothesized that siRNA-RNA interactions, instead of siRNA-DNA interactions help stabilize heterochromatin formation and spreading. This is because RITS recruits an RNA-dependent RNA polymerase following H3K9me-mediated target recognition (Sugiyama et al., 2005). A positive feedback loop ensues, to amplify the number of siRNA species generated (Sugiyama et al., 2005). In short, when the Clr4 methyltransferase complex enables spreading of H3Kme2 heterochromatin, the RNA-dependent RNA polymerase transcribes a new round of dsRNAs from original DNA target sequences, which are also processed into siRNAs by Dcr1 (Sugiyama et al., 2005; Zhang et al., 2004). Collectively, these processes contribute to a constitutively heterochromatinized state of RITS targets in fission yeast (Sugiyama et al., 2005).

1.640 Xist: lncRNA-mediated dosage compensation of the female X chromosome in mammals

A functionally and structurally well-characterized long non-coding RNA species directly involved in RNA-guided and epigenetically-mediated silencing, is the mammalian Xist RNA. Female mammalian species have developed a mechanism that simultaneously enable the presence of an active X chromosome, and target inactivation of the other X chromosome within the same nucleus. This is achieved by Xist. This ~17 kb RNA is processed similar to mRNAs, in that it has a 7-methylguanosine cap on the 5' end, is spliced, and has a 3' poly-A tail (reviewed by Bernstein & Allis, 2005). Yet it remains in the nucleus (Bernstein & Allis, 2005). This process is initiated at the chromosomal region known as the X inactivation center (Xic) (Bernstein & Allis, 2005). On the X chromosome, Xic is transcribed into the Xist RNA, which then remains in the nucleus and coats the female X chromosome (Bernstein & Allis, 2005). As a regulatory mechanism, the Xic can also transcribe the antisense Xist RNA named Tsix (Bernstein & Allis, 2005). Once post-transcriptionally processed, the 40kb Tsix transcript can bind Xist, sequestering it from the nucleus and repressing Xist from X inactivation (Bernstein & Allis, 2005).

Xist-mediated silencing is contingent on the epigenetic modification of the X chromosome. Poly-A pulldown of Xist using antisense sequences, and subsequent mass spectrometry, identified that Xist interacts with SHARP, a complex that binds HDAC3 (McHugh et al., 2015). The involvement of an HDAC is implicated in the preliminary steps of X chromosomal inactivation. Xist can direct HDAC3, enabling hypoacetylation of the X chromosome on histone H3 and H4, initiating nucleosome compaction and heterochromatinization (McHugh et al., 2015). With the help of the nuclear scaffolding protein SAF-A, the Xist RNA physically binds to the methylated

states of H3K9me2, H3K27me3 and H4K20me on the female X chromosome (Helbig & Fackelmayer, 2003).

The mammalian Polycomb repressive complex 2 (PRC2) is the histone H3K27 methyltransferse (Plath, 2003). The histone-lysine N-methyltransferase G9a is implicated in the dimethylation of H3K9, while Set7 is implicated in H3K20 monomethylation (Kohlmaier et al., 2004; Rougeulle et al., 2004). The PRC1 complex which contains two RING-type zinc finger proteins, is also necessary for the ubiquitination of histone H2A lysine 119 (H2AK119ub1), and is temporally-linked to the H3K27 methylation status of the inactive X (de Napoles et al., 2004). Transcription of Xist and binding to the inactive X chromosome begins on heterochromatinized and silent genes, but eventually spreads to euchromatinized regions of the X chromosome (Cerase et al., 2015). Accumulation and spreading of Xist transcripts across the chromosome are reliant on an essential RNA structure. Two repetitive poly-A stemloops linked by a U-rich linker in the 5' region of Xist is conserved among most mammalian species, and functions to recruit the Polycomb PRC2 complex (Maenner et al., 2010). Mutations in this repetitive region completely block X inactivation (Maenner et al., 2010). It is hypothesized that the poly-A stemloops may function as a scaffold, whereby they coordinate the interaction between different proteins within the PRC2 complex (Johnsson et al., 2014).

1.650 Towards the Identification of Novel DRBPs

Many chromatin-associated proteins, especially transcription factors, have been identified to bind mRNA products, and this function may regulate their turnover by preventing translation (Castello et al., 2012). Other RNAs in DRBP complexes may also function as an effector, recruiting proteins to a specific DNA locus, or a scaffold, by acting as a co-activator to promote binding of chromatin-associated proteins (Cassiday & James, 2002; Hudson & Ortlund, 2015). It

remains unclear how most DNA:RNA:protein complexes interact *in vivo*, as it is unlikely that RNAs form R-loops with DNA (i.e. DNA:RNA hybrids) (Skourti-Stathaki & Proudfoot, 2014). This is because the resulting displaced, single-stranded DNA from R-loop interactions are unstable, and more amenable to DNA mutations; it would be evolutionarily disadvantageous for co-activating RNAs to promote genomic instability (Skourti-Stathaki & Proudfoot, 2014). It is clear however, that DRBPs play dynamic and diverse roles in gene expression, developmental regulation, and disease (reviewed by Hudson & Ortlund, 2015). Thus expanding the knowledge of chromatin-associated proteins, and their potential regulation by RNA species, is critical to the understanding of fundamental cellular processes.

In 2012, two groups were monumental in expanding the DRBP compendium with novel high-throughput methods (Baltz et al., 2012; Castello et al., 2012). Castello et al. (2012) characterized the mRNA-binding protein interactome in HeLa cells, in an "mRNA interacome capture". The researchers identified 860 mRNA binding proteins, a vast majority of which are non-classical RNA-binding proteins (Castello et al., 2012). Not only were 499 novel RBPs identified, but many were classical chromatin-binding proteins, DNA helicases, and/or single-stranded DNA-binders (Castello et al., 2012). The authors noted that nucleic-acid binding zinc finger proteins were enriched in the mRNA capture, with 31/67 of these zinc-containing proteins having both DNA and RNA-binding GO annotations (Castello et al., 2012). Hendrickson et al. (2016) analyzed 28 of the chromatin-associated proteins with non-classical RNA binding domains identified in the mRNA capture. RNA immunoprecipitation and sequencing analyses identified widespread binding to structurally diverse lncRNAs and mRNAs (G Hendrickson et al., 2016). Interestingly, Chd4 (a Chd3 interactor within the NuRD complex) bound the human lncRNA, Xist - identifying Chd4 as a novel XIST-binding protein and DRBP (G Hendrickson et al., 2016).

The animal model, *Tetrahymena thermophila*, is a ciliate protozoan that undergoes developmentally-regulated and RNA-guided DNA elimination. These developmental processes are contingent on the orchestration of siRNAs, lncRNAs, epigenetics, and chromatin-associated proteins. *Tetrahymena* are suited for investigating DRBPs for three main reasons: i) their unique cellular physiology, ii) the availability of DNA and RNA binding proteins throughout development, and iii) the presence of diverse RNA species during conjugation. A brief introduction to *Tetrahymena*, in addition to the presence of DRBPs in relation to irreversible genome silencing mechanisms, are discussed below.

1.70 *Tetrahymena thermophila*

1.710 A Historical Model

Tetrahymena thermophila is a ciliated, unicellular protozoan (**Figures 6A and 6B**). This non-parasitic and fresh-water protist has a molecular complexity comparable to humans, making it a useful model organism to study complex eukaryotic molecular mechanisms (Frankel, 2000). One reason for this is the organisms large transcriptome (J Xiong et al., 2012). During vegetative growth, the *Tetrahymena* transcriptome represents 55% of the active genome (J Xiong et al., 2012). This is more than ten times larger than the human transcriptome, since less than 5% of the human genome is transcribed (Caron et al., 2001). Yet this ciliate is simpler to work with than mammalian model organisms. As compared to the mouse model, *Mus musculus,* the high homologous recombination frequency of *T. thermophila* allows for highly-efficient functional analysis of gene knockouts and knockins (G. L. Yu, Hasson, & Blackburn, 1988). Discoveries of fundamental eukaryotic mechanisms using *T. thermophila* have paved the way for many

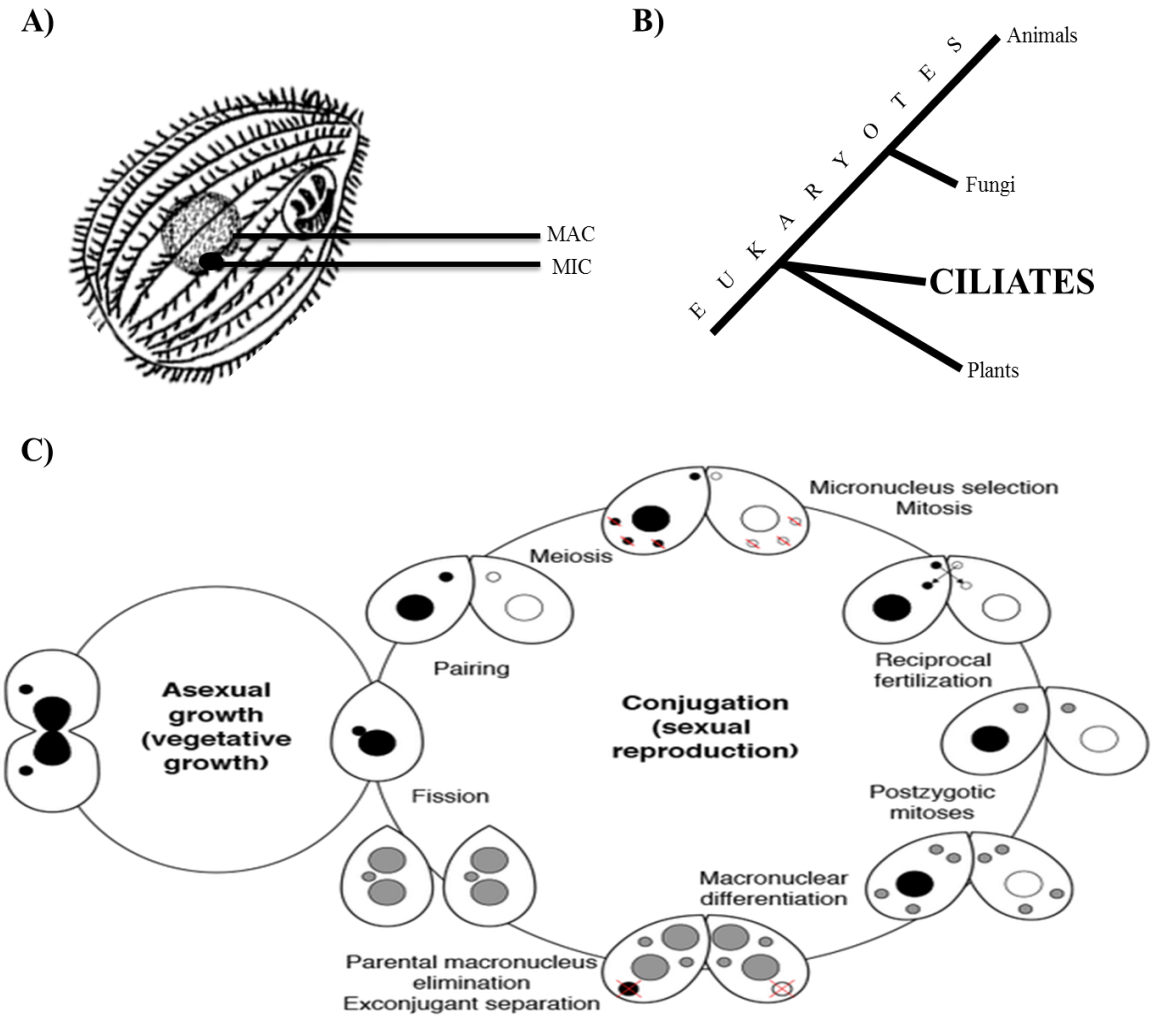


Figure 6. *Tetrahymena thermophila is an animal model useful for studying chromatin dynamics.* **A) and B)** *T. thermophila* is a nuclear dimorphic ciliate that contains two structurally and functionally distinct nuclei within the same cytoplasm. The large polyploid MACronucleus is transcriptionally active (euchromatinized) during vegetative growth, while the small diploid MICronucleus is transcriptionally silent (heterochromatinized). **C)** Vegetatively, the cell divides amitotically. Two cells of different mating types can undergo sexual development (i.e. conjugation) during starvation conditions. In short, a series of mitotic events, meiotic events, nuclear exchange and nuclear degradation result in the generation of exconjugants that are genetically distinct to pre-conjugants. Image adopted from Yin, Gater and Karrer (2010).

important discoveries in modern molecular biology proving the humble ciliate to be a valuable and historically significant model organism for studying eukaryotic cells. One such noteworthy breakthrough was the Nobel-prize-winning discovery of catalytic RNA, or ribozymes (T R Cech, 1990). This discovery uncovered one of the many multifaceted roles of non-coding RNAs which would lead to future studies on RNA functions and their molecular applications and initiated ideas about an "RNA world" (Thomas R Cech, 2015). The Nobel-prize winning discovery of telomeres and telomerase, the identification of linker histone H1 as a crucial element to chromatin compaction, the demonstration of histone acetylation as a transcriptional activator, and identification of small RNA-mediated genome rearrangements are among other remarkable discoveries using *Tetrahymena* (Brownell et al., 1996; Greider & Blackburn, 1985; Mochizuki et al., 2002*;* Shen & Gorovsky, 1996). *Tetrahymena* is a model organism that can be grown in diverse laboratory conditions. In rich medium, the ciliate divides vegetatively. The cells have a rapid doubling time of about 2.5 hours, and can grow exponentially to a concentration of 10^6 cells/mL, optimally at 28^{o}C (Cassidy-Hanley, 2012). *Tetrahymena* can also reproduce sexually via conjugation. A cell from one of the seven mating types can pair with a different mating type under nutrient starvation. During conjugation, many remarkable molecular processes occur - involving genome rearrangements, chromatin remodeling, and processes involving RNAi-like mechanisms.

1.720 Nuclear Dimorphism

Similar to other members of the phylum Ciliophora, *T. thermophila* has two nuclei within the same cytoplasm. This phenomenon is referred to as nuclear dualism, or nuclear dimorphism. *T. thermophila* has a small, diploid germinal micronucleus (MIC), and a large polyploid somatic macronucleus (MAC). The presence of two nuclei, in two distinct chromatin states, makes this

eukaryote an ideal model for studying chromatin dynamics. Genomic sequencing of the MAC was completed in collaboration with The Institute of Genomic Research (TIGR). Prof. Pearlman's lab at York University was an important contributor to this genome project (Coyne et al., 2008; Eisen et al., 2006). Sequencing of the MIC was done in collaboration with the Broad Institute of Harvard and MIT (E. P. Hamilton et al., 2016). Sequencing and microarray data on gene expression throughout vegetative growth, starvation and conjugation, are publicly available on the *Tetrahymena* Genome Database (TGD - www.ciliate.org) and the *Tetrahymena* Functional Genome Database (TFGD - http://tfgd.ihb.ac.cn/) (Miao et al., 2009; J Xiong et al., 2012).

1.730 Life Cycle

The polyploid MAC (~45 copies) is transcriptionally active throughout vegetative growth, and is composed of regularly positioned nucleosomes downstream of transcription start sites, while the MIC is transcriptionally silent and features delocalized nucleosomes (Gorovsky, 1973; Jie Xiong et al., 2016). During vegetative growth, cells divide amitotically. The five diploid chromosomes of the MIC are only transcriptionally active during a short period of time during conjugation (Gorovsky, 1973). As mentioned, when two *Tetrahymena* cells of different mating types pair to reproduce sexually, sexual recombination occurs alongside a series of complex genome rearrangements (Cervantes et al., 2013). An illustration of *Tetrahymena* conjugation is provided in **Figure 6C**.

The process occurs as follows: the MIC undergoes meiotic division, producing four haploid pronuclei. One of these nuclei divides once mitotically, whilst the other three are degraded. The selection and degradation of nuclei is epigenetically-driven, whereby the selected pronucleus is acetylated on H3K56 (Akematsu et al., 2017). One of the new, mitotically-derived daughter

nuclei is exchanged with that of the other mating cell. These two exchanged haploid pronuclei fuse (a process called reciprocal fertilization), forming a new diploid zygotic nucleus. This zygotic nucleus is identical in each mating cell. The diploid nucleus then undergoes two further rounds of mitosis resulting in four diploid nuclei. One nucleus is degraded, one matures into the new MIC and the last two form new MACs. The old MAC in the cell is degraded via autophagy (Akematsu, Pearlman, & Endoh, 2010). The development of the new MAC, a stage called Analgen, requires many programmed genome rearrangements (Mochizuki & Gorovsky, 2004a).

The ultimate goal of conjugation is to produce offspring genetically unique to the parents. Therefore, both the MIC and MAC in conjugating *Tetrahymena* cells require genome rearrangements to occur. During Analgen formation, rearrangements allow the formation of the large polypoid MAC genome, from the small diploid MIC genome. There are two main forms of genome rearrangements in *Tetrahymena*: i) chromosome breakage, and ii) the removal of internally eliminated sequences (IESs) (Yao et al., 1990; Yao et al., 1984). Chromosome breakage involves the fragmentation of the five MIC chromosomes at specific chromosome breakage sequences (CBSs). This cutting is coupled with telomere formation; the end result is the formation of ~280 new MAC chromosomes of approximately 700 kB (Hamilton et al., 2005). The second form of genome rearrangement involves irreversible genome silencing in the form of the removal of micronuclear IESs. The MIC genome contains around 6000 IESs (Nikiforov et al., 2000). These highly repetitive, heterochromatinized regions of DNA are likely degenerate transposons, and they are excised during the formation of the MAC (Cherry & Blackburn, 1985; Yao et al., 1984).

As mentioned, RNA Inference (RNAi) is an evolutionarily conserved process that can inhibit gene expression in *Tetrahymena*, as well as in other eukaryotes. *Tetrahymena* undergoes an

RNAi-like process during conjugation. The mechanism occurs during formation of Analgen, when IES-containing sequences are excised from the MIC, during formation of the new MAC. This irreversible genome silencing was first described by Mochizuki et al. (2002), and was later named the scan RNA model (Mochizuki & Gorovsky, 2004b). An illustration of this mechanism is provided in **Figure 7**. In this model, DNA is first bi-directionally transcribed in the transiently transcriptionally-active MIC, forming long dsRNAs. These dsRNAs are cleaved by the *Tetrahymena* Dicer analog, Dcl1p (Mochizuki & Gorovsky, 2005). The result is ~28bp RNA fragments, called scan RNAs (scRNAs) (Mochizuki & Gorovsky, 2005). These scRNAs bind to Twi1, a *Tetrahymena* Argonaute protein with PPD domains (Mochizuki et al., 2002). This RNA-protein complex (similar to the RISC-guide siRNA complex), localizes to the MAC, where scRNAs homologous to the parental MAC are degraded (Mochizuki et al., 2002). Non-homologous scRNAs remain intact and bound to the Twi1 protein, and move to the new developing MAC (Analgen) (Mochizuki et al., 2002). The scRNA-Twi1 complex binds chromodomain proteins named Pdds (**P**rogrammed **D**NA **D**egradation) ultimately leading to the recognition and degradation of IESs (Aronica et al., 2008; Madireddi et al., 1996; Mochizuki et al., 2002).

IES heterochromatinization via chromatin remodeling, specifically though H3K9 and H3K27 methylation, is central to the recognition and removal of IESs (Liu et al., 2004; Yifan Liu et al., 2007). The primary histone lysine methyltransferase responsible for H3K27 methylation in *Tetrahymena* is the protein Ezl1, and H3K27 methylation regulates H3K9 methylation (Yifan Liu et al., 2007). Pdd proteins have DNA-binding chromodomains, and bind both methylated H3K9 and H3K27 initiating DNA elimination (Liu et al., 2007; Taverna et al., 2002). The Pdd proteins present within scanRNA:Twi1 complexes are able to bind IES targets via scanRNA

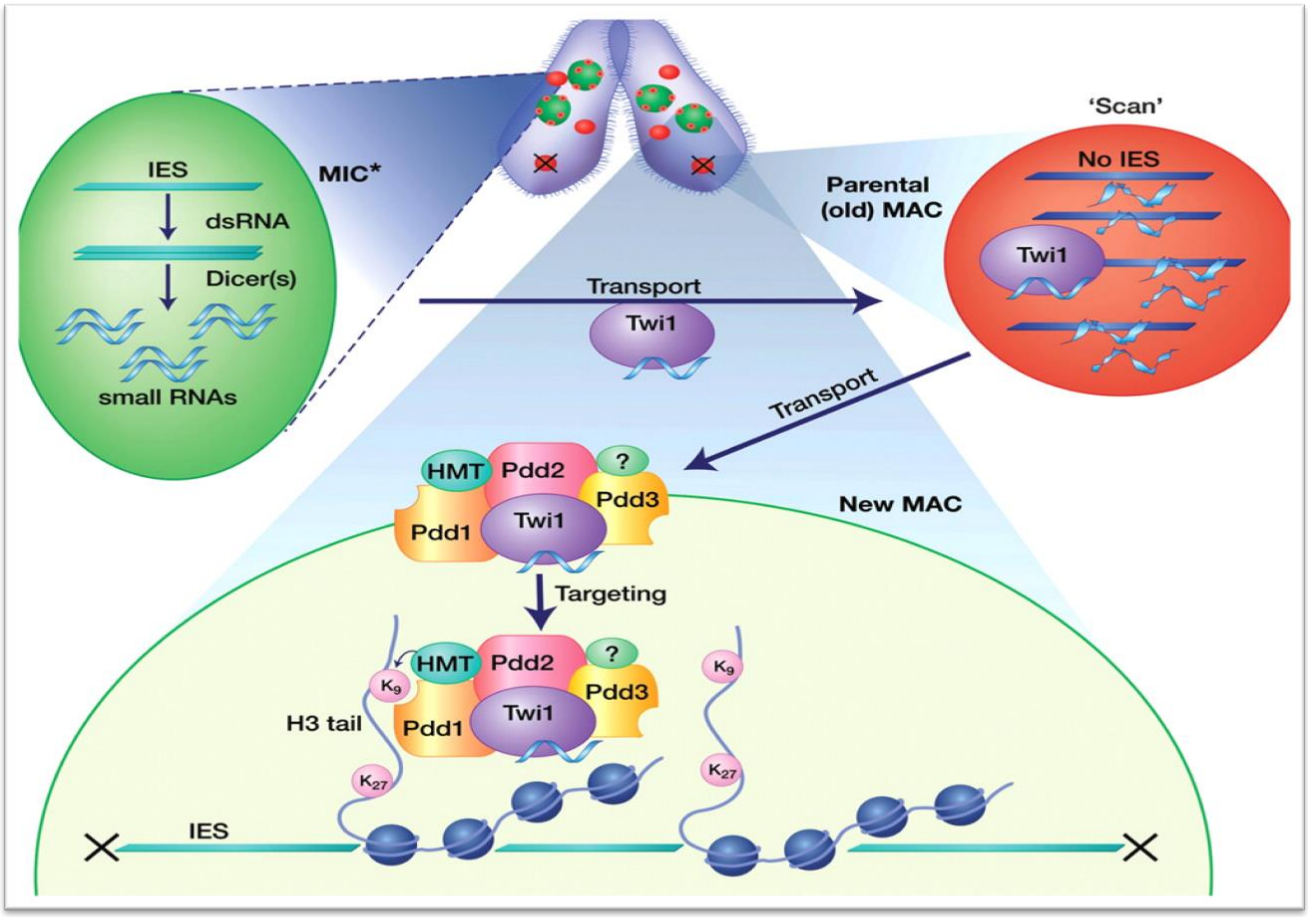


Figure 7. *Development of new macronuclei during Tetrahymena conjugation: the ScanRNA Model.* In an RNAi-like process, development of new macronuclei from the parental micronuclear genome is mediated by RNA-guided DNA elimination (irreversible genome silencing). In short, the parental MIC is bi-directionally transcribed and processed into ~28nt scan RNAs (scRNAs). One of the scRNA strands binds the Piwi protein Twi1 and the complex is transported to parental macronuclei. In homology-directed selection, scRNAs that are homologous to the old MAC are degraded. Non-homologous (i.e. IES-containing) scRNAs are transported by Twi1 to the developing macronuclei where they form a multi-protein complex which selectively binds and leads to the excision of genomic IESs. The illustration of the Scan RNA model is by Bernstein and Allis (2005).

sequence homology, bend the DNA, and move the ends of IESs together for efficient excision (Coyne et al., 1999; Madireddi et al., 1996; Nikiforov et al., 2000). *Tetrahymena* mutants with inhibited histone methylation, have a significantly lowered efficiency of IES removal (Y Liu et al., 2004).

IESs themselves have a modest AT-rich bias, but some are flanked by 5' G5 DNA tracts that identify sequence boundaries (Carle et al., 2016). Recent studies have indicated that transcription of 3' C5 strands in developing macronuclei leads to G5-containing lncRNAs that form G-rich quadruplexes with the unwound 5' G5 DNA (Carle et al., 2016). The excision of IESs is contingent on the formation of this DNA:RNA hybrid, as the protein RNA helicae protein Lia3 specifically binds the G-quadruplex and recruits the domesticated transposase Tpb2, which ultimately catalyzes the final step of IES excision (Carle et al., 2016; Vogt & Mochizuki, 2013). Furthermore, recent studies have suggested the presence of Type I and Type II IESs, which are excised by early and late scanRNAs, respectively (Noto et al., 2015). The transcription of late scanRNAs from developing macronuclei are enhanced *in trans*, by early scanRNAs (Noto et al., 2015). As a result, there is the production of long non-coding RNAs both during the early stages of development (from parental micronuclei), and the later stages of development (from developing macronuclei), to generate the two forms of IESs (Aronica et al., 2008; Noto et al., 2015). The presence of long dsRNAs in parental macronuclei have also been observed, however their functional relevance remains elusive (Aronica et al., 2008; Woo, Chao, & Yao, 2016). It is also hypothesized that these lncRNAs interact with scRNAs in an uncharacterized, homology-directed regulatory mechanism of gene expression within parental macronuclei (Aronica et al., 2008). The presence, function, structure, and sequence bias of the lncRNA species in *Tetrahymena* development are uncharacterized (Woo, Chao, & Yao, 2016). Further, how the

highly heterochromatinized IESs become accessible to chromatin excision machinery has yet to be identified.

1.740 Tetrahymena Chd family proteins are candidate DRBPs

The *Tetrahymena* Chromodomain Helicase DNA-binding proteins Chd3 and Chd7 are hypothesized to be essential, pre-emptive factors to IES excision, and the mechanisms by which they mediate developmental processes remain unknown (Fillingham et al., 2006). In a screen for ATP-dependent chromatin remodeling proteins in *T. thermophila*, the Pearlman Lab previously identified and partially characterized two conserved Chd proteins of the SNF2 superfamily, Chd3 and Chd7. Chd proteins are ATP-dependent enzymes which physically and directly modify histone-DNA interactions by mobilizing nucleosomes, to alter gene expression. As previously described, Chds are critically important ATP-dependent chromatin remodeling proteins, and are involved in a wide variety of cellular functions and human diseases, such as chromatin regulation, oncogenesis, stem cell pluripotency, and rRNA transcription (Gaspar-Maia et al., 2011; Schnetz et al., 2009; Zentner et al., 2010). These Chd proteins are candidate DRBPs for five reasons.

First, as mentioned, an emerging view of DNA-binding proteins suggests these proteins bind and/or are regulated by RNAs. Many chromatin-associated and ATP-dependent chromatin remodelers bind diverse RNAs in processes such as mRNA post-transcriptional regulation, as well as ncRNA-mediated transcriptional silencing (G Hendrickson et al., 2016). For example, the SWI/SNF chromatin remodeling complex in *A. thaliana* occurs by lncRNAs-mediated transcriptional silencing (Wierzbicki et al., 2013). Secondly, *Tetrahymena* Chd proteins are tentative DRBPs through bioinformatics comparison of the mRNA::protein interactome in mammalian cells (Baltz et al., 2012; Castello et al., 2012). Third, Chd3 and Chd7 have unique

expression profiles. RNAseq and microarray data suggest the Chd proteins peak in gene expression almost exclusively during RNA-guided genomic rearrangements within *Tetrahymena* conjugation. In *Tetrahymena,* DNA-binding chromatin remodelers promote recognition and removal of IESs during irreversible genome silencing. The abundance of both long and short RNAs during this complex process, suggest DRBPs likely play a role in these genome rearrangements (Woo et al., 2016)

Fourth, Chd3 and Chd7 have unique domain architectures, containing classical DNA-binding CHROMO, BROMO, PHD, and ZMIZ domains. Chromodomains and zinc finger proteins have been noted to bind both DNA and RNA in diverse eukaryotic families. For example, the ATP-dependant SWI/SNF chromatin remodeling complex in *A. thaliana* is guided by lncRNAs-mediated transcriptional silencing (Y. Zhu, Rowley, Böhmdorfer, & Wierzbicki, 2013). Chromodomains and zinc fingers can bind both DNA and RNA. The *D. melanogaster* protein MSL (male specific-lethal) must bind the lncRNA Rox2 via it's chromodomain for dosage compensation of the male X chromosome (Akhtar et al., 2000). Fourth, tChd3 and tChd7 are candidate DRBPs due to their expression profiles available on the Ciliate Genome Database, www.ciliate.org. DNA microarray and RNAseq expression data suggest the Chds peak in gene expression almost exclusively during RNA-guided genomic rearrangements. Lastly, preliminary lab analyses on tChd3 and tChd7 suggested these chromatin remodelers co-localize to sites of these rearrangements, and peak in protein expression during this landmark stage.

1.80 Thesis Project: Macromolecular Interactions of Tetrahymena CHD proteins during growth and development

1.810 General overview

The main objective of this research project is to perform functional analyses on CHD proteins to better elucidate their developmental functions and *in vivo* mechanisms. The first approaches include the analysis of novel macromolecular interactions with CHD proteins. The elucidation of novel CHD interactors is of special interest, because the role of these chromatin remodeling proteins, especially in the context of the *Tetrahymena* scRNA model, is not fully understood. *In vivo* and *vitro* models were used to characterize tChd-binding macromolecules on the RNA, chromatin and protein level. For *in vivo* analyses, a three-way immunoprecipitation protocol is used. **P**rotein **I**mmuno**p**recipitation (PIP), **Ch**romatin **I**mmuno**p**recipitation (ChIP), and **R**NA **I**mmuno**p**recipitation (RIP) will identify DRBP interactions, in an experimental setup called IP3 (**Figure 8A**). Protein-tagged *Tetrahymena* strains for *CHD3* and *CHD7* had been previously constructed using CU428 (*Mpr*/*Mpr* [VII, mp-s]) and B2086 (*Mpr+*/*Mpr+* [II, mp-s]) cell strains, of the inbred family B. These strains were constructed by transformation using gene gun biolistic bombardment, by shooting FZZ (3X FLAG, TEV cleavage site and Protein A domain) containing plasmids into *Tetrahymena* cells and recombining with the *Tetrahymena* genome, and tagging the *CHD* genes of interest with this epitope marker. These engineered CHD tagged strains are essential for the affinity purification of CHD-macromolecule complexes, because an enrichment and high-yield purification of CHD proteins is vital to the identification of novel interactors. RNA interactors were identified by Urea-PAGE analysis. Chromatin interactions were identified by qPCR. Protein interactor identification was completed via mass spectrometry, with collaborators from the Lunenfeld-Tanenbaum Research Institute at Mount Sinai Hospital. Previous PIP studies in the Pearlman Lab identified a MIZ/SP-RING finger protein, here referred to as Miz1, to interact with Chd3 in vegetative growth. This novel interaction is analyzed in depth.

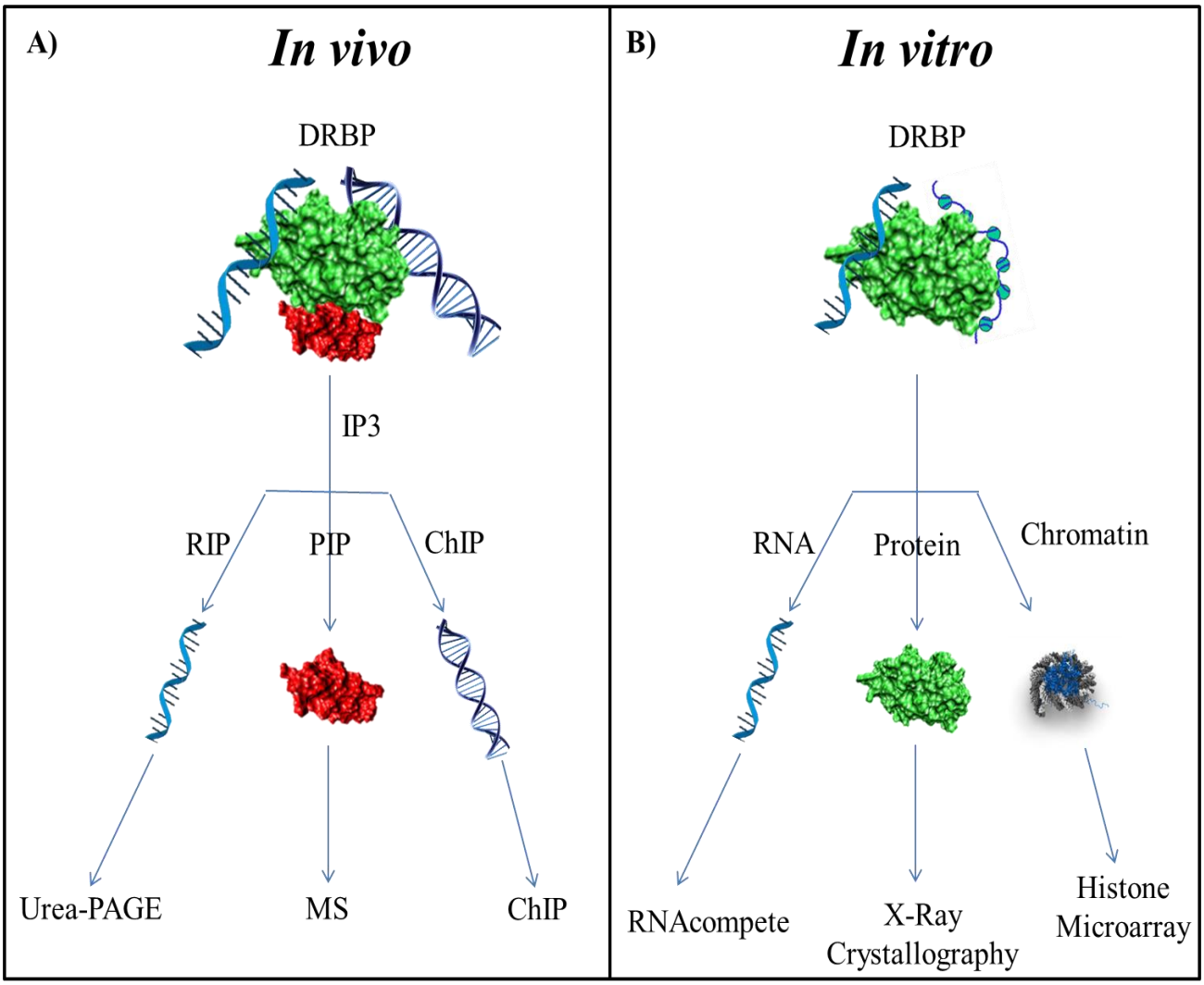


Figure 8. *Macromolecular interactome analysis for Tetrahymena CHD family proteins.* To identify macromolecular interactions of the candidate DRBPs tChd3 and tChd7, *in vivo* and *in vitro* approaches were used to characterize RNA, DNA and protein interactions. *In vivo* analyses involved three-way immunoprecipitation (IP3) of tChd proteins. Specifically, protein immunoprecipitation (PIP), RNA immunoprecipitation (RIP), and Chromatin immunoprecipitation were conducted, followed by macromolecular-specific functional analyses. *In vitro* analyses of tChd7 involved protein crystallography trials, RNAcompete for motif-binding analysis, and Histone Peptide Array analysis.

In parallel to *in vivo Tetrahymena* IP3 assays, two synthetic recombinant constructs of the CHROMO-BROMO-CHROMO (CBC) and PHD-PHD-PHD (PPP) regions of Chd7 were commercially generated, for *in vitro* functional analyses (**Figure 8B**). tChd7 was selected for *in vitro* analyses because of it's domain architecture which contains both tentative epigenetic reader domains and RNA-binding domains in a unique organization. The CBC and PPP regions were expressed and purified in *E. coli* using 6xHis and GST-containing vectors. *In vitro* chromatin interactions were analyzed by histone peptide microarray. The commercial peptide microarray, originally developed by Nady et al. (2008), contains 384 histone modification sites. Histone PTMs can act as docking sites for chromatin-binding proteins which read and write the "histone code". This array screened hundreds of potential interaction sites such as H3K9 or H3K27, for Chd7 binding.

RNA interactions were analyzed by the RNA motif assay called RNAcompete. RNAcompete is an *in vitro* method for HTP identification of RNA-binding specificity of RBPs, developed by Debashish Ray and Timothy Hughes from University of Toronto, Department of Molecular Genetics. The assay works by co-incubating GST-purified proteins with a 7-mer RNA pool. Following incubation, bound RNAs are purified and identified by microarray analysis. Lastly, *in vitro* analyses on the protein level were conducted by HTP X-ray crystallography trials in collaboration with the Hui Lab, at the Structural Genomics Consortium, in Toronto. A series of supplementary analyses on tChd knockouts (KOs) examined gene essentiality, provided a global view of tChd function within the cell. In short, the goal of this project was to explore the binding affinities of these chromatin remodelers, and conduct functional analyses guided by unique protein structure, to better understand the role of Chds within the cell.

2.00 Materials and Methods

Specifications of *Tetrahymena* strains used in this thesis are provided in **Table 1**. Centrifuge models for all small-scale centrifugations (1.5 ml volumes or less) and large-scale centrifugations are listed in **Table 2**. Primer sequences used for PCR, qPCR, and *in vitro* transcription are listed in **Table 3**. PCR reactions were conducted using the T100 Thermal Cycler (BioRad). Bacterial cultures were grown using the Innova 2300 platform shaker (New Brunswick Scientific). *T. thermophila* were grown using the G10 gyrotory shaker (New Brunswick Scientific).

2.10 Preliminary Analyses

2.110 Bioinformatic mRNA/Proteome Comparison

Supplementary information from Castello et al. (2012) identified 860 RNA-binding proteins (RBPs) with UV-crosslinking, followed by oligo dT-pulldown and mass spectrometry. Many proteins that were identified were also DNA-binders, such as transcription factors or chromatin remodeling proteins. *Tetrahymena* homologs were identified by annotating NP RefSeq identifiers of strong mRNA-binding mammalian proteins. The 860 mammalian proteins were then subjected to BLASTp sequence analysis using the *Tetrahymena* Genome Database (www.ciliate.org) to identify TTHERM_ protein annotations. At BLASTp $p=e^{-100}$, Chd3, Chd7 Chd2, SNF2-related proteins and other chromatin remodelers were identified. A statistical algorithm developed by Castello et al. (2012) for predictive power, scored both Chd3 and Chd7 as significant RBPs, but with low precision.

2.120 Cytoscape Gene Expression Maps

T. thermophila gene co-expression data for *CHD3*, *CHD7* and *MIZ1* were extracted from TFGD http://tfgd.ihb.ac.cn/. The *Tetrahymena* Gene Network was from microarray expression data with a minimum Z-score threshold 3.49 (Miao et al., 2009). The larger of a Z-score between two genes indicated the more reliable interaction (more similar expression profile). The downloaded networks were generated using Cytoscape 3.5.1, as undirected gene networks using the top 30 highest gene co-expressers.

2.130 Immunofluorescence

Tetrahymena samples were collected at every stage of the life cycle; vegetative growth, starvation and up to 14 hours of conjugation. Cells in 1.5 mL of culture were fixed with 2:1 mercuric chloride:ethanol solution. Cold acetone was added to permeablize cells, which were then pelleted by centrifugation at maximum speed, and washed with 1xPBS (1.86mM NaH_2PO_4, 8.41 mM Na_2HPO_4, 175 mM NaCl). The primary Mouse M2 antibody was suspended in 1xPBST (1XPBS + 0.1% Tween20), and samples were incubated at room temperature for one hour. After washing in 1xPBS, fluorescently conjugated secondary antibody (Rhodamine Red, goat anti-mouse) was added and incubated at room temperature for one hour in the dark. 1xPBS washed cells were incubated with DAPI (4',6-diamidino-2- phenylindole) solution. Wet mount slides were prepared and visualized by confocal fluorescence microscopy (Olympus BX51) at 400x total magnification (40X objective, 10X eyepiece), using Z plane superimposition.

2.20 *In vivo* binding affinities: IP3

2.210 T. thermophila growth for immunoprecipitation

For vegetative growth, tagged *T. thermophila* strains were grown overnight at 30^{o}C in 150 mL of SPP media (2% proteose peptone, 90 µM sequestrene) supplemented with 1X PSF antibiotic cocktail (ThermoFisher, Cat. 15240062). Cells were harvested at approximately 2-3x10^{5} cells/mL. Cultures were centrifuged at 3000 rpm for two minutes. To study vegetative growth, cell pellets were washed once with 10 mL 10mM Tris-HCl pH 7.5 and immediately frozen at -80^{o}C for downstream immunoprecipitation analysis. For starvation, the logarithmic cell pellets were washed three times with 10mM Tris-HCl pH 7.4. The cells were re-suspended in 300 mL of 10mM Tris-HCl pH 7.4, and the cells were starved at 30^{o}C for 10-18 hours without shaking. The pellet collected as above was immediately stored at -80^{0}C. For mating or conjugation, starving cell cultures of different mating types with equal cell concentrations, were mixed (0 hours). Cells were collected by centrifugation at the desired time point in mating, and the cell pellet was immediately frozen at -80^{o}C for downstream immunoprecipitation.

2.220 TCA Protein Precipitation

A 1 mL aliquot of *Tetrahymena* (vegetative, starved, or conjugating) were centrifuged at maximum speed for 3 minutes at room temperature. Cell pellets were washed with 10mM Tris-HCl pH 7.4. Cells were re-centrifuged and the supernatant was removed to leave 100µL. 10uL of 100% TCA was added and the cell lysate was incubated on ice for 30 minutes. Following incubation, lysate was centrifuged at maximum speed for two minutes at room temperature. The supernatant was discarded and pellets were resuspended in 50-100µL of 2X SDS loading dye (4% SDS, 160 mM Tris-HCl, pH 6.8, 20% glycerol, Bromophenol blue, 10% BME). The pH was adjusted until the Bromophenol pH indicator returned to a blue colour, with 5-10µL of 1N

NaOH. Samples were boiled for five minutes, vortexed, and centrifuged at maximum speed for 5 minutes at room temperature for SDS-PAGE analysis.

2.230 SDS-PAGE Analysis

Protein samples were loaded and electophoresed through 6-12% SDS-polyacrylamide gels (5% stacking gel: 5% polyacrylamide, 0.12% SDS, 150mM Tris-HCl pH 6.8, 0.1% APS and 0.05% TEMED) at 100V in Western Running Buffer (25 mM Tris, 190 mM glycine, 0.1% SDS) for 1.5 hours. SDS-PAGE gels were transferred onto nitrocellulose blotting membrane (Pall Life Sciences BioTrace NT) overnight at 15 volts in Western Transfer Buffer (25 mM Tris, 192 mM glycine, 20% methanol). Membranes were blocked with 5% Blotto (5% skim milk powder diluted in 1XPBS), on a rotary shaker at 80 rpm (New Brunswick Gyrotory Water Bath Shaker, Model G76) for one hour at room temperature. Primary antibody (anti-rSUMO, anti-IgG, anti-mBetaActin, anti-mH3K36me3, anti-mH3K36Ac, anti-mFLAG) were diluted 1:10,000 in 5% Blotto at room temperature for one hour. Following incubation, membranes were washed three times in PBST for ten minutes at room temperature. Blots were incubated with secondary antibody (goat-anti rabbit or goat-anti mouse) diluted 1:10,000 in 5% Blotto at room temperature for one hour. Following incubation, membranes were washed three times in PBST for ten minutes at room temperature. Western blots were visualized using 2 mL of Chemiluminscent HRP Antibody Detection Reagent (BioRad, 1705060) for five minutes. Blots were exposed to autoradiography film (Denville Scientific Inc. E3018) as necessary. Semi-quantification of Western blots was performed using ImageJ software. The signal of each band (anti-IgG or anti-FLAG) was subtracted from background signals, and divided over beta-actin for normalization. Individual replicates were averaged, and plotted on Microsoft Excel.

2.240 Affinity Purification (Protein immunoprecipitation)

For one step-immunoprecipitation, cell pellets were resuspended in lysis buffer (10 mM Tris–HCl pH 7.5, 1 mM $MgCl_2$, 300 mM NaCl, 0.2% NP40, and yeast protease inhibitors (Sigma)). Centrifugation-cleared lysates (16000xg, 30 minutes, 4^oC) were applied to 50uL anti-FLAG agarose beads (Sigma, A4596). The resin and lysate was left for 4h with end-to-end rotation. Following incubation, protein:bead complexes were washed five times, once with IPP300 (10 mM Tris–HCl pH 8.0, 300 mM NaCl, 0.1% NP40), three times with IPP300 without NP40, and three times with IPP100 (10 mM Tris–HCl pH 8.0, 100 mM NaCl). Immunoprecipitants were eluted with 750uL 0.5M NH_4OH. Protein input samples (centrifugation-cleared lysates) and immunoprecipitants following ammonia elution were either analyzed by silver staining (Proteo Silver Stain Kit, Sigma) or analyzed by Western blot. For Western blot, proteins were electrophoresed through 8-10% SDS-PAGE, transferred, and probed with anti-mBeta Actin, anti-mIgG and/or anti-mFLAG.

Tandem purifications for SUMO trials were performed as described above. However, lysis buffer was supplemented with 50mM N-ethylmalemide (Sigma E3876), and centrifugation-cleared lysates were first applied to 100uL anti-IgG beads (ThermoFisher, 20333). Following washes, beads were incubated with 30μg TEV protease overnight at 4^o. The supernatant containing FLAG-tagged protein was applied to anti-FLAG beads. The rest of the procedure is as described above.

2.250 Mass Spectrometry (in collaboration with the Gingras Lab)

Mass spectrometry was completed by Dr. Jean-Phillipe Lambert, in collaboration with the Gingras Lab at the Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital. In short,

immunoprecipitants were dessicated at 4°C by speed vacuum and re-suspended in concentrated HCl. Denatured peptides were subsequently trypsinized, and loaded onto a 75 µm reversed-phase high-performance liquid chromatography capillary column, which was equilibrated with 2% acetonitrile and 0.1% formic acid (Dr-Maisch GmbH; Germany). The protein species were ionized, and analyzed with a linear trap quadrupole (LTQ) mass spectrometer by tandem mass spectrometry (MS/MS).

Peptides were identified by their mass to charge ratio, and the resulting data were analysed by Mascot, using the *T. thermophila* protein database (NCBI). Protein interactors were bioinformatically filtered by SAINT (Significance Analysis of the Interactome, (Choi et al., 2011)). In short, individual spectral counts of identified protein interactors are quantitatively ranked. The SAINT algorithm probabilistically scores protein spectra, and aims to identify significant and reproducible interactors by accounting for negative control and replicate data.

2.260 RNA immunoprecpitation

RNA immunoprecipitation was conducted as described above for two-step immunoprecipitation. However, lysis buffer was supplemented with 1unit/µL Murine RNase Inhibitor (NEB, M0314). Purified and TEV-cleaved protein eluates were suspended in TriZOL reagent, and RNA immunoprecipitants were extracted. RNAs were electrophoresed at 150V for 1.5h on 12% Urea-PAGE in 1X MOPS running buffer (7M urea, 1X TEB (45 mM Tris-borate/1 mM EDTA), 12% polyacrylamide, 0.1% APS and 0.05% TEMED). Urea-Page gels were pre-run for 30 minutes, and wells were flushed out with MOPS buffer every 5 minutes, prior to loading of RNA samples. Urea-PAGE gels were stained in 1X SybrGold (ThermoFisher S11494) diluted in RNase-free 1X TEB buffer. Gels were imaged by UV transillumination.

2.270 Chromatin Immunoprecipitation

A 100 mL culture of exponentially-growing *Tetrahymena* was crosslinked with 37% formaldehyde, at a 1:40 dilution. Cells were mixed on a platform shaker for 30 minutes, and the reaction was quenched with 300 mM glycine. Crosslinked pellets were resuspended twice with ice-cold Tris-HCl pH 7.4 and subsequently lysed in 1 mL ice-cold lysis buffer (50mM Tris-HCl pH 8.0, 5mM EDTA, 1% SDS, 1mM PMSF and Sigma Protease Inhibitor Cocktail). Lysates were sonicated on ice (Fisher Scientific, Model 120) for 10 cycles, 25 times (0.3 seconds on and 6 seconds off) at 30% amplitude. Sonicated samples were centrifuged at 15000xg for 30 minutes at 4^{o}C. During centrifugation, 50uL packed anti-FLAG beads per sample were equilibrated with dilution buffer (50mM Tris-HCl pH 8, 150mM NaCl, 5mM EDTA, 1% Triton X, 0.1% NaDeoxycholate), Aliquots of soluble extracts were taken for analysis of protein integrity by anti-mFLAG and anti-mIgG Western blot, and chromatin integrity by PCR of the Alpha Tubulin gene (*ATU1*), rDNA, M-element and R-element). The remaining cleared lysate was applied to anti-FLAG beads, and left for overnight rotation at 4^{o}C. Beads were washed once with FA Buffer (1mM EDTA pH 8.0, 50mM HEPES pH 7.5, 140mM NaCl, 0.5% Triton X-100, 0.5% NaDeoxycholate), LiCl buffer (0.25 M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate) and finally 1X TE. The chromatin was purified by incubating the beads in 50ug/mL Proteinase K diluted in Proteinase K buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5% SDS), at 42^{0}C for 2h and 65^{0}C for 8h.

2.280 qPCR

Quantitative PCR analysis of immunoprecipitated chromatin from Chd3, Chd7, and Miz1 expressing cells, and wildtype cells was conducted. Data was expressed in IP/In enrichment, using the Pfaffl Method of Normalization. In short, this method of qPCR data analysis is reliant

on the generation of a DNA standard curve, or calibration line, using three or more different dilutions of DNA.

First, purified input and immunoprecipiated DNA were assessed for quality and quantity with an OD280nM NanoDrop spectrophotometer reading. The normalization of input DNA with respect to the immunoprecipitated chromatin was done for every repetition. This accounts for differences in PCR efficiency, primer pairs, and signal-to- background noise ratios. To assess quantitative differences in ChIP-qPCR samples, a standard curve using input chromatin was generated using three dilutions of input chromatin: undiluted input DNA, 1:10 diluted input DNA, and 1:100 diluted input DNA. The curve is essential as qPCR efficiency curves normally deviate from 100% efficiency reactions, even with minimal technical errors.

The three points of each standard curve and the normalized sample were conducted for every primer pair, which include the following: *ARS2* (alanine tRNA synthetase), *BTU* (beta-tubulin), *CIT2* (citrate synthetase), *HMGB* (transcription factor), *HSP70* (chaperone), *PDD1* (developmentally-regulated Programmed DNA Deletion protein, negative control), rDNA (ribosomal DNA), and *RPS2* (ribosomal protein 2 of the small subunit). Gene targets for qPCR were selected based on literature searches of *CHD3/7* or *CHD3/7*-complex bound loci, as well as gene targets that are co-expressed with the Chd family proteins during growth.

Quantitative PCR runs were performed using Bio-Rad Sybr Green Supermix (Cat. #1708880) in duplicate, and reactions were analyzed with the Bio-Rad CFX 96-well Real-Time System. In each qPCR run, input dilutions (to generate the curve), along with input and immunoprecipitated chromatin diluted to 2.5 ng/µL were run. Raw Cq values for input DNA and IP DNA were analyzed using the BioRad Prime PCR program, which normalizes these data to the generated

standard curve that we represented as % with respect to the input DNA. Ultimately, these normalized ChIP data are expressed as fold increase, by dividing normalized IP over normalized Input. A gene target was classified as 'successful', and a bona-fide chromatin immunopreciptant, if there is a minimum of 5-fold increase over IP/input, with minimal-to-no fold increase within wildtype samples. Heatmaps were generated using Microsoft Excel. Fold increase levels were converted to the log10 scale.

2.30 In vitro binding affinities of Tetrahymena Chd7

2.310 Induction, Expression and Purification of Tetrahymena Chd7 CBC and PPP Regions

A construct of the CHROMO-BROMO-CHROMO (CBC) region of tCHD7 was commercially generated based on sequence data provided by the Pearlman lab by DNA 2.0 (Menlo Park, California). The CBC region was cloned with an N-terminal 6xHis tag into the pJ441 expression vector, for inducible expression and purification. This vector was transformed into BL21 DE3 expression cells (NEB). Colonies were inoculated into a 10 mL YT+Kan (0.01% Bacto Tryptone, 0.005% yeast extract, 0.005% NaCl, 50 μg/ml Kanamycin) starter culture, and grown overnight at 37°C. The starter culture was used in a 1:50 dilution for 2L of TB+Kan (0.012% tryptone, 0.025% yeast extract, 0.004% glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4, 50 μg/ml Kanamycin) the following day. The culture was grown for 4 hours (OD_{600}=~0.4) at 37^0C . Cells were induced with 0.4mM IPTG, and grown overnight (12-18 hours) at 16^0C . Samples were electrophoresed through 10% SDS-PAGE, and gels were stained with Coomassie Blue. A construct of the PHD-PHD-PHD (PPP) region of tChd7 was also generated by DNA 2.0, however the protein was poorly expressed. A GST-tagged construct was generated by ligating the PPP region of pJ441 into pTH6838 using SacI and SalI sites, and transforming the construct into DH5 alpha (NEB).

Plasmids from successful transformants were subcloned into BL21 DE3 for expression and purification. This was also carried out to generate a GST-tagged CBC construct.

2.320 Histone Protein Microarray - MODArray

MODified Histone Peptide Array (Active Motif, Cat. 13005) and MODified Protein Domain Binding Kit (Active Motif, Cat. 13007) were used to analyse CBC histone binding interactions. The protein microarray has 384 sites, each with a different unmodified histone peptide, posttranslationaly modified peptide, or a combination of multiple modified peptides. Analogous to a Western blot, the array is first blocked with 5% milk for 1 hour, washed with TBS 3 times, and then incubated with the protein of interest (6xHis-CBC), for 1 hour. This is followed by 3 TBS washes, primary anti-6His antibody incubation (1 hour), 3 TBS washes, and secondary anti-rabbit HRP antibody incubation (1 hour). After a final set of 3 TBS washes, the array is incubated with ECL to detect CBC:Histone interactions, and imaged. The final image is processed with MODarray Array Analyze Software to detect CBC:Histone-binding interactions.

2.330 Histone Co-immunoprecipitation

Co-immunoprecipiation of tChd7 with H3K36me3 was adapted from Kacrich et al. (2016). In short, *Tetrahymena* pellets were lysed (10mM Tris-HCl pH 8.0, 150mM NaCl, 1mM DTA, 0.5% NP-40, 1mM PMSF and Sigma Protease Inhibitor Cocktail). 100U of Bensonase (Sigma) and 1mM $MgCl_2$ were added, and lysates were incubated for 1 hour at 4^0C. Following incubation, lysates were centrifuged at 15000xg for 30 minutes at 4^0C. The supernatant was collected, and applied to anti-FLAG beads for 2 hours of end-to-end rotation at 4^0C. Beads were washed once with Buffer 1 (20mM Tris-HCl pH 7.5, 500mM NaCl, 5mM EGTA and 1% Triton), and Buffer 2 (20mM Tris-HCl pH 7.5, 500mM NaCl, and 5mM EGTA). Samples were directly dissolved in

Laemmli sample buffer (300mM Tris-HCl pH 6.8, 10% SDS, 50% glycerol and 0.05% Bromophenol Blue), and boiled for 5 minutes. Samples were electrophoresed through 10% SDS-PAGE, transferred, and probed for H3K36me3 (Cell Signaling).

2.340 Protein Crystal Trials of CBC-6xHIS

The CBC-6xHIS construct was purified with Takara Nickel Resin Kit (Cat. 635677). Immediately following elution, 1mM DTT was added to increase protein stability by preventing disulphide bond formation of the cysteine-rich construct. Overnight dialysis in 2 litres of crystallization buffer (10 mM HEPES pH 8, 500 mM NaCl) was performed. Proteins were concentrated with Amicon 10KMW Size Exclusion columns (Cat.UFC900308), to a minimum of 5 mg/mL. Immediately before plating the protein into crystal trial plates, 1mM TCEP (Tri(2-carboxyethyl)phosphine hydro-chloride) reducing agent was added. In Sitting-Drop 96-well plates, one drop of protein and one of crystallization solution were added to each well (Hampton Research) using Phoenix Crystallization Robots (Art Robbins Instruments). Plates were observed at 24h, 48h, 96h, and then once a week for a month using the Nikon sm21500 Light Polarizing microscope. Pictures were taken with the Lleica m216a microscope.

2.350 Purification of CBC-GST and PPP-GST constructs

Induced cells from a 250 mL culture were centrifuged at 3700rpm for 15 minutes (4ºC). Cells were lysed in 35 mL of lysis buffer (20 mM HEPES pH 7.5, 0.1 mM EDTA, 1 M NaCl, 10 mM B-mercaptoethanol, 1 mM PMSF). To the lysate, 150µL of 100 mg/mL lysozyme was added, and the lysate remained on ice for 30 minutes. Extracts were sonicated on ice (Fisher Scientific, Model 120) 8 times (5 seconds on and 15 seconds off) at 50% amplitude. The whole cell extracts were centrifuged at 4^0C, 15000rpm for 20 minutes. Meanwhile, 250µL GST resin per protein

was transferred into a 15 mL Falcon tube. Beads were washed in 6 mL cold PBS three times, by inversion and centrifugation at 3700rpm for 1 minute (4^0C). With the final wash, the buffer was removed, and the supernatant from sonicated samples was added. The samples were mixed by end to end rotation at 4^0C for 2 hours. Following incubation, beads were washed once with Buffer I (2M NaCl, 10mM β-mercaptoethanol), once with Buffer II (1M NaCl, 10mM β-mercaptoethanol) and twice with Buffer III (0.1M NaCl, 20mM HEPES pH 7.5, 0.1mM EDTA, 10mM β-mercaptoethanol). Following washes, GST-purified proteins were eluted with 300uL Elution Buffer (50mM Tris-HCl pH 8.8, 30mM Reduced Glutathione, 20% glycerol, 10mM β-mercaptoethanol, 250mM NaCl). Eluates and beads were centrifuged at 3700rpm for 2 minutes (4^0C), and frozen at -80^0C .

2.360 RNACompete

RNACompete was conducted on the CBC-GST and PPP-GST constructs as described (Ray et al., 2009). In short, GST-purified proteins were incubated with an RNA pool encompassing a diverse and large set of RNA 7-mers. RBPs were purified, and RNAs are identified by microarray analysis. RNAcompete data produced motifs, consensus sequences and Z-scores for each RBP. RNA motif logos were created by aligning the top ten high-scoring 7-mers. RNAcompete experiments are judged to be successful, on the basis of (a) correlation among high-scoring 7-mers between replicate pools, and (b) the presence of a clear motif among the high-scoring 7-mers.

2.370 Probe labelling and purification

The 28nt DNA oligos were transcribed into RNA using the MEGAshortscript™ Kit High Yield Transcription Kit (Cat. AM1354). 20μL reaction mixes used for *in vitro* transcription included

2uL of 3000 Ci/mmol α-^{32}P GTP (PerkinElmer), 2 µL T7 10X Reaction Buffer, 8 µL nucleotide solution (75 mM), and 10 µM of template. The reaction was incubated overnight at 37^{0}C. Alternatively RNAs (G10 and A10) were radiolabeled with γ-^{32}P ATP (PerkinElmer) by T4 Polynucleotide Kinase (1X T4 Polynucleotide Kinase Buffer (NEB, B0201), two µL 200 µM γ-^{32}P ATP (PerkinElmer), 10U of T4 Polynucleotide Kinase (NEB, M0201), 25µM RNA Sequence (IDT), and ddH2O), and incubated for two hours at 37^{0}C.. To the radiolabelled RNA probes, 5uL of Formamide RNA loading buffer (10mM EDTA, 80% formamide, 0.06% bromophenol blue and 0.06% xylene cyanol) was added. The contents were heated for 10 minutes at 95^{o}C, and the entire contents were loaded onto a 12% urea gel. RNAs were electrophoresed through 100V at 4^{o}C for 90 minutes in 1X TEB buffer. Gels were exposed to HyBlot CL Autoradiograph Films (Denville Scientific, Cat. E3012). The addition of Glogos ® II Autoradmarkers (Agilent technologies Cat. 420201) beside gels acted as guides for gel purification of RNA probes, following development of films. Radioactively labelled RNA bands of the correct size were excised from the gels. RNAs were eluted from the gel by dissolution into 150-200µL 0.5M NaCl overnight at room temperature. Purified RNAs were assessed for radioactivity with a scintillation counter (HiDex, 300XL).

2.380 Electrophoretic Mobility Shift Assay (EMSA)

An Electophoretic Mobility Shift mastermix (300mM KCl, 0.1M EDTA, 200mM Tris-HCl pH 7.6 and 1mM βME), 8µL of bromophenol blue dye was added to the labelled RNA probes, and the contents were incubated at 95°C for 5 minutes, and crash cooled on ice to 37°C. The Chd7-PPP protein was serially diluted in EMSA dilution buffer (1X EMSA mastermix, 10% glycerol), and co-incubated with 3000cpm of each RNA, at 37^{0}C for 20 minutes, with 1uL of RNase Inhibitor. The protein:RNA mix was chilled on ice for 5 minutes, electrophoresed through a 6%

native polyacrylamide gel (1X TEB, 40% polyacrylamide, 0.08% APS, 0.037% TEMED) at 100V for 1 hour, at 4^0C. The gel was dried for 30 minutes at 80^0C (BioRad Gel Drier), and exposed to a phosphorimager screen overnight. The next day, films were scanned with the Phosphor Imager Typhoon Trio+ (GE Healthcare).

2.40 Knockout analysis of t*CHD3*, t*CHD7* and t*MIZ1*

2.410 DAPI Staining

Cells in 1.5 mL of culture were fixed with 2:1 $HgCl_2$:EtOH solution. Cold acetone was added to permeablize cells. Cells were pelleted, washed with 1xPBS and incubated with DAPI (4',6-diamidino-2- phenylindole). Wet mount slides were prepared and visualized by fluorescence microscopy (Olympus BX51) at 400x total magnification.

2.420 Genomic DNA extraction, and PCR of Knockouts

A 1 mL aliquot of exponentially growing *Tetrahymena* was re-suspended in 500µL of *Tetrahymena* lysis solution (0.35 M NaCl, 0.01 M Tris-HCl pH 7.4, 0.01 M EDTA, 1% SDS, 42% urea), and subsequently 600µL of phenol:chloroform (1:1). The lysate was centrifuged at room temperature for 1 minute at 13000 rpm. The aqueous layer was transferred to a new micro-centrifuge tube, 600µL of phenol:chloroform (1:1) was added, and the sample was re-centrifuged. The top layer was again separated, 300µL of chloroform was added, and the sample was centrifuged. 200µL of 5M NaCl was added to the aqueous layer, followed by 800µL of isopropanol. The sample was centrifuged at room temperature for 10 minutes at 13000 rpm. The supernatant was discarded, and the DNA pellet was washed with 70% ethanol. The dried DNA pellet was re-suspended in 100µL of sterile ddH2O.

2.430 RNA extraction, and qPCR of Knockouts

For total RNA extraction from *Tetrahymena*, a 1 mL aliquot of exponentially growing cells were pelleted and resuspended in 1 mL of TriZOL Reagent (Invitrogen). Following a one minute incubation, 200µL of chloroform was added, and the sample was shaken vigorously by hand. The sample was centrifuged for 15 minutes at 3,000 rpm, at 4˚C. To the aqueous phase, 500µL of isopropanol was added. This was followed by centrifugation for 10 minutes at 10,000 rpm at 4˚C. The RNA pellet was washed with ice cold 100% ethanol, and resuspended in 50 µL RNase-free water. Takara PrimeScript qPCR kit (Cat. RR014A) was used for gene expression of endogenous genes in knockout strains.

Table 1. *Tetrahymena* cell strain identifiers and description

Strain	**Description**
B2068	Inbred B Family Strain ; Compatible with CU428; (*Mpr*+/*Mpr*+ [II, mp-s])
CU428	Inbred B Family Strain; Compatible with B2068; (*Mpr*/*Mpr* [VII, mp-s]);
CHD3B-FZZ	Originally B2068; Chd3 protein tagged with FZZ
CHD3C-FZZ	Originally CU428; Chd3 protein tagged with FZZ
CHD7B-FZZ	Originally B2068; Chd7 protein tagged with FZZ
CHD 7C-FZZ	Originally CU428; Chd7 protein tagged with FZZ
MIZ1B-FZZ	Originally B2068; Chd7 protein tagged with FZZ
MIZ1C-FZZ	Originally CU428; Chd7 protein tagged with FZZ

Table 2. Centrifuge equipment models and specifications

Model	**Volume**	**Temperature**
Eppendorf 5424 centrifuge	0-1.5 mL (Eppendorf)	Room Temperature
Sorvall Legend Micro 21R microcentrifuge	0-1.5 mL	4^{o}C
Sorvall Legend RT	1.5mL-15 mL	4^{o}C
IEC Centra CL3	10- 50mL	Room Temperature
Beckman Coulter Avanti J-301	50-500 mL	4^{o}C

Table 3. DNA Primers used for PCR amplification, qPCR and *in vitro* transcription all listed 5' to 3'

Assay	Gene Target	Primer
ChIP-qPCR	Intragenic ARS2 (alanyl-tRNA synthetase 2)	F - CACTGCGATAACACCTCTGAA R - GAGATCTTTCACCAGCAACAAAG
ChIP-qPCR	rDNA	F - GTCACGGGTAACGGAGAATTAG R - CTGGCTCCCTGAATTAGGATTG
ChIP-qPCR	CIT2 (Citrate Synthase)	F - GGAAATGTCCTCCGTGTCTAC R - CTAATACACCCAAGTCACCATCA
ChIP-qPCR	HMGB1 (TF)	F - TGCTAGCCAATACGCAACTC R - GGGCACTCAGTTCTTCTCTTG
ChIP-qPCR	HSP70 (chaperone)	F - CTTTCGATGTCTCCCTCCTTAC R - CTTCACCACCCAAGTGAGTATC
ChIP-qPCR	BTU2 (cytoskeletal)	F - CTTGGTCTCTGCTGCTATGT R - CGAGGGAAGGGAATCAAGTT
ChIP-qPCR	Twi1 (*Tetrahymena* Piwi protein)	F - CAAGCACCGCAATGAAAGATAA R - CCATAACCCAAGGAACACCA
ChIP-qPCR	RPS22 (Ribosomal Protein)	F - GTTTCCAAGGTCGTCCTCAA R - TCCACTTTTCGAAGTCAGCA
KO Cloning	CHD3 5'UTR	F - GGGCCATGTAGCGGTATATCAGCATTTTTT R – CTCGAGGATTCTTCTTCTTCTTCTTGCATG
KO Cloning	CHD3 3'UTR	F- CCGCGGGGATTAGTACACAAAAACAAAAGACA R – GAGCTCTCATTCTTCCTTCTTTCCTTCATC
KO Cloning	CHD7 5'UTR	F- GCGCCCATGTCAGCATCTGAGGATTCAATT R- CTCGAGCTTACACCGTCGTGAAGCTTT
KO Cloning	CHD7 3'UTR	F-CCGCGGGAAGAGAAGGTAGAATGACTCGTTCT R- GAGCTCTCAAACCTTGTTTATTTTTTAGCCA
KO Cloning	MIZ1 5'UTR	F- CGGGGTACCACTGGACCTCCATAATAAATACA R- CCGCTCGAGCAGACATTCCTTAACCTATTT
KO Cloning	MIZ1 3'UTR	F- GCGGCGGCCGCAGGTATAAGTGAGTTAATAGCTT R- GCGGAGCTGATCTGCTTCAAAAAAATAAAT
KO RT-PCR	CHD3	F- TCATTCTTCCTTCTTTCCTTCATCT R- TCCACAGAACACTGCATTGTC
KO RT-PCR	CHD7	F- GTTCAGAAAGCGAATATGAGTATCTAG R- GAAGCTGATGCATTATCTCTTTCA
KO RT-PCR	MIZ1	F- GGAAAGTAAGCTGGGGAAGC R- GAAGATGCTGGATTCAAAATTG
KO Cassette	NEO4	F - GTGATTCACGATTTATGCATGATCCA R – GCTTTAATGTTAATATAATAC
Cloning GST	CBC	F- CGGGAGCTCGGATCCATCTTCTACCCTGAAATT R- CGGGAGCTCGGATCCAAGCTGTGTTACTACTGT
Cloning GST	PPP	F- CGGGAGCTCGGATCCAAGCTGTGTTACTACTGT R- CGGGTCGACTTACGGGTGATCTTCGCACAGCAT
RNAComp	Control	CACCCCAAACACCCCAAACACCCCA
RNAComp	Experimental	GGGGGGGAAGGGGGGGAAGGGGGGG

3.0 Results

3.10 *In vivo* Analyses of Chd Proteins

3.110 Bioinformatic and functional analyses indicate tChds are candidate DRBPs

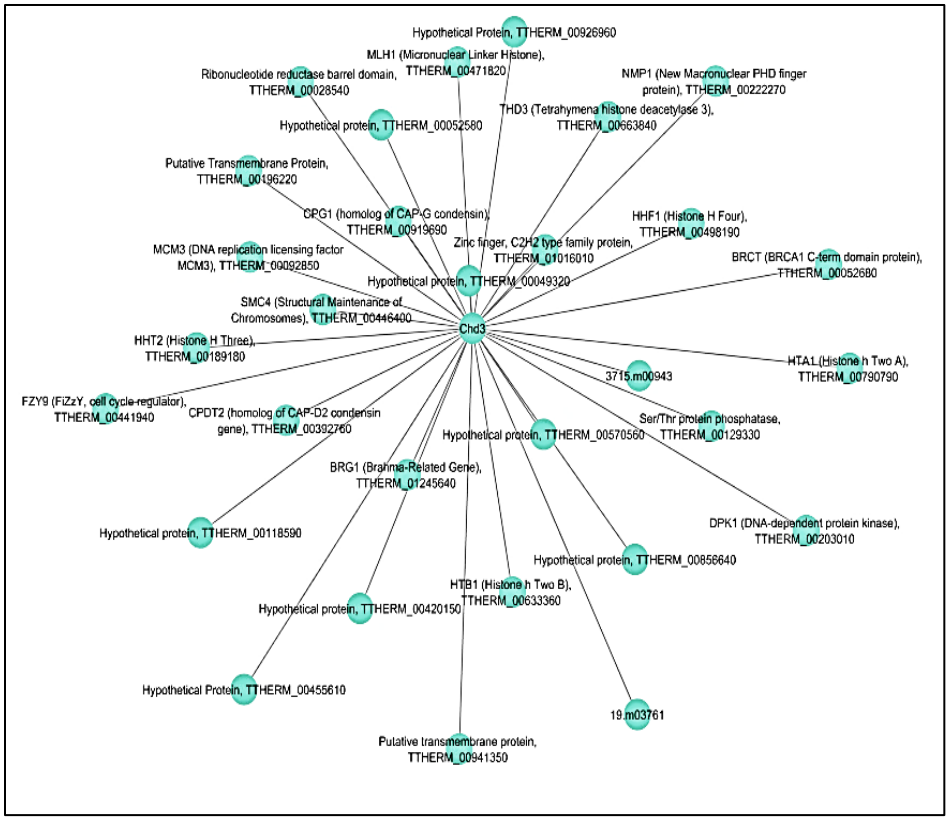
To assess *Tetrahymena* proteins as DRBPs, both bioinformatic and functional analyses were conducted. *Tetrahymena* protein sequences were compared to bona fide mammalian DRBPs previously identified by high-throughput mRNA-proteome binding studies (Castello et al. 2012). The researchers identified 860 RNA-binding proteins (RBPs) via UV-crosslinking, oligo dT-pulldown, and mass spectrometry. Many identified RNA-binding proteins were known DNA-binders, such as transcription factors and chromatin remodelers. From the exhaustive list of the mammalian mRNA-binding proteome, I identified *Tetrahymena* orthologs by BLASTp searches. This was done by annotating the NP RefSeq identifiers of strong mRNA-binding mammalian proteins. These 860 mammalian proteins were subjected to BLASTp sequence analysis using the *Tetrahymena* Genome Database (www.ciliate.org) to identify TTHERM_ protein annotations. At BLASTp E-value=e^{-100}, Chd3, Chd7 and other chromatin remodeler proteins were identified as possible RBPs (**Table 4)**. When compared to the original exhaustive list by Castello et al. (2012), the researchers scored both Chd3 and Chd7 as significant RBPs, based on a statistical algorithm they developed for predictive power.

To assess co-expression of genes with *CHD* family members, and ultimately gene function, Cytoscape analyses of *CHD3* and *CHD7* based on publically available microarray data were conducted (Miao et al., 2009). Results indicated co-expression with other conjugation-specific genes (**Figure 9**).

Table 4. *Chromatin-associated Tetrahymena orthologs of the human mRNA-binding proteome.* Castello *et al*. (2012) identified more than 800 mRNA-binding proteins in an HTP "mRNA capture" assay. The comprehensive list was subjected to BLASTp analysis to identify *Tetrahymena* orthologs. At an E-value of -100, ~350 high scoring (i.e. orthologous) *Tetrahymena* proteins were identified, and could be generally classified as post-translational modifying proteins, chaperones, axonemal proteins, translation elongation factors, chromatin-related proteins, ribosomal proteins, helicases and mRNA processing factors. The ~42 chromatin-associated proteins within the list are presented here, which include *Tetrahymena* Chd3 and Chd7.

Score	*Mammalian RBP*	*Tetrahymena Ortholog*
1217	gi|40217847|ref|NP_054733.2|	TTHERM_01034400 type III restriction enzyme, res subunit family protein
1003	gi|193211480|ref|NP_056175.3|	TTHERM_00316180 DEAD/DEAH-box helicase
1003	gi|4505939|ref|NP_000928.1|	TTHERM_00538940 DNA-directed RNA polymerase
958	gi|221316723|ref|NP_078938.2|	TTHERM_00518510 P-loop ATPase fused to an acetyltransferase, putative
883	gi|68509926|ref|NP_001349.2|	TTHERM_01044780 DEAH-box nuclear pre-mRNA splicing factor
793	gi|23510338|ref|NP_003325.2|	TTHERM_00997780 ubiquitin-activating enzyme E1
689	gi|4505941|ref|NP_000929.1|	TTHERM_00085650 DNA-directed RNA polymerase; RPC2
541	gi|20149629|ref|NP_057439.2|	TTHERM_00095530 DEAD/DEAH-box helicase
527	gi|55953087|ref|NP_036473.2|	TTHERM_00242500 nucleolar GTP-binding protein
523	gi|4507635|ref|NP_003926.1|	TTHERM_00497920 DNA topoisomerase I
517	gi|158420731|ref|NP_001005271.2|	TTHERM_00049310 SNF2 family amine-terminal protein
509	gi|4505939|ref|NP_000928.1|	TTHERM_00037340 DNA-directed RNA polymerase
498	gi|13514831|ref|NP_004389.2|	TTHERM_00219180 DEAD/DEAH-box helicase family protein
490	gi|4826690|ref|NP_004932.1|	TTHERM_00120930 DEAD/DEAH-box helicase, putative
490	gi|4507635|ref|NP_003926.1|	TTHERM_00464990 DNA topoisomerase I
485	gi|4503529|ref|NP_001407.1|	TTHERM_01197120 DEAD/DEAH-box helicase
484	gi|4826690|ref|NP_004932.1|	TTHERM_00419730 DEAH-box nuclear pre-mRNA splicing factor
473	gi|256000749|ref|NP_001157711.1|	TTHERM_00120930 DEAD/DEAH-box helicase, putative
471	gi|7661920|ref|NP_055555.1|	TTHERM_01197120 DEAD/DEAH-box helicase
468	gi|5803092|ref|NP_006829.1|	TTHERM_00624400 type II methionine aminopeptidase
461	gi|68509926|ref|NP_001349.2|	TTHERM_00120930 DEAD/DEAH-box helicase, putative
454	gi|50659095|ref|NP_004719.2|	TTHERM_00105060 DEAD/DEAH-box helicase family protein
452	gi|68509926|ref|NP_001349.2|	TTHERM_00419730 DEAH-box nuclear pre-mRNA splicing factor
449	gi|4758138|ref|NP_004387.1|	TTHERM_00190830 DEAD/DEAH-box helicase ; DRH1 (DExD/H box RNA helicase 1)
441	gi|4758138|ref|NP_004387.1|	TTHERM_00569290 DEAD/DEAH-box helicase; LIA2 (localized in macronuclear anlagen 2)
437	gi|7706254|ref|NP_057018.1|	TTHERM_00050590 nucleolar protein
431	gi|61676188|ref|NP_113584.3|	TTHERM_00161820 HECT-domain (ubiquitin-transferase) protein
430	gi|11225260|ref|NP_003277.1|	TTHERM_01299650 DNA topoisomerase I
426	gi|32483374|ref|NP_006383.2|	TTHERM_00558040 nucleolar protein 5a
426	gi|20336302|ref|NP_064547.2|	TTHERM_00120930 DEAD/DEAH-box helicase, putative
426	**gi|118421089|ref|NP_001262.3|**	**TTHERM_00193800 CHD7 (Chromo-Helicase-DNA-binding domain)**
422	**gi|118421089|ref|NP_001262.3|**	**TTHERM_00049310 CHD3 (Chromo-Helicase-DNA-binding domain)**
421	gi|171460958|ref|NP_079170.2|	TTHERM_00160670 nucleolar protein
418	gi|4507635|ref|NP_003926.1|	TTHERM_00216140 DNA topoisomerase
407	gi|51094101|ref|NP_076977.3|	TTHERM_00449630 type III restriction enzyme, res subunit family protein
407	gi|52630322|ref|NP_005843.2|	TTHERM_00193800 type III restriction enzyme, res subunit family protein
400	gi|118421089|ref|NP_001262.3|	TTHERM_01245640 HSA protein ; BRG1 (Brahma-Related Gene)
392	gi|158420731|ref|NP_001005271.2|	TTHERM_00137610 SNF2 family amine-terminal protein
390	gi|18375623|ref|NP_542165.1|	TTHERM_00469380 DEAD/DEAH-box helicase
381	gi|224593278|ref|NP_060365.7|	TTHERM_00187140 DEAD/DEAH-box helicase
375	gi|118421089|ref|NP_001262.3|	TTHERM_00137610 SNF2 family amine-terminal protein
345	gi|31542656|ref|NP_061943.2|	TTHERM_00095530 DEAD/DEAH-box helicase

A)



B)

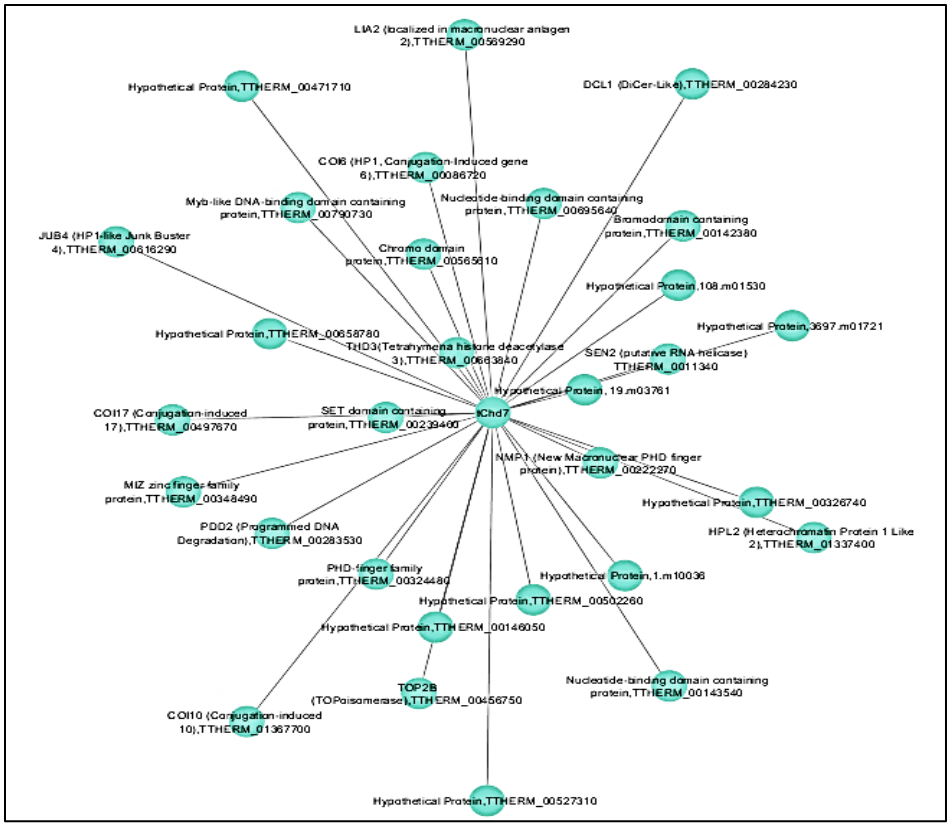
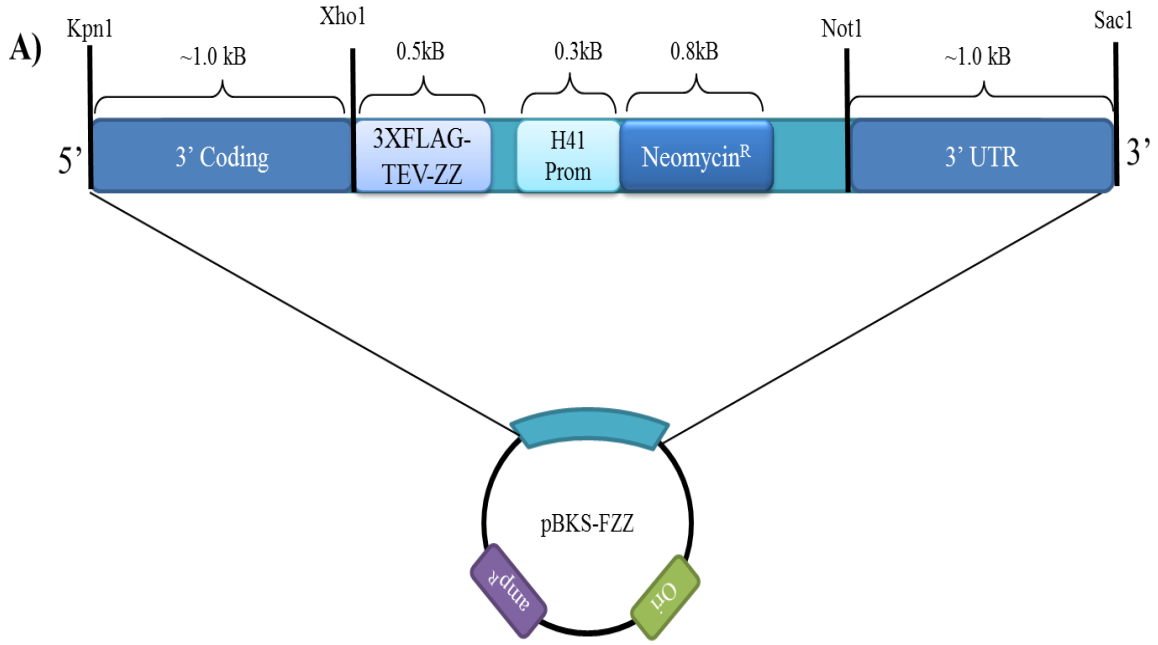


Figure 9. Cytoscape analysis of *CHD3* and *CHD7*. The top 30 highest co-expressing genes with **A)** *CHD3* and **B)** *CHD7* were subjected to unidirectional network analysis. The central node represents the associated *CHD3* or *CHD7* gene; distance between the central node and outward nodes is representative of their estimated co-expression relatedness. As expected, networks show developmentally regulated, co-expressed genes. Most genes, roughly 50% for both *CHD3* and *CHD7*, are involved in chromatin dynamics. Networks were especially enriched for both epigenetic-binding proteins (e.g. HP1 proteins and SET domain proteins, conjugation-specific histone proteins) and genes involved in RNA-directed DNA elimination (e.g. *PDD2*, *DCL1*, and *LIA2).* TTHERM_ numbers were annotated with their corresponding *Tetrahymena* gene

Almost half of all co-expressors for both *CHD3* (14/30) and *CHD7* (15/30), were other chromatin-associated and chromatin remodeling proteins with epigenetic-binding domains (e.g. HP1 proteins, SET domain proteins, conjugation-specific histone proteins). Further, *CHD7* co-expressed with other genes directly involved in the production and function of scan RNAs such as *PDD2*, the primary chromodomain protein encoding gene which binds RNAs in a large multi-protein complex, *DCL1*, the gene encoding dicer protein which processes scanRNA, and *LIA2*, a recently characterized DEAD/DEAH box helicase involved in RNA production, IES excision and chromatin remodeling.

Initial functional analysis of Chd3 and Chd7, i.e. *in vivo* localization and expression patterns, were contingent on the generation of *Tetrahymena* strains endogenously tagged at the C-terminus of the gene locus of interest. This was conducted by biolistic bombardment transformation of *Tetrahymena* using the pBKS vector. This plasmid contains a 3xFLAG-TEV-ProteinAProteinA (FZZ) tag (**Figure 10A)**. Including the 18 kDa tag, Chd3-FZZ and Chd7-FZZ are 203 kDa and 274 kDa, respectively. The protein sizes of FZZ-tagged strains for Chd3 and Chd7 were confirmed by anti-IgG Western blot analysis of whole cell extracts (**Figure 10B**). Tagged proteins in strains of the two mating types used resolved to the correct size, and were used for all downstream *in vivo* procedures.

To optimize immunoprecipitation assays for high protein expression and protein yield, a developmental Western blot was done (**Figure 11**). The overall trend of protein expression data for both Chd3 and Chd7 correlates to the publically available *Tetrahymena* Genome Database (TGD) microarray and RNAseq data. These proteins are developmentally regulated. Both



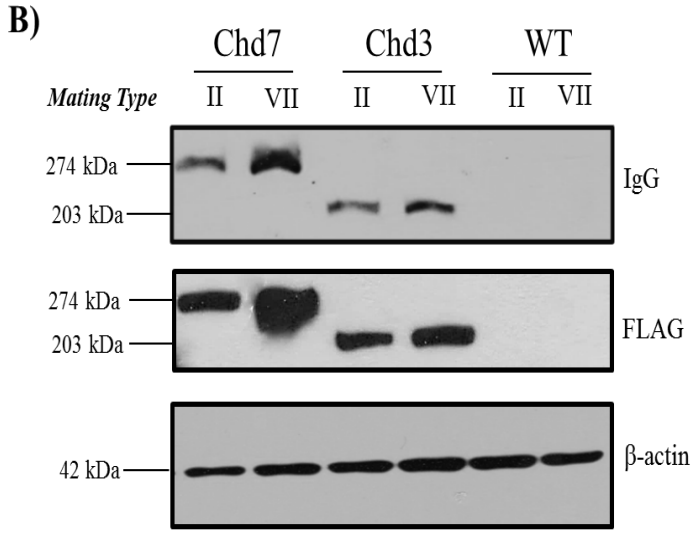
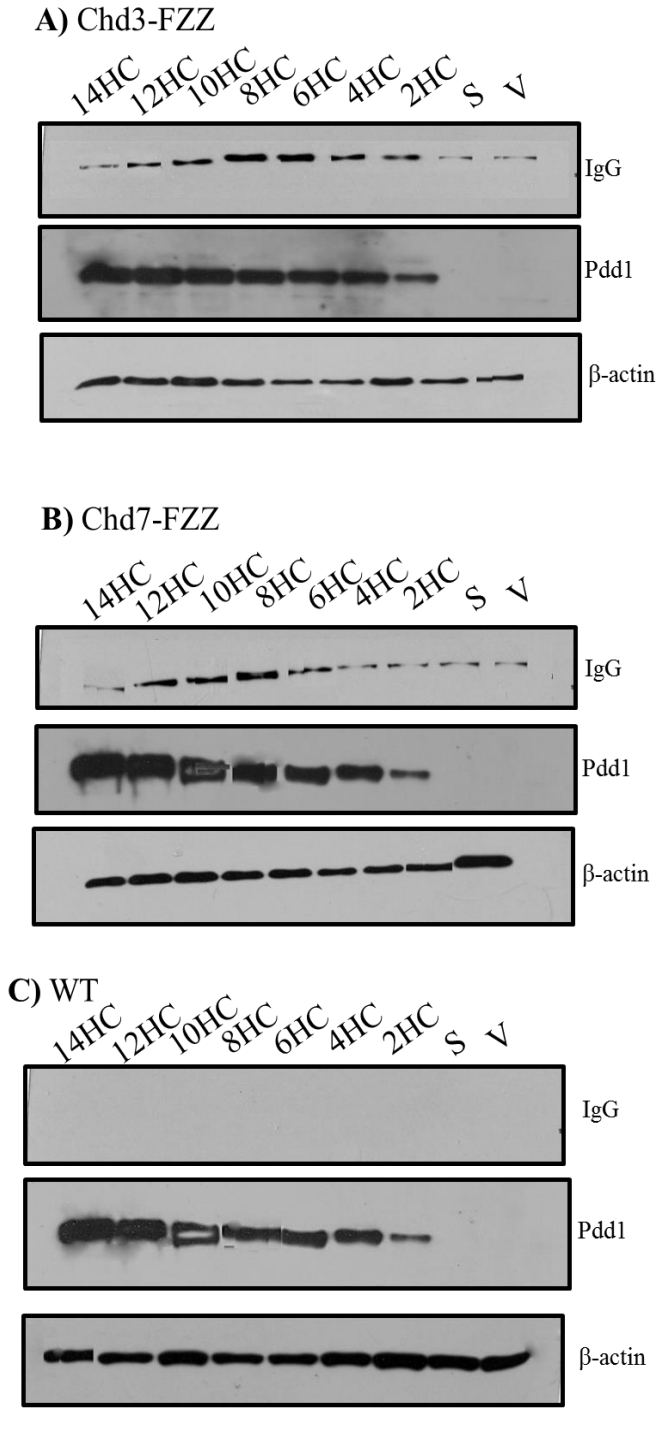


Figure 10. *Confirmation of successful FZZ-tagged Chd3 and Chd7 constructs by Western blot of whole cell extracts.* **A)** *Tetrahymena* genes *CHD3* and *CHD7* were endogenously and C-terminally tagged at the loci of interest by cloning into the pBKS-FZZ vector, and Biolostic Gene Bombardment of wildtype *Tetrtaymena*. **B)** Successful transformants in two different mating types (B2086, MTII and CU428, MTVII) were confirmed with Western blot analysis of whole cell extracts, using anti-IgG which binds the Protein A moiety of FZZ and anti-FLAG which binds the 3xFLAG moiety of FZZ antibodies. Chd7-FZZ is roughly 274 kDa, while Chd3-FZZ is ~203 kDa. Anti—β actin was used as a loading control.



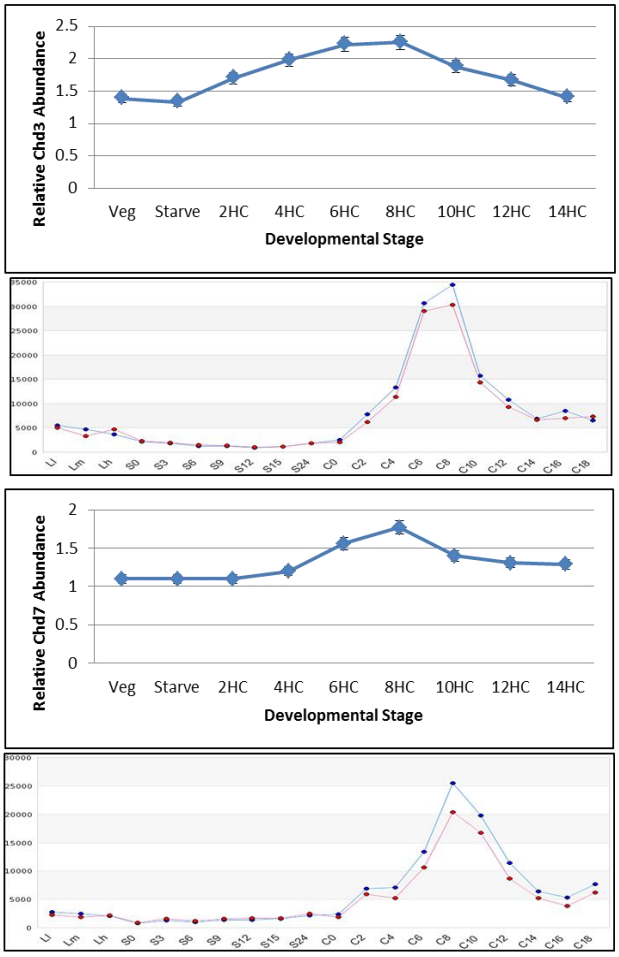


Figure 11. *Developmental Western blot of whole cell extracts.* Protein expression levels of **A)** Chd3-FZZ, **B)** Chd7-FZZ, and **C)** WT cells were analyzed by collection of *Tetrahymena* during vegetative growth, starvation and conjugation. Conjugating cell samples were collected every 2 hours, for a total of 14 hours. Anti-IgG, a marker of Protein A within the ZZ strains, anti-Pdd1, a marker of conjugation), and anti- β-actin, a positive loading control, were used. Peak expression of Chd3 and Chd7 are 7 and 8 hours respectively. ImageJ semi-quantification of each Western blot is provided on the top right with protein normalized to β-actin, alongside publically available microarray expression data below, www.ciliate.org. Strains had similar protein expression patterns as compared to microarray data. The +/- duplicate SEM values are provided at every time point.

chromatin remodelers are minimally expressed during vegetative growth and starvation. There is an increase in protein expression immediately after cell pairing, starting at 2 hours of conjugation. However, unlike the steady increase observed in expression data for Pdd1, a key developmental regulator of *Tetrahymena* conjugation, Chd3 and Chd7 are expressed during vegetative growth and starvation, and they exhibit a bell curve-like pattern. Chd3 had a peak protein expression between 6 and 8 hours of conjugation, while Chd7 had peak protein expression at approximately 8 hours of conjugation. ImageJ freeware was used to provide semi-quantitative data analysis of protein expression patterns for each stage of development. The ratio of peak protein expression during conjugation, over the vegetative protein expression for Chd3 and Chd7 indicated a fold increase of 2.12+/-0.05 and 1.77 +/-0.06, respectively.

Figure 12 illustrates the protein localization patterns of Chd3 and Chd7 throughout the life cycle. These chromatin remodeling proteins exclusively localized to the macronuclei. From vegetative growth up until 14 hours of conjugation, there was no expression of these proteins within micronuclei at any point. During vegetative growth and starvation, Chd3 and Chd7 localize to parental macronuclei. During development, Chd3 and Chd7 localization remains exclusively in parental MACs from 2 hours to 6 hours of conjugation. Between 6-8 hours of conjugation, protein expression of Chd3 and Chd7 was lost in the old MAC. At this time point, there was localization in a discrete, switch-like manner to the developing MAC. These developing macronuclei are known as Anlagen. Localization remained within Anlagen for the remainder of cell pairing measured up to 14 hours, even when the old MAC was not yet degraded. At no point during the sexual lifecycle was protein localization of Chd3 or Chd7 observed in the micronuclei, cytosol or membrane structures, indicating an expression pattern known as zygotic macronuclear expression. These expression and localization patterns of Chds

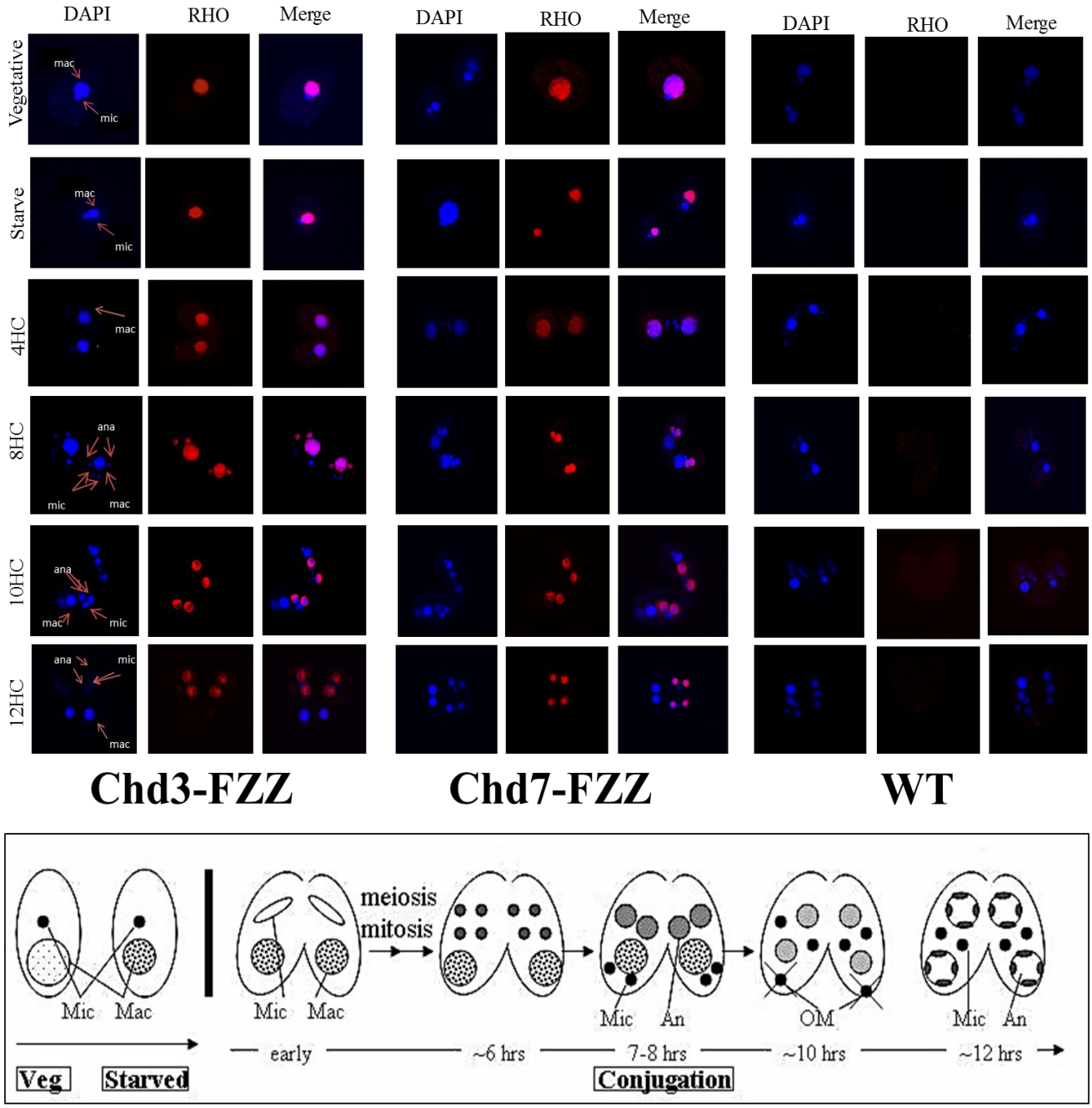


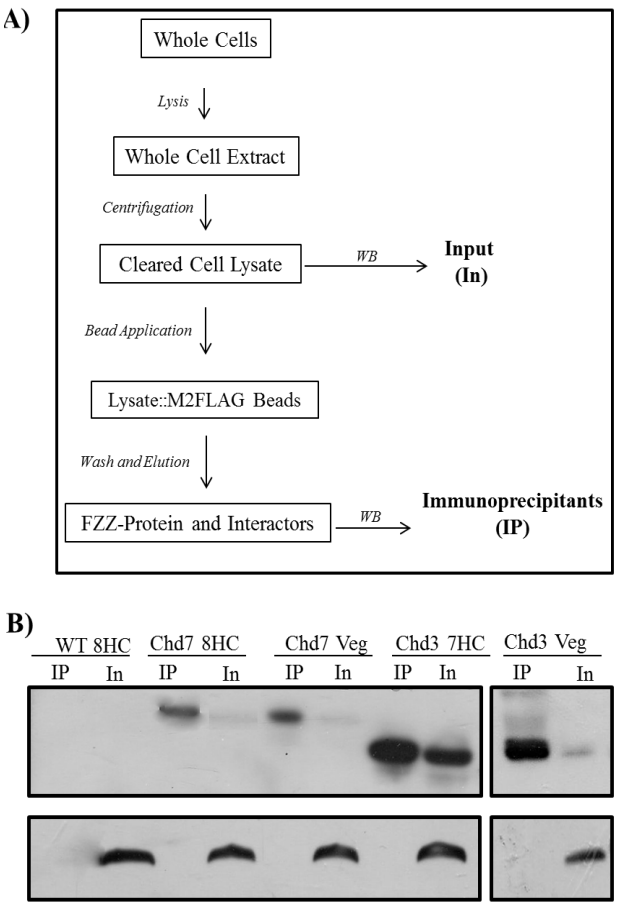
Figure 12. *Indirect immunofluorescence analysis of Chd3-FZZ , Chd7-FZZ and untagged WT cells, throughout Tetrahymena development.* Chd3-FZZ and Chd7-FZZ have similar protein localization patterns. Nuclear localization of Chds occurs exclusively in the MAC throughout the lifecycle. This includes localization of Chd3 and Chd7 to Analgen. From 6 to 8 hours of conjugation, there is a distinct switch in localization pattern from the parental MAC, towards the developing MAC. Immunofluorescence was done in duplicate, along with wildtype cells as a negative control for rhodamine fluorescence. An image adopted from the Ciliate Genomics Consortium (http://faculty.jsd.claremont.edu/ewiley/) is provided below, to illustrate temporal and spatial nuclear differentiation during *Tetrahymena* conjugation.

(i.e. peak in protein expression and localization to Anlagen during 6-8 hours of conjugation) are temporally and spatially correlated to RNA-guided DNA elimination events in *T. thermophila.*

3.120 Affinity Purification and MS indicates Chds bind chromatin-related proteins, and confirms a novel Chd3:Miz1 interaction

The Chd3-FZZ and Chd7-FZZ prey proteins were immunopurified from *Tetrahymena* strains by affinity purification (**Figure 13A**). Successful immunoprecipitation from both vegetative and conjugating FZZ-tagged cell strains were confirmed by Western blot (**Figure 13B**). Raw AP–MS interaction data were analyzed using SAINT (Significance Analysis of INTeractome) (Choi et al., 2011). This statistical tool uses unfiltered, quantitative spectral counts and assigns a confidence value for each interacting protein pulled down by affinity purification. The SAINT algorithm estimates a probability for each interactor, or prey protein. This numerical estimate is entirely contingent on data from control trials (AP-MS on untagged cells), the number of experimental biological replicates (minimum of two), and the presence of other AP-MS data from FZZ-tagged proteins. In sum, SAINT is a tool that enables the discrimination between true and false interactions (Avg P>0.8). AP-MS and SAINT data for Chd3 and Chd7 revealed a number of interesting interactions.

Affinity purification and mass spectrometry on Chd7 during vegetative growth identified the Chd7 bait, and Krs1 (lysyl-tRNA synthetase) as a potential, minimally-enriched interactor (see **Appendix A).** During conjugation, AP-MS and SAINT analysis of Chd7 identified histone H2B, microtubule subunits, and ribosomal proteins as interactors. Non-SAINT validated proteins included Hhf2 (histone H4) (79 hits), Tgp1 (*Tetrahymena* G-DNA binding protein) (84 hits), H3 (Histone H3) (112 hits), Hta3 (Histone H2A) (39 hits), transcription factor S-II,



C)

Chd3 Veg

TTHERM Number	SAINT	Protein Annotation
TTHERM_00049310	268	CHD3 **BAIT**
TTHERM_00442270	277	MIZ zinc finger **PREY**

Chd3 7HC

TTHERM Number	SAINT	Protein Annotation
TTHERM_00049310	297	CHD3 **BAIT**
TTHERM_00442270	256	MIZ zinc **PREY**

Chd7 Veg

TTHERM Number	SAINT	Protein Annotation
TTHERM_00193800	1691	CHD7 **BAIT**

Chd7 8HC

TTHERM Number	SAINT	Protein Annotation
TTHERM_00193800	162	CHD7 **BAIT**

Figure 13. *Protein immunoprecipitation of Chd3-FZZ and Chd7-FZZ and mass spectrometry.* **A)** Affinity purification of FZZ-tagged strains involved cell lysis, soluble protein extraction, application of lysates to anti-FLAG bead, bead washes, and final protein elution under denaturing, basic conditions. Throughout the procedure, small aliquots were harvested after centrifugation (soluble input proteins) and after purified protein elution (immunoprecipitated proteins). **B)** These input and immunoprecipitated samples were electrophoresed through 10% SDS-PAGE, and probed with anti-IgG, to confirm successful purification for mass spectrometry analysis. **C)** Mass spectrometry and statistical analysis of protein interactors indicated Chd3 (the bait protein) binds a previously uncharacterized MIZ-SP/RING protein (a novel prey protein) during vegetative growth and conjugation. While mass spectrometry of Chd7 was successful in that bait was recovered and enriched, subsequent statistical analyses did not identify significant protein interactors of functional interest.

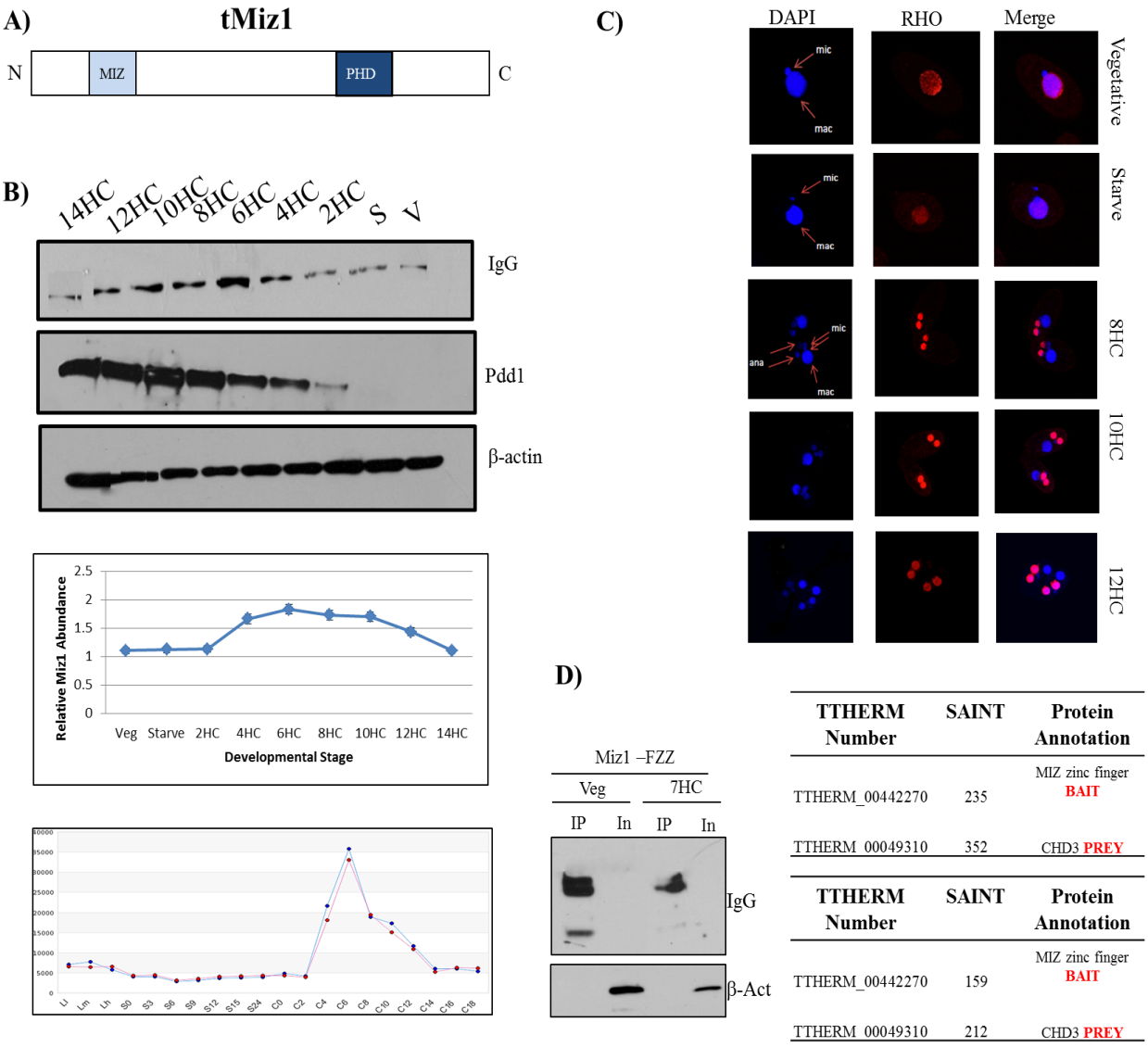
TFIIS (8 hits), and Translation elongation factor EF-1 (5 hits). Coupled with AP-MS analyses, affinity purified Chd3 and Chd7 during conjugation were electrophoresed through SDS-PAGE and silver stained (**Appendix B**). None of the Chd7 interactors were of functional interest, and thus not pursued in this study.

Chd3-FZZ bound a novel and previously uncharacterized MIZ/SP-RING finger protein, here called Miz1. This binding interaction occurs both during vegetative growth and conjugation (**Figure 13C**). Other SAINT-validated Chd3 interactors during conjugation include proteins largely involved in protein metabolism (e.g. ribosomal proteins, proteases, tRNA synthatases) and enzymes involved in cellular respiration (**Appendix C**). Proteins that were pulled down with

higher spectral count values than wildtype controls, but failed SAINT analysis stringency measures include: Hsp82 (34 hits), Eef2 (Eukaryotic Elongation Factor) (21 hits), Hsp70 (15 hits), Vma1/2 Vacuolar Membrane ATPase (14 and 21 hits), ADP/ATP Carrier Acc1 (25 hits), and translation elongation factor EF-1 (23 hits).

Miz1 has a MIZ-SP/RING domain near the N-terminus, as well as a central zinc-containing PHD finger (**Figure 14A**). To validate the Chd3:Miz1 interaction, Miz1 was endogenously tagged in the same manner as Chd3 and Chd7. The size of Miz1-FZZ is ~163 kDa, however on SDS-PAGE the protein resolves to a higher molecular weight (~220kDa). Expression patterns indicated Miz1 peaked in expression at around 8 hours of conjugation, similar to Chd7, and slightly after the peak expression time period of Chd3 (**Figure 14B**). Semi-quantification of Miz1 indicates a fold increase at its peak during conjugation over vegetative growth, at ~1.83 +/- 0.07. The *in vivo* localization pattern of Miz1 mirrors that of the Chds, where localization is exclusively observed in the macronuclei, and there is a distinct switch from localization from the parental macronuclei to zygotic micronuclei at 6-8 hours of conjugation (**Figure 14C**).

Mass spectrometry of reciprocally-tagged Miz1-FZZ identified Chd3 as a strong *in vivo* prey. Both bait (Miz1) and prey (Chd3) were abundantly recovered with a Miz1:Chd3 ratio of 235:352 and 159:212 SAINT hits during vegetative growth and conjugation, respectively (**Figure 14D**). A complete list of SAINT validated Miz1 interactors are provided in **Appendices D and E.** In these data, ribosomal proteins were also observed to bind Miz1 in SAINT-validated AP-MS.



TTHERM Number	SAINT	Protein Annotation
TTHERM_00442270	235	MIZ zinc finger **BAIT**
TTHERM_00049310	352	CHD3 **PREY**

TTHERM Number	SAINT	Protein Annotation
TTHERM_00442270	159	MIZ zinc finger **BAIT**
TTHERM_00049310	212	CHD3 **PREY**

Figure 14. *Reciprocal tagged and functional analysis of Miz1-FZZ.* **A)** Miz1-FZZ is a previously uncharacterized MIZ-SP/RING protein, with a MIZ RING domain in the N-terminus and PHD finger in the C-terminus. **B)** Developmental expression patterns of Miz1 on the protein level indicates a peak in expression at around 6 hours of conjugation (top and middle), which roughly correlates to gene microarray data (bottom). **C)** Indirect immunofluorescence of Miz1 indicates it has an identical protein localization pattern as it's binding partner Chd3, as well as Chd7. The zinc finger is expressed throughout development, however it is exclusive to the MAC. During the development of the new MAC, Miz1 localizes to Analgen and is lost in parental macronuclei. **D)** Successful affinity purification of vegetative and 7HC Miz1-FZZ was confirmed by anti-IgG Western blot of input and immunoprecipitated protein samples. Mass spectrometry data for vegetative and conjugating Miz1-FZZ strains from two replicates was filtered by SAINT, an algorithm which converts protein spectral hits to a numerical probabilistic score. Chd3 was the most abundant and specific interactor, confirming the presence of a Miz1:Chd3 complex throughout development.

3.130 Miz1 is a hypothetical SUMO E3 ligase, which may mediate temporally-regulated SUMOylation events during conjugation

SUMO, or Small Ubiquitin-Like Modifier, is a small protein that is able to be post-translationally conjugated to target proteins in eukaryotes. Similar to ubiquitination, SUMOylation of protein substrates involves a highly organized series of steps including an E1 activating enzyme, E2 conjugating enzyme and E3 ligating enzyme. Unlike ubiquitination however, the addition of SUMO to proteins does not exclusively signal protein degradation. Rather SUMOylation is involved in diverse cellular processes. Miz1 is a RING zinc finger protein. Eukaryotic orthologs based on protein sequence similarity identified through www.ciliate.org include the E3-SUMO ligases Gei17 (*Caenorhabditis elegans*), SIZ1 (*S. cerevisiae*), and PisA (*Dictyostelium discoideum).* This protein is therefore a candidate Class I member of the Siz1/PIAS family of E3 SUMO ligases (**Figure 15A**). SUMOylation of substrates usually occurs on the canonical consensus motif **"ψ-K-X-E"** (where ψ is a hydrophobic amino acid and X is any amino acid). Additionally, SUMOs can non-covalently bind other proteins through targeting specific SIMs. Prediction of SUMOylation sites and SUMO-interaction Motifs (SIMs) on the Miz1 amino acid sequence was completed using GPS-SUMO freeware (**Figure 15B**). GPS-SUMO computational analysis of Miz1 indicated the protein has both SUMO interaction and SUMOylation motifs. GPS-SUMO computational analysis of tChd3, the validated Miz1 protein interactor, also indicated the presence of both SUMO interaction and SUMOylation motifs (data not shown).

To test whether Miz1 has SUMO-binding affinity as a potential E3-SUMO ligase, an anti-SUMO polyclonal antibody gifted by the Forney Lab (Purdue University) was used. The antibody was made from amino acids 21-43 (FFKIKKTTQFKKLMDAYC) of the *Tetrahymena* SUMO

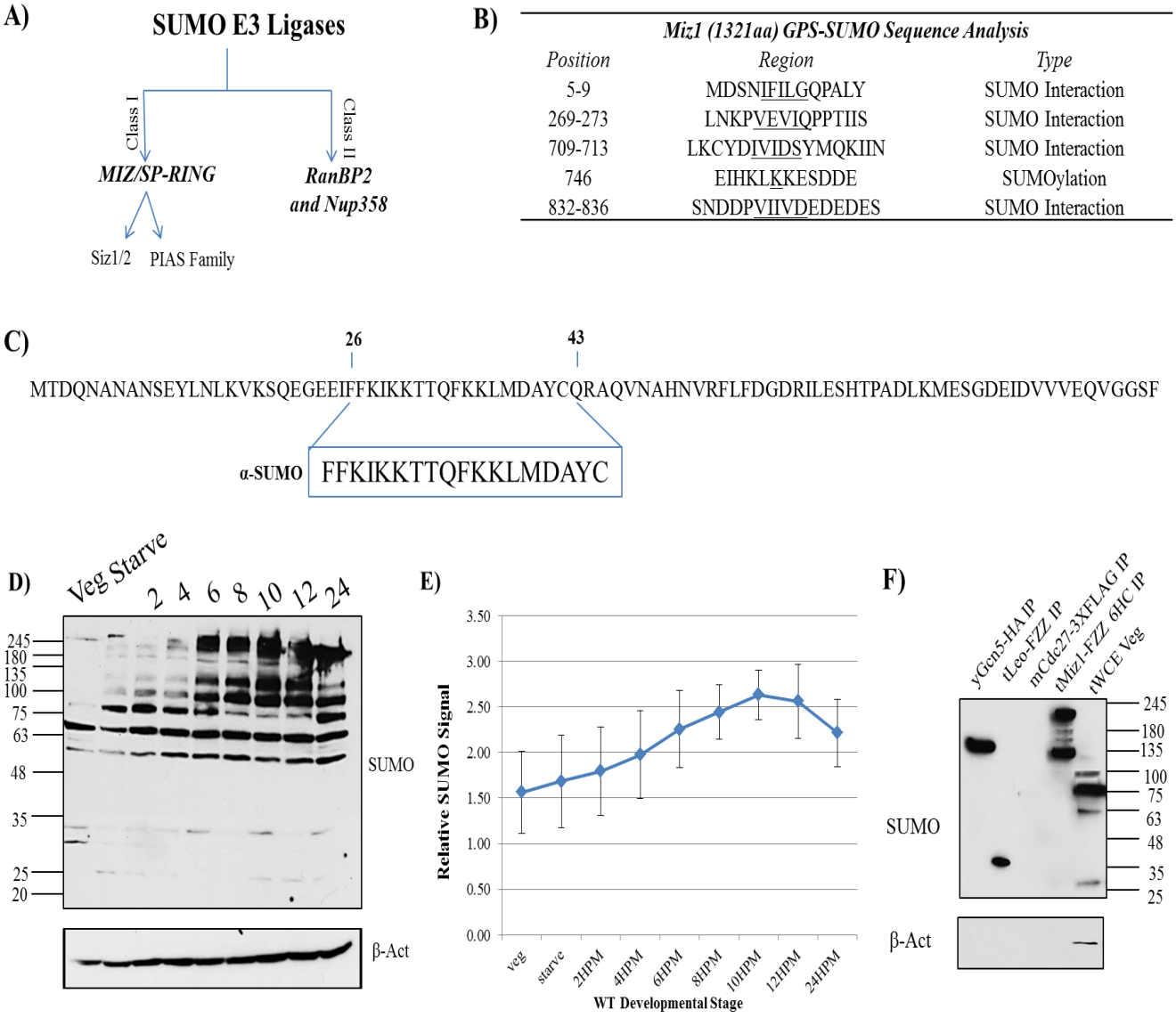


Figure 15. *tMiz1 may function as an E3 SUMO ligase in developmentally regulated Tetrahymena SUMOylation.* **A)** There are two classes of SUMO E3 Ligases. Miz1 is a Class I member, of the MIZ/SP-RING family. **B)** SUMO ligases frequently have SUMO interaction motifs (SIMs). GPS SUMO sequence analysis of the 1321 amino acid Miz1 sequence identifies four potential SUMO Interaction Motifs and one potential SUMOylation site. **C)** *SMT3* encodes the post-translational modification SUMO (Small ubiquitin-related modifier). Collaborators at the Forney Lab (Purdue University) developed a polyclonal *Tetrahymena* SUMO-antibody, made from amino acids 26-43 of the 90 aa full-length tSmt3 protein . **D)** Developmental Western blot analysis of wildtype *Tetrahymena* with anti-SUMO indicates SUMOylation is developmentally regulated. **E)** Semi-quantification of the developmental Western blot shows a peak in SUMO protein expression between 8-12h of conjugation, n=3. **F)** The *Tetrahymena* SUMO antibody is non-specific to the IgG moiety of the Miz1-FZZ tag, however it does not bind HA or 3xFLAG tags.

protein, tSmt3 (**Figure 15C**). The N-terminal region of Smt3 shows divergence from other eukaryotes, however most of the protein sequence is largely conserved including a diglycine motif present on the mature C-terminus of SUMO proteins. Total SUMOylation in *Tetrahymena* peaks between 8-12 hours of development, as demonstrated in **Figures 15D and E.** Expression at ~8hrs indicated co-expression of SUMO with Miz1 and Chd3.

Full length Miz1-FZZ is 163 kDa. However, the approximate size of Miz1-FZZ on Western blots is ~220 kDa, making it ~60kDa larger than expected. Since a single SUMOylation unit can increase the size of a protein by ~15 kDa, this size difference may be due to a multiply SUMOylated form of Miz1, specifically, the post-translational addition of four SUMO proteins. To directly test potential Miz1:SUMO interactions, affinity purification-Western blot (AP-WB) with the anti-SUMO antibody was conducted. Initial AP-WB trials on Miz1-FZZ suggested the protein bound tSmt3 strongly *in vivo*. However, it was identified that the antibody binds non-specifically to Miz1, via it's affinity to the ZZ portion of the FZZ tag (see **Figure 15F**). The antibody does not bind FLAG or HA tags, but binds all FZZ-tagged proteins. Therefore, a tandem two-step immunoprecipitation of Miz1-FZZ resulting in TEV cleavage of the IgG portion was conducted. **Figure 16** shows that in vegetative and conjugating cells, purified Miz1-FLAG binds tSmt3 minimally. This was not observed in starved cells, even though developmental Western blot analysis indicated SUMOylation also occurs during starvation in *Tetrahymena*.

Note that developmental Western blot analysis using anti-SUMO on *MIZ1* and *CHD3* knockout cells (knockout constructs will be discussed at length in section 3.30) was also conducted. However, no difference in SUMO expression was observed when compared to wild type (data

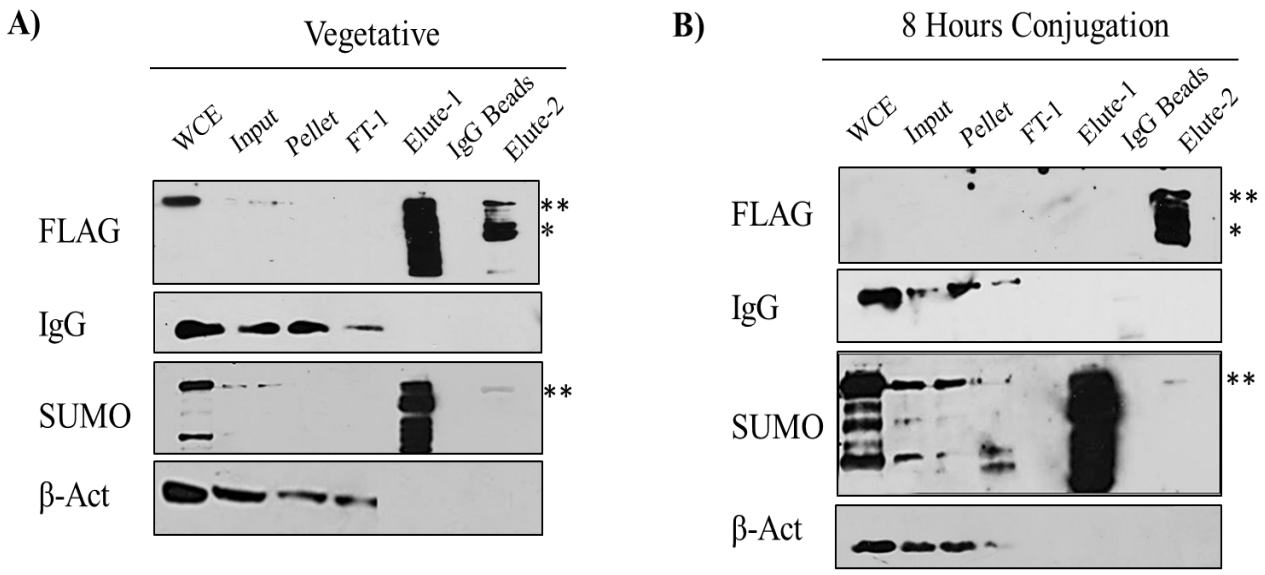


Figure 16. *SUMOylation of Miz1.* The predicted E3 SUMO-ligase Miz1 was immunopurified by a two-step immunoprecipitation process, which ensures the IgG moiety of the FZZ tag is cleaved (anti-SUMO non-specifically binds IgG).Whole cell extracts (WCEs), soluble input nuclear fractions (Input), insoluble fractions (Pellet), flow through from first step IgG bead purification (FT-1), elute from first step IgG bead purification (Elute-1), post TEV-cleaved IgG beads (IgG Beads) and final FLAG-taggd protein elute (Elute 2) were electrophoresed through 10% SDS-PAGE. During **A)** vegetative growth **B)** in conjugation a weak anti-SUMO signal was visible. No anti-SUMO signal was observed during starvation (data not shown). Full length Miz1-FZZ is ~170 kDa (*). However, the approximate size of immunoprecipitated Miz1-FZZ is ~220 kDa, possibly due to a SUMOylated form (**). Upon cleavage of the ZZ moiety of the FZZ tag, the protein should be 8 kDa smaller from either the unSUMOylated or SUMOylated forms.

not shown). Additionally, tandem two-step purification and anti-SUMO probing of Chd3-FZZ was also conducted, but no signal was observed (data not shown).

3.140 RNA immunoprecipitation and Urea-PAGE analysis suggest Chds bind long RNAs and other distinct RNA species

Protein immunoprecipitation was optimized for RNA extraction of a known *Tetrahymena* RNA-binding protein, to allow intact protein purification and high-RNA yield. Optimization was conducted using a ZZ-tagged construct of the essential *Tetrahymena* Piwi protein, Twi12 (kindly gifted by the Collins Lab at UC Berkeley) (**Figure 17**). Twi12 binds tRNA fragments derived

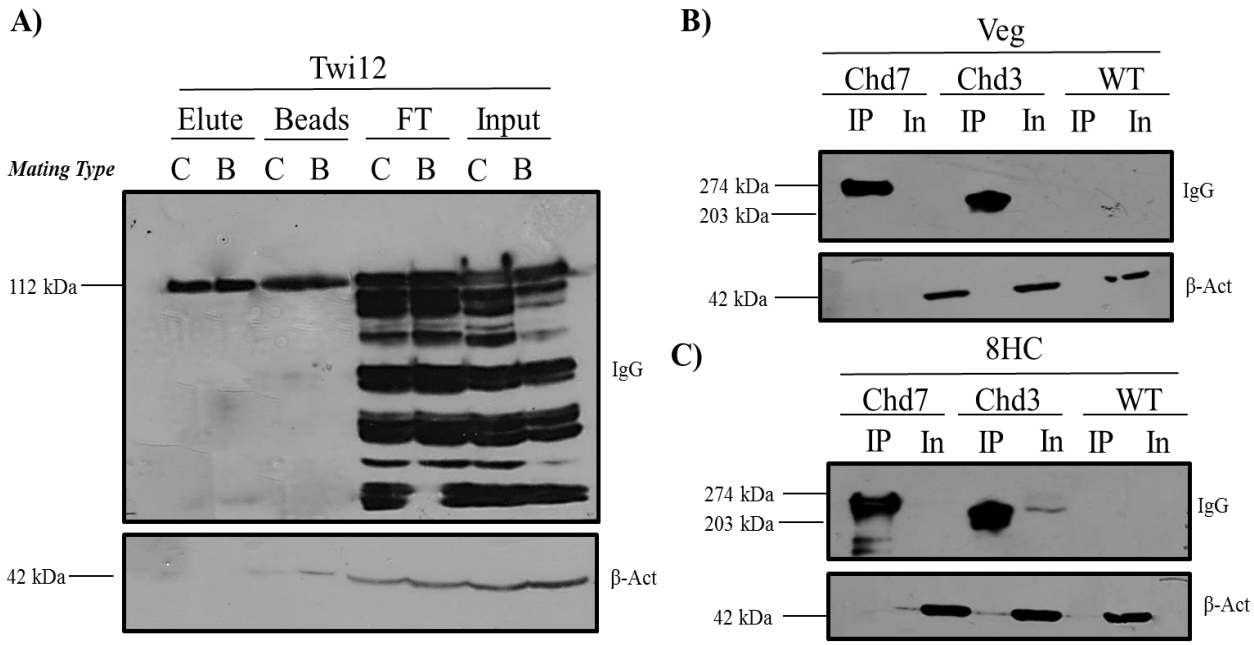


Figure 17. *Optimization of RIP on Chd3 and Chd7.* **A)** Protein immunoprecipitation was optimized for RNA purification using ZZ-Twi12, a vegetative tRNA binding protein, gifted by Kathleen Collins (UC Berkeley). ZZ-Twi12 was purified using IgG conjugated beads, and samples were collected throughout the purification process. Intact proteins were eluted with anti-FLAG peptide. Chd3-FZZ and Chd7-FZZ were successfully purified with the same Twi12 protocol, during **B)** vegetative growth and **C)** conjugation.

from the 3' ends of mature tRNAs during vegetative growth, and are observed as RNAs smaller than 70nts (i.e. the size of *Tetrahymena* tRNAs). Following successful purification of CHDs and TriZOL-elution of bound RNA species, RNA immunoprecipitants electrophoresed though urea-PAGE suggested Chd3 and Chd7 bind diverse RNA species *in vivo* (**Figure 18A**). Chds exhibited similar RNA band patterning throughout the life cycle. Long RNA species (i.e. above 300nt) predominated Chd3 and Chd7 binding. During conjugation, both Chd3 and Chd7 bound an RNA doublet at around 80nts. Unique to vegetative growth was an RNA doublet between 80nt-150nt, and a single RNA species at ~80nt. No RNAs below 50nt were observed to bind Chds in vegetative growth or conjugation. RNase treatment was used to confirm RNA immunoprecipitants observed by urea-PAGE analysis were not contaminated with DNA (**Figure 18B**).

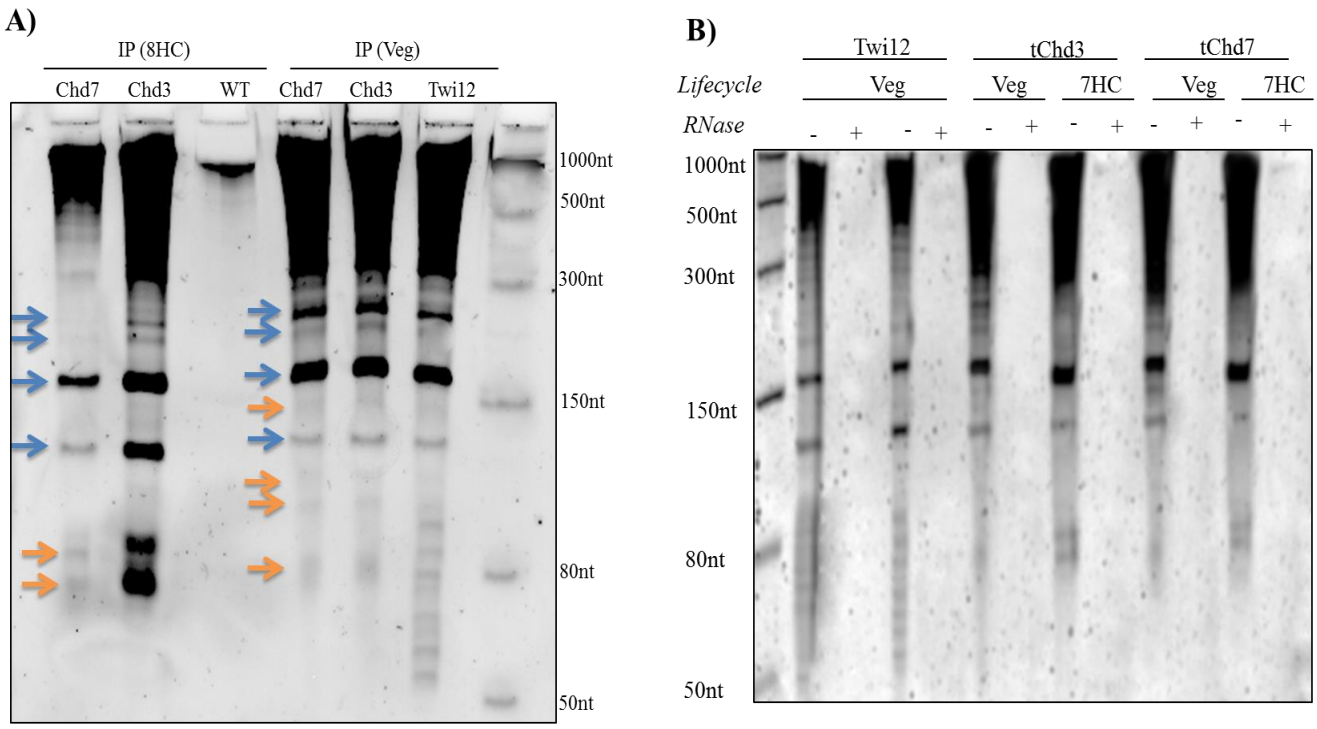


Figure 18. *RNA Immunoprecipitation of Vegetative and Conjugating Chd3 and Chd7.* RNA was immunopurified from CHD3/7-FZZ complexes, and electrophoresed through 12.5% Urea-PAGE alongside a LMW ssRNA ladder. Sybr Gold was used for staining. Twi12, a tRNA fragment binding protein, was used as a positive control, and untagged wildtype cells were used as a negative control. **A)** In both vegetative growth (V) and 8 hours of conjugation (7HC), Chds bind multiple species. Orange arrows indicate RNA species unique to the Chds, and blue arrow indicate RNA species also pulled down in Twi12. Chds throughout the life cycle exhibited similar RNA patterning, except for two unique RNA species in vegetative growth, at ~100nt. **B)** RNase treatment of RNA immunoprecipitants was used to confirm the nucleic acids pulled down upon TriZOL extraction did not result from DNA contamination.

3.150 Chromatin Immunoprecipitation of Chds and Miz1 indicate the proteins bind constitutively expressed genes during vegetative growth, and Chd3:Miz1 might be a repressive regulator of the conjugation-specific PDD1 gene

In the last component of IP3, ChIP was optimized for vegetative *T. thermophila* cells, in the analysis of Chd3, Chd7 and Miz1 chromatin-binding affinities. Bd-FZZ is a *Tetrahymena* bromodomain protein that has undergone ChIP NGS and qPCR analysis with collaborators at the Fillingham Lab, Ryerson University. This cell strain was used as a positive control for all ChIP experiments, while wildtype cells were used as a negative control. Sonication of whole cell

extracts from each cell strain was conducted, to generate 500bp-1kb fragments of genomic DNA, the optimal chromatin shearing size for quantitative PCR analysis (**Figure 19A**). Intact chromatin integrity was confirmed by PCR amplification of input DNA samples, using functionally and spatially diverse gene elements present in the *Tetrahymena* genome*:* ATU (alpha tubulin), rDNA, M-element (IES) and R-element (IES) **(Figures 19B and 11C)**. Successful amplification of target genes in input chromatin samples indicated high quality chromatin was present post-sonication. Protein integrity following sonication was confirmed by Western blot analysis. Subsequent immunoprecipitation of the proteins of interest was conducted, and proteins successfully purified were also analyzed by Western blot (**Figure 19D and 11E)**.

Recovered protein immunoprecipitants were used for downstream qPCR analysis. Gene targets for qPCR of ChIP samples were selected based on literature identifying Chd3, Chd7 or Chd1 (a close tChd7 ortholog) genomic binding sites in yeast and mammalian cells. For Chd1, primers for the intragenic region of the alanyl tRNA synthetase, or *Tetrahymena ARS2* gene were used. Human Chd1 is known to bind intergenic regions of active tRNA synthetase genes including alanyl tRNA synthetase in an RNA pol II-dependent manner (Siggens et al., 2015). The gene target for Chd7 was the coding region of the *Tetrahymena* rDNA locus. Mammalian Chd7 is known to be an essential regulator of ribosomal RNA biogenesis, and has the capacity to bind both coding and non-coding regions of rDNA, specifically at hypomethylated, active rDNA (Zentner et al., 2010). For *Tetrahymena* Chd3, the gene target selected was the Citrate Synthase gene *CIT2*. In yeast, the Chd3-containing SAGA/SLIK complex binds regions in the 5' upstream

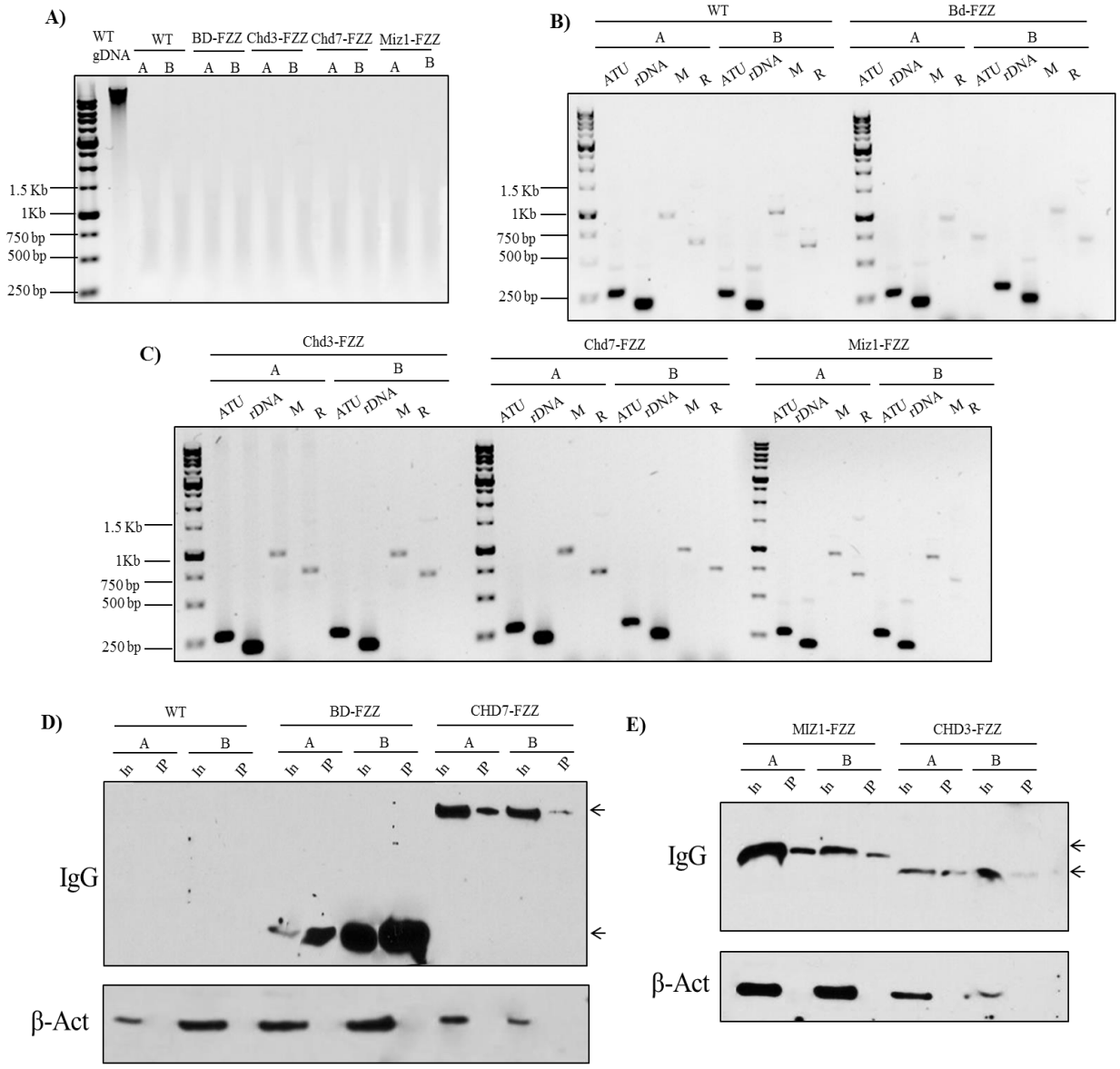


Figure 19. *Confirmation of purified input chromatin and protein for chromatin immunoprecipitation.* **A)** Whole cell lysates of biological replicates (A and B) were sonicated for 25 pulses (0.3s), 10 times. Sheared input chromatin was purified and was visualized on a 1% agarose gel. Chromatin shearing was targeted for 500bp-1kb fragments for downsteam qPCR. To confirm integrity of input DNA, PCR amplification of *Tetrahymena* α-Tubulin (*ATU*, 293 bp), ribosomal DNA (*rDNA*, 227bp), IES M-Element (*M*, 1008 bp) and IES R-Element (*R*, 862 bp) were conducted. This was performed both on **B)** controls (untagged WT cells, and the positive control Bd-FZZ), and **C)** experimental samples (Chd3-FZZ, Chd7-FZZ and Miz1-FZZ). Input (In) and immunopurified (IP) protein samples from **D)** untagged WT, Bd-FZZ (50 kDa), Chd7-FZZ (274 kDa), and **E)** Chd3-FZZ (203 kDa), Miz1-FZZ (>200 kDa) were used for Western blot analysis. Arrows indicate the protein of interest. Although immunoprecipitations of FZZ-tagged proteins following sonication were not highly enriched, the proteins of interest were successfully purified and used for downstream qPCR analysis.

portion of *CIT2* and *GAL10* genes (Pray-grant et al., 2002). In addition to these three candidate genes, six other target genes of diverse functional classes were selected. Rationale for these targets were based on their likelihood of being constitutively expressed and/or regulated by chromatin remodeling proteins/complexes in the macronucleus during vegetative growth. These include the HMGB transcription factor (HMGB1), the HSP70 protein chaperone (Hsp70), the beta-tubulin gene (*BTU2*) and the conjugation-specific programmed DNA deletion protein (*PDD1*). Quantitative PCR was also conducted for the gene encoding the Ribosomal Protein 22 of the Small subunit (*RPS22*). Bd-FZZ is known to bind RPS22 by NGS and qPCR, acting as a positive control.

The qPCR data for the candidate gene interactors was expressed as a fold enrichment of immunoprecipitated chromatin, over input chromatin. Bona fide interactors had a minimum of a fivefold increase, with minimal-to-no binding affinity in wildtype. Results indicated that Chd7 bound to *HMGB1, BTU2, PDD1 and HSP70,* but when compared to wildtype background data, these bindings were likely insignificant. Strong DNA-binding affinities were seen with both Chd3 and Miz1, at the same gene loci. Both proteins bound beta-tubulin (*BTU2*) and citrate synthase (*CIT2*) strongly. Chd3 binding at *BTU2* was ~5500 fold enriched (4725 and 6411 fold per replicate), and Miz1 binding was ~ 1.67×10^{10} fold enriched (1.59×10^{11} and 1.75×10^{9} per replicate). At *CIT2,* Chd3 was enriched by ~564,600 fold (433,746 and 695,543 per replicate), while Miz1 was enriched ~2913 fold (4764 and 1063 per replicate). Interestingly, both Chd3 and Miz1 bound the developmentally regulated gene *PDD1*. Chd3-binding was enriched ~159 fold (195 and 123 fold per replicate), and Miz1 was enriched ~629,390 fold (293,274 and 965,506 fold per replicate). The fold enrichment data were converted to a log10 scale, and displayed as a heat map (**Figure 20**). It is important to note that qPCR was conducted on the positive control

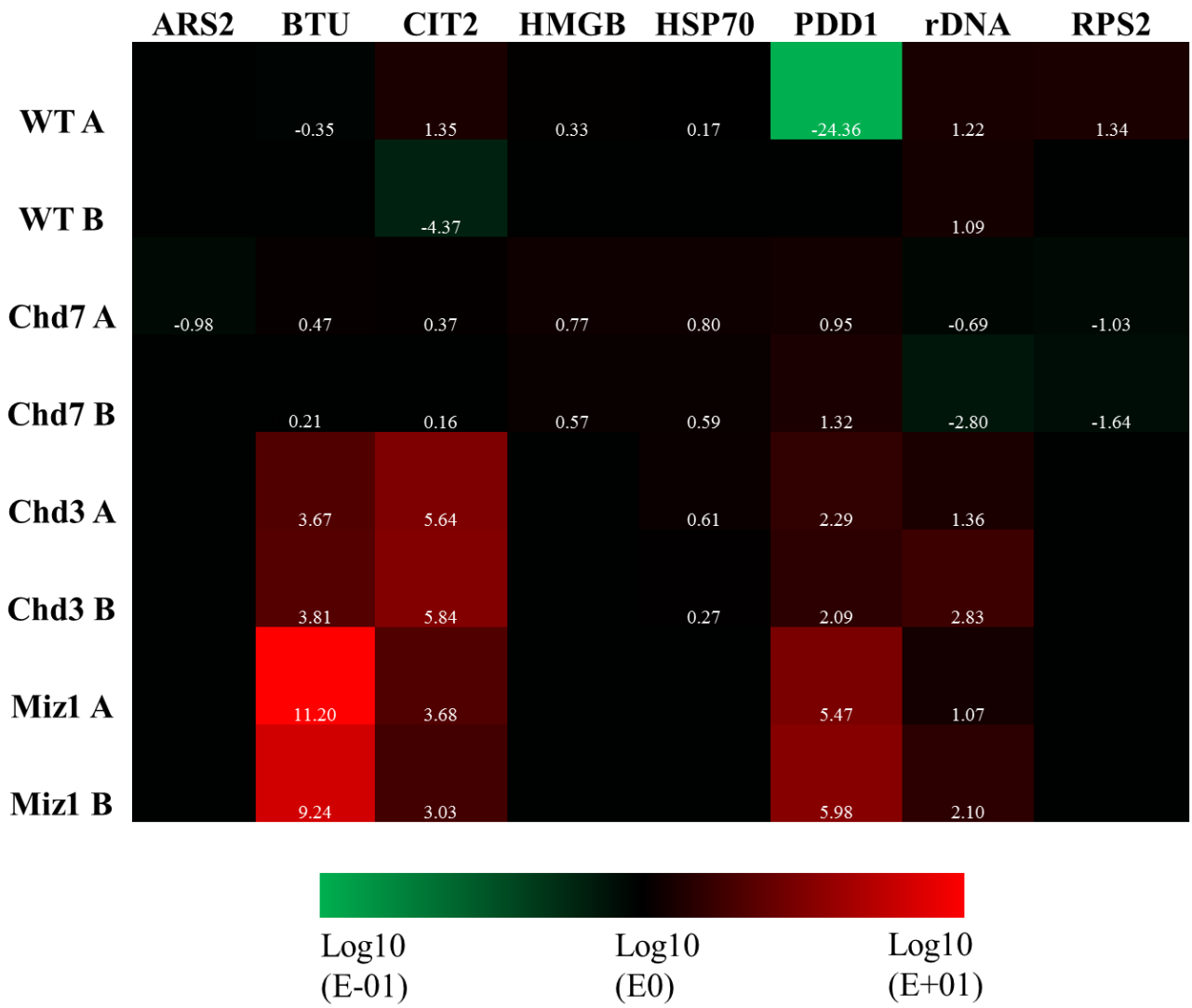


Figure 20. *Quantitative PCR analysis of chromatin from immunoprecipitated Chd3, Miz1 and Chd7.* Gene targets are listed horizontally and included *ARS2,* rDNA, *CIT2*, *HMGB1, HSP70, BTU2, PDD1,* and *RPS22.* Standardized qPCR data for two replicates (A and B) of the immunoprecipitated protein (listed vertically) was converted to fold difference of IP over input. Fold increases were converted to a log10 scale (white), and displayed as a heat map. Black spots indicate zero gene-binding, while spots towards green gradients indicated a minimal, E-01 fold increase. Spots towards red gradients indicated E+01 fold increases with the highest localization of Chd3 and Miz1 towards the *BTU2, CIT2* and *PDD1* genes. Note that green indicates E-01 and red indicates E+01 on the log10 scale.

Bd-FZZ for *RPS22*, and qPCR fold changes of 1.639 x 10^4 and 2.9321 x 10^4 were observed. However, Bd-FZZ qPCR was not conducted for the target genes, and thus was not included in the heat map data.

3.20 *In vitro* Analyses of Chds

3.210 In vitro binding affinities of Tetrahymena Chd7: the synthetic construct

In vitro studies to characterize DNA, RNA and protein interactions of the *Tetrahymena* Chd7 protein were conducted for the following reasons. First, synthetic constructs of two regions of tChd7, the Chromo-Bromo-Chromo and PHD-PHD-PHD regions, were previously generated by DNA 2.0. Chd7 was of interest, since the domain architecture of this protein is unique to Apicomplexan species (including *Tetrahymena)* (**Figure 21A**). Second, AP-MS data on tChd7 were largely uninformative; additional assays on Chd7's binding affinities would help elucidate the chromatin remodeler's function. Third, an *in vitro* model of macromolecular interactions would provide a broader view of tChd7's protein, DNA/chromatin, and RNA interactions.

Following subcloning into BL21 DE3, positive transformants were cultured and protein expression was induced with IPTG. Different induction methods for expression of the CBC and PPP were tested, to assess protein toxicity and sequestration of proteins within *E. coli* into inclusion bodies. Transformants were grown at different temperatures, different [IPTG], and different induction times. During optimization, a DNA 2.0 positive control vector was well induced under every condition tested. The Chd7 CBC expression patterns suggested the protein was only induced at 16^{o}C for 16-20 hours. Subsequent purification by Immobilized Metal Affinity Chromatography (IMAC) enabled successful purification from nickel columns by the poly-histidine tag (**Figure 21B**). The 6xHIS-PPP protein was not expressed under any condition.

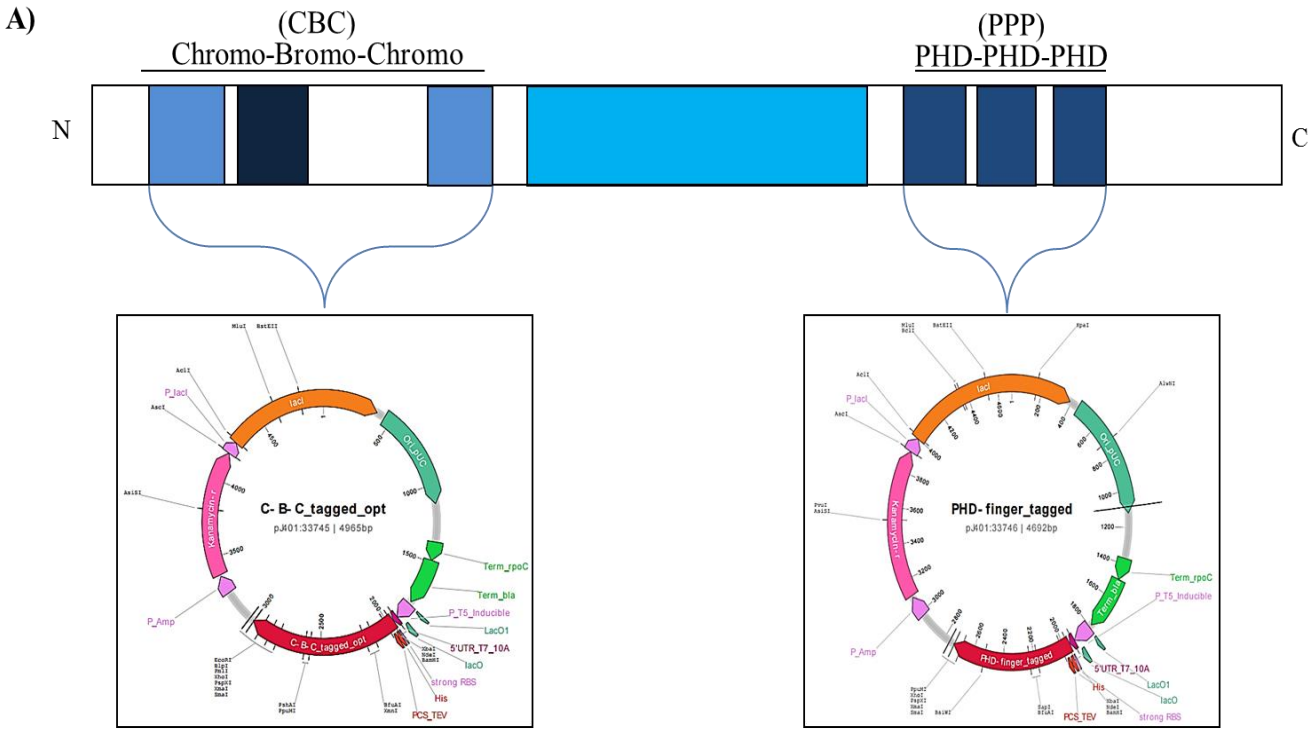


Figure 21. *In vitro analyses of tChd7 guided by unique domain architecture.* **A)** Two N-terminal 6xHis-tagged synthetic genes were cloned by DNA 2.0 (Menlo Park, California) into the pJ401 vector, using the Chromo-Bromo-Chromo (CBC) and PHD-PHD-PHD (PPP) regions of *Tetrahymena* Chd7. The expression vector was transformed into *E. coli* BL21 DE3 for protein expression and purification. **B)** BL21 DE3 transformants were induced at 4°C for 18 hours, with 0.4mM IPTG. Only the CBC construct was well expressed, and the 41.19 kDa protein was purified on a nickel column. Fractions of uninduced (UI) and induced (I) whole cell extracts (WCE), cell lysate (Lys), filtered lysate (Lys$_{Fi}$), flow-through (FT$_{1/2}$), wash, (W) and eluates (E$_{1}$/$_{2}$) were electrophoresed through 8% SDS-PAGE and stained with Coomasie Blue.

3.220 The CHROMO-BROMO-CHROMO region of tChd7 bind H3K36, and four of H3K36 epigenetically modified states

To study chromatin and epigenetic interactions, a commercial Histone Peptide Array was used for HTP interaction analysis. The Active Motif MODified Histone Peptide Array is able to screen a candidate 6xHis-tagged, epigenetic-binding protein against hundreds of candidate 19mer histones and associated histone PTMs. The array is composed of 384 unique histone modification combinations in duplicate - including acetylation, methylation, phosphorylation, and citrullination modifications of the N-terminal tail of core histone proteins. The 6xHIS-CBC construct of Chd7 was expressed and purified in *E. coli*, thus it was selected for the protein microarray analysis.

The domain architecture of the CBC also indicated it was a good target for binding acetylated or methylated histones. Array results indicated that the CBC region of tChd7 bound Histone H3, Lysine 36 (H3K36), and all four of its epigenetic signatures with specificity (**Figure 22A**). In decreasing order, Active Motif microarray processing software indicated that the top four epigenetic marks bound to the CBC were H3K36Ac, H3K36me3, H3K36me2 and H3K36me1 (**Figure 22B**). These top four hits had the highest "specificity factor", a ratio of the averaged spots of interest, divided by the average of all the spots not containing the mark.

All the single modified peptide hits are displayed in **Figure 22C.** While H3K36 modifications (Ac, me3, me2, me1) were abundant, the protein also bound H3K27me1/2/3 and H4K20Ac minimally. In addition to characterizing unmodified and epigenetically-modified histone binding proteins, Active Motif software provided Peptide Intensity Scaling (i.e. measurement of background and peptide signal), as well as Error Scaling (i.e. variation measurements between

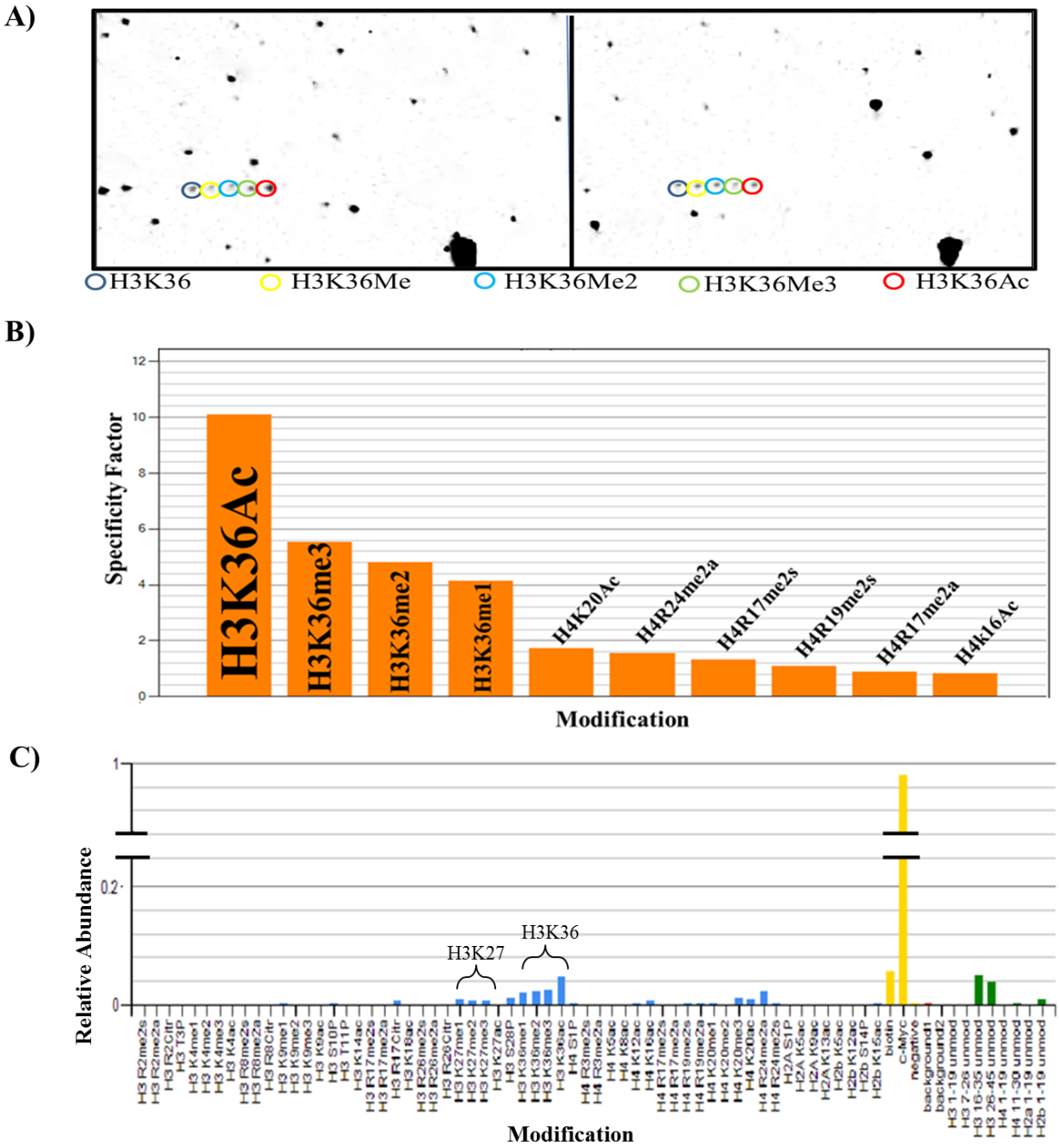


Figure 22. *CBC epigenetic-binding interactions using the Active Motif Histone Peptide Array.* **A)** The purified CBC protein was probed in duplicate using the epigenetic protein microarray. The CBC reproducibly and specifically bound unmodified human histone peptide Histone 3 Lysine 36 (H3K36), as well as all four epigenetic modifications available at this site, acetylation (Ac), monomethylation (Me1), dimethylation (Me2), and trimethylation (Me3). **B)** Active Motif software analysis of the top ten CBC-binding epigenetic marks based on the Specificity Factor (i.e. averaging specific and positive epigenetic marks, and dividing this average by all spots not containing the mark) is provided. Epigenetically-modified H3K36 comprises the top four hits, with H3K36Ac as the most abundant and specific mark. **C)** Analyses of all single modified peptides indicated the CBC protein also bound unmodified H3K27 and its three methylation states (mono-, di- and tri-methylated H3K27), H4K20 (acetylated, di-methylated and monomethylated), H3R17Citr, H4K20me3, H4K20Ac, and H4R24me2. Co-incubation of the array with C-myc was used as a positive control.

the left and right sides of the array). While background and peptide intensities were both slightly low, spot intensity error was in the ideal 0-5% range (i.e. minimal variation of the left and right array).

As an epigenetic mark, H3K36 plays an important role in regulating transcriptional machinery (**Figure 23A**). The *Tetrahymena* region flanking H3K36 shows divergence from mammalian H3K36 in the form of an I35V variation (**Figure 23B)**. However, both isoleucine and valine are non-polar, branched-chain amino acids with minimal structural differences. Thus, AP-WB using commercially available mammalian anti-H3K36Ac and H3K36me3 antibodies were used to probe purified Chd7-FZZ. This would corroborate the high throughput peptide microarray data, *in vivo*. Results suggested Chd7-FZZ bound H3K36me3 minimally in vegetative growth and conjugation (**Figure 23C)**. H3K36Ac was also pulled down, and was highly enriched in co-immunoprecipitants. Thus Chd7 binds H3K36Ac and H3K36me3 *in vivo* throughout the life cycle.

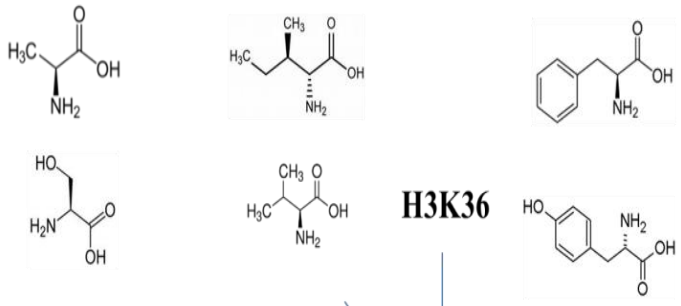
3.230 Crystallization trials of the CBC construct generated microcrystals following HTP screening

To study tChd7 on the protein level, HTP crystallization trials in collaboration with the Hui Lab at the Structural Genomics Consortium (University of Toronto) were conducted. HTP crystallization trials were completed on the CBC region of tChd7 (**Figure 24A**). As a control, crystallization plates were also set up for the *Cryptosporidium* kinase Cgd1, which was a previously crystalized protein by the SGC. Unfortunately, only microcrystals of the CBC were produced after trials of six different 96-well crystallization solution plates – or 576 unique buffer conditions. These buffer conditions included a broad portfolio of crystallization solutions of

A)

H3K36Ac	H3K36Me1	H3K36Me2	H3K36Me3
Transcriptional activation. Present at promoters of RNA polymerase II-transcribed genes, at 5` regulatory regions. Inversely related to H3K36me2.	Transcriptional elongation. Mediates Rpd3 HDAC activity by promoting recruitment to elongating RNA pol II.	Represses intergenic transcription. Present in the coding region and 3′ UTR of genes. Antagonist of H3K27me3. Deposited at DSB sites.	Involved in transcriptional elongation, HR repair and defining exon boundaries. Antagonist of H3K27me3.

B)




```
tH3           1 MARTKQTARKSTGAKAPRKQLASKAARKSAPATGGIKKPHRFRPGTVALREIRKYQKSTD
hH3           1 MARTKQTARKSTGGKAPRKQLATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQKSTE
                ************ ******** ******** *** ***** *********** *****

tH3          61 LLIRKLPFQRLVRDIAHEFKAELRFQSSAVLALQEAAEAYLVGLFEDTNLCAIHARRVTI
hH3          61 FLIRKLPFQRLVREIAQDFKTDLRFQSAAIGALQEASEAYLVGLFEDTNLCAIHAKRVTI
                ************ **  **  ***** *  ***** ****************** ****

tH3         121 MTKDMQLARRIRGER
hH3         121 MPKDIQLARRIRGER
                * ** **********
```

C)

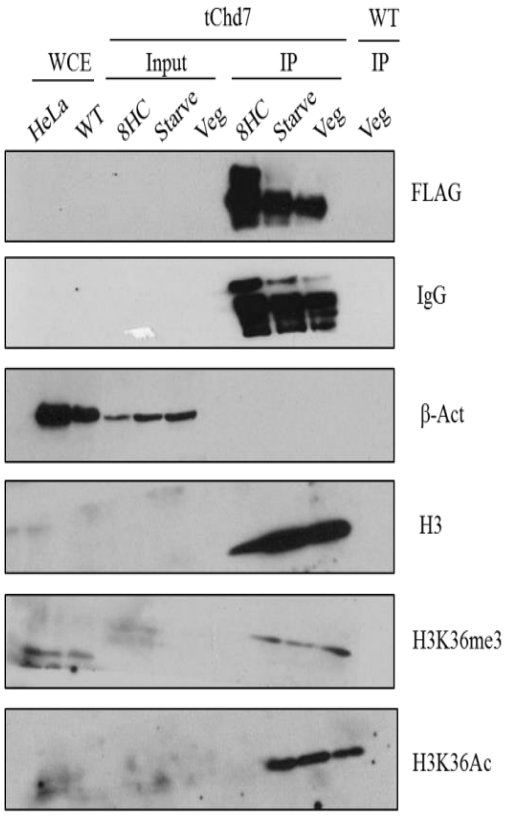


Figure 23. *H3K36 and its epigenetic modifications in T. thermophila*. **A)** H3K36 can be epigenetically modified by acetylation, monomethylation, dimethylation, and trimethylation. **B)** A human and *Tetrahymena* H3 protein sequence alignment indicates 85.2% identity/similarity between the 135 histone H3 residues. There are three amino acids immediately flanking H3K36, which show divergence between the species. These include H3K31, H3K35 and H3K41. **C)** Anti-H3K36me3 and Anti–H3K36Ac Western blot on whole cell extracts (WCEs) and Chd7-FZZ immunoprecipitation fractions suggest Chd7 binds both epigenetic marks *in vivo,* with a selective preference towards H3K36Ac.

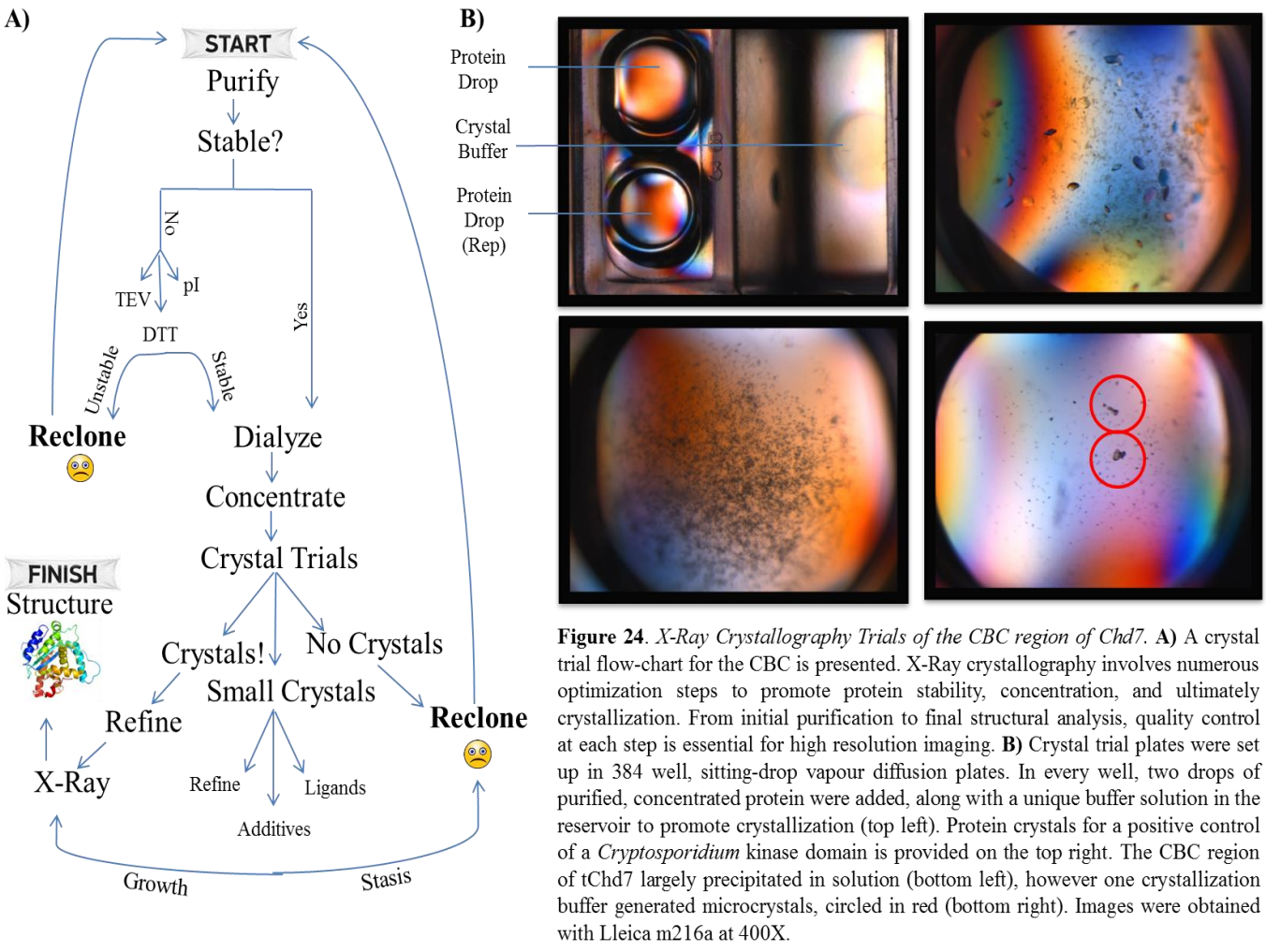


Figure 24. *X-Ray Crystallography Trials of the CBC region of Chd7.* **A)** A crystal trial flow-chart for the CBC is presented. X-Ray crystallography involves numerous optimization steps to promote protein stability, concentration, and ultimately crystallization. From initial purification to final structural analysis, quality control at each step is essential for high resolution imaging. **B)** Crystal trial plates were set up in 384 well, sitting-drop vapour diffusion plates. In every well, two drops of purified, concentrated protein were added, along with a unique buffer solution in the reservoir to promote crystallization (top left). Protein crystals for a positive control of a *Cryptosporidium* kinase domain is provided on the top right. The CBC region of tChd7 largely precipitated in solution (bottom left), however one crystallization buffer generated microcrystals, circled in red (bottom right). Images were obtained with Lleica m216a at 400X.

diverse salts, organic acids, polymers, and pHs. **Figure 24B** shows the microcrystals observed in crystallization buffer #321 (0.2M $MgCl_2$ Hexahydrate, 0.1M HEPES pH 7.5, and 30% isopropanol). Microcrystals in this buffer solution were seen within sitting-drop plates, both for TEV cleaved (i.e. the 6xHis tag was excised) and uncleaved (the 6xHis tag was intact) proteins. Buffer condition #321 was used for a subsequent additive screen - which includes small bioactive ligands that can be added to the buffer of interest, to promote protein crystallization and growth. The library of reagents promote the solubility and crystallization of proteins by manipulating sample-sample, and sample-solvent interactions. The additives include salts, amino

acids, dissociating agents, linkers, polyamines, co-factors, reducing agents, chelating agents, polymers, carbohydrates, detergents and more. Unfortunately, larger crystals did not form following the additive screen, nor was the protein more stable in a buffer condition optimized for protein isoelectric point (pI). Ultimately, the CBC-6xHIS construct was deemed unfit for further crystallization trials

3.240 RNACompete analysis of the CBC region of tChd7 indicates strong poly(G)-binding

RNAcompete is an *in vitro* method for HTP identification of RNA-binding motifs and ultimately sequence preference, for a GST-tagged protein of interest. The assay was developed by Debashish Ray and Timothy Hughes from University of Toronto, Department of Molecular Genetics (Ray et al., 2009). The CBC-6xHIS and PPP-6xHIS constructs of tChd7 were cloned into the pTH6838 GST-containing vector gifted by the Hughes Lab (**Figures 25A-C**). With the GST-tag, both the CBC and PPP constructs were expressed and purifiable, at 77kDa and 65 kDa respectively (**Figure 25D-E**). RNAcompete was conducted on both the CBC and PPP regions of tChd7 to study *in vitro* RNA interactions. In short, the CBC-GST and PPP-GST constructs were co-incubated with a 7-mer RNA pool. Following incubation, the bound RNAs were purified and identified by microarray analysis of RNAs enriched in the bound fraction, as compared to the initial RNA pool. Microarray data were divided into 2 sets (SetA and SetB). The data were transformed into scatterplots to indicate the correlation between 7mers identified in both sets.

RNAcompete data for the CBC region of tChd7 suggested the protein binds poly(U)-sequences, but statistical analyses of Sets A and B generated a cloud-shaped scatterplot, and low Z-scores for associated sets (data not shown). This experiment was considered unsuccessful, due to the

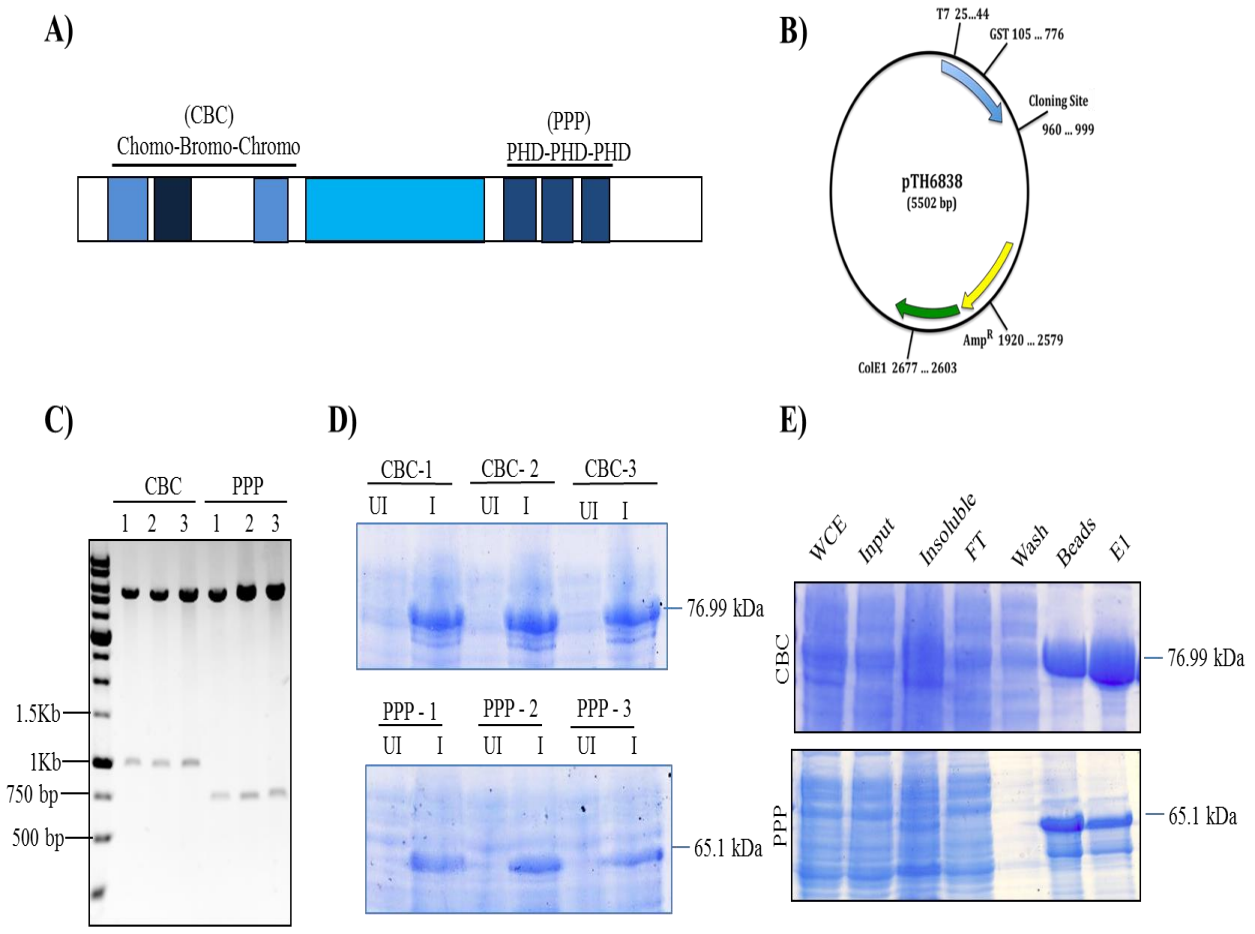


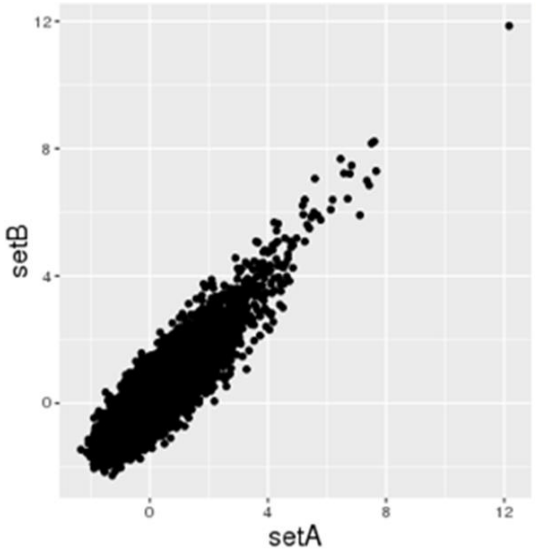
Figure 25. *Cloning the CBC region of Chd7 from the pJ401 6xHis vector, to the pTH6838 GST vector.* **A)** The 6xHis-tagged Chromo-Bromo-Chromo (CBC) and PHD-PHD-PHD (PPP) regions of tChd7 were originally generated by DNA 2.0, and were subsequently cloned into a GST-containing vector for RNAcompete analysis. **B)** The GST vector pTH6838 was used for RNAcompete. **C)** CBC and PPP regions in 6xHis vector were digested with SacI and SalI. Colony PCR of three CBC and PPP transformants in the GST vector corresponded to the expected sizes of 1008bp and 729 bps respectively. **D)** Following sub-cloning into BL21 DE3, uninduced (UI) *E. coli* were induced (I) at 16°C for 18 hours with 0.4mM IPTG for protein expression. The CBC region of Chd7 (41.19 kDa), including the GST tag (35.8 kDa) was observed at the expected size of 76.99 kDa. The PPP region (29.3 kDa), including the GST tag (35.8 kDa) was observed at the expected size of 65.1 kDa. **E)** The CBC and PPP constructs were successfully purified from 1L cultures using GST-conjugated beads. Proteins were eluted with glutathione containing buffer.

CBC's low specificity of RNA-binding. However, **Figure 26** shows the successful RNAcompete data for the PHD-PHD-PHD region of Chd7. Z-score values for all bound RNAs were above 5.5, and a strong correlation between Set A and Set B was observed as indicated by the linear correlative distribution. Numerous spots near the top right corner of the plot indicated highly enriched 7mers for the corresponding RNA motif. For the PHD protein, a clear motif was consistent across replicates. Data suggested that Chd7 bound poly(G) sequences almost exclusively via it's triplicate PHD finger region.

To corroborate RNAcompete data, Electrophoretic Mobility Shift Assays (EMSAs) were conducted using an RNA probe concatenating three copies of 7mer motifs tested by the HTP assay (see **Figure 27A**). Two *in vitro* transcribed RNA probes were tested: the first included the highest scoring RNAcompete motif, a poly-G sequence (Z-score = 12.41), and the second was composed of a low-scoring CA-rich motif (Z-score = -1.344). Between 0-26.88 μM of the triplicate PHD finger protein were tested with the radioactively labeled probes. The G-rich probe formed a distinct RNA-binding complex with the PHD finger region of Chd7 at ~3.84 μM of protein. RNA-binding affinity of the protein was saturated at ~15.36 μM. Free RNAs, which migrated much quicker than the RNA:protein complexes were observed at the bottom of each EMSA, however the progressive loss of free RNA species with the development of RNA:protein complexes was not observed. In an effort to see a stronger shift and the sequestration of free RNAs at higher protein concentrations, an EMSA with new RNA probes was conducted (see **Figure 27B**). G10 and A10 end-labeled RNA probes were incubated with the PHD finger protein. In these analyses, again at 3.84 μM, the PHD finger preferentially bound the G10 probe, as indicated by the sequestration of free G10 with increasing protein. However, the presence of a distinct RNA:protein complex shift was not observed.

A)

B) 7-mer scatter plot



Top 7-mers by Z-score

Set A score		Set B score		Set A+B score	
GGGGGGG	12.17	GGGGGGG	11.85	GGGGGGG	12.41
AGGGGGG	7.66	GGGAGGG	8.22	GGGAGGG	8.14
GGGAGGG	7.60	GGAGGGG	8.16	GGAGGGG	8.09
GGAGGGG	7.51	GGGGGGC	7.68	AGGGGGG	7.74
GGGGGGU	7.42	GGGGGGA	7.47	GAGGGGG	7.41
GAGGGGG	7.35	AGGGGGG	7.29	GGGGGGA	7.40
GGGGGAG	7.11	UGGGGGG	7.22	GGGGGGU	7.39
GGGGGGA	6.83	GUGGGGG	7.21	GGGGGGC	7.33
GUGGGGG	6.77	CGGGGGG	7.06	GUGGGGG	7.22
GGGGAGG	6.70	GAGGGGG	6.99	UGGGGGG	7.09

Figure 26. *RNAcompete analysis for the PPP region of Chd7.* **A)** Following incubation of the PHD-PHD-PHD region with the RNAcompete pool, the protein was identified to preferably bind G-rich sequences, and generated a 7-mer poly(G) motif. Numerous high-intensity spots on microarrays indicated the Chd7 PPP bound RNA well (microarray data not shown). **B)** One-sided *Z*-test scores for each motif were significantly higher than the mean of 5.5. RNA motif logos were created by aligning the top 10 high-scoring 7-mers, and the motifs generated were consistent between replicates (Set A+ Set B). The tChd7 CBC experiment was considered unsuccessful.

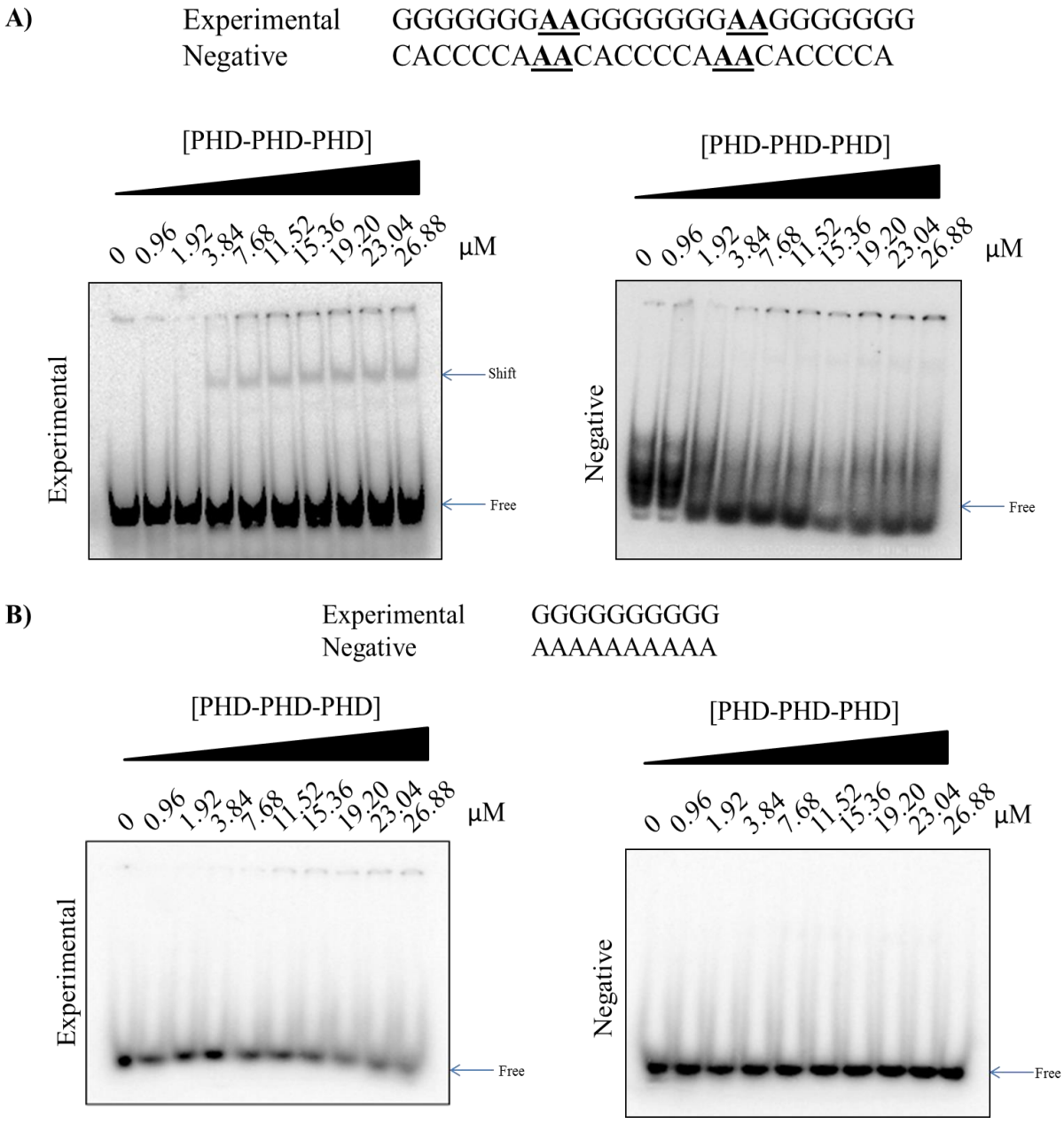


Figure 27. *Electrophoretic mobility shift assay of the triplicate PHD finger of Chd7.* **A)** RNA probes concatenating 3-4 copies of the 7mer motifs (seperated by AA linkers) were generated for analysis by Electrophoretic Mobility Shift Assay (EMSA). The highest scoring RNAcompete motif, a poly-G sequence, was used as the experimental probe (Z-score = 12.41). A negative control composed of a concatenated "CACCCA" sequence was generated, which was a low-scoring motif identified by RNAcompete (Z-score = -1.344). The triplicate PHD finger was used for EMSA using these RNA probes, and electrophoresed through a 6% native PAGE. The formation of an RNA-binding complex composed of the PHD finger and G-rich probe is observed at 3.84 µM. **B)** EMSA analyses of the PHD finger with G10 and A10 probes suggest the protein preferentially binds G-rich sequences, as indicated by the reduction of free G10 RNA species with increasing protein concentration.

3.30 Functional analysis of knockout constructs

3.310 Confirmation of gene knockout Tetrahymena strains

In addition to *in vivo* and *in vitro* functional analyses of RNA, protein, and chromatin interactions of the *Tetrahymena* Chd family of proteins, functional analyses on knockout strains were conducted to examine gene essentiality. Knockouts were generated by complete replacement of the tChd3, tChd7 or tMiz1 endogenous coding regions with a neomycin cassette (**Figure 28A**). The neomycin cassette confers resistance to paromomycin, and is under a $CdCl_2$-inducible MTT1 promoter. Following transformation, genetic assortment, and selection of a single *Tetrahymena* transformant, confirmation of successful gene replacement was conducted on the unique population. Positional PCR amplification of the NEO4 cassette, and either the 5'UTR or 3'UTR region for the gene of interest was conducted on transformants in two different mating types (**Figure 28B**).

3.320 Essentiality

To test for gene essentiality, *Tetrahymena* knockout transformants were grown with and without paromomycin as a selective pressure, for 200 generations. The endogenous genes of interest were PCR amplified. For *CHD3*, *CHD7*, and *MIZ1*, copies of the endogenous genes were present in both drug and no drug conditions post-assortment, suggesting the genes are essential for *Tetrahymena* viability (**Figure 29A)**. More copies of the endogenous genes were present in wild type cells as compared to knockouts. Furthermore, the relative number of endogenous gene copies were also greater in transformants grown without drug for 200 generations, as compared to those grown with drug. Positional PCR amplification of either the 5'UTR or 3'UTR with the NEO4 cassette, showed increased copies of the NEO4 cassette in cells grown with drug (**Figure 29B**). Cells grown for 500 generations without drug eventually exhibited complete loss of the

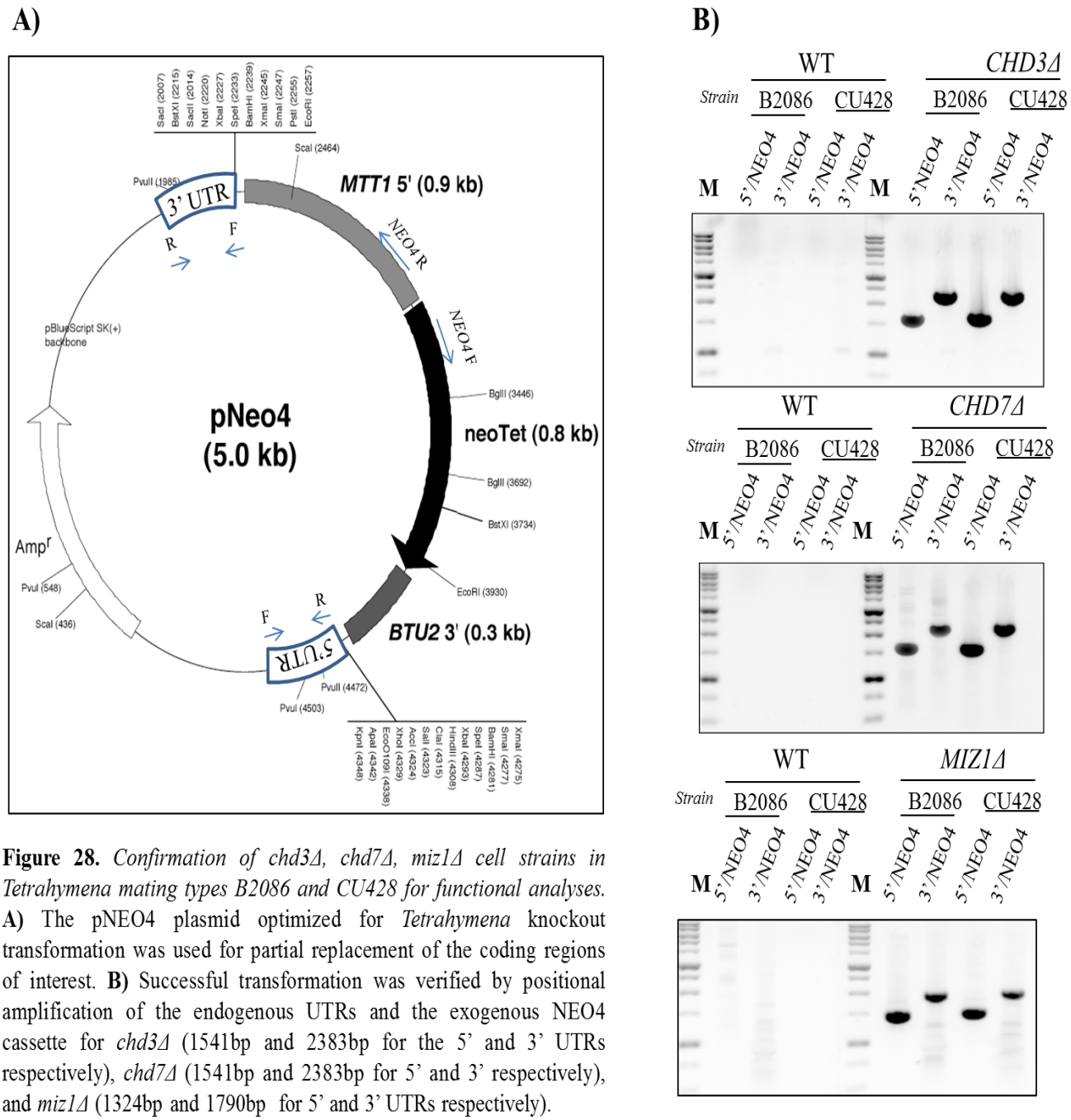


Figure 28. *Confirmation of chd3Δ, chd7Δ, miz1Δ cell strains in Tetrahymena mating types B2086 and CU428 for functional analyses.* **A)** The pNEO4 plasmid optimized for *Tetrahymena* knockout transformation was used for partial replacement of the coding regions of interest. **B)** Successful transformation was verified by positional amplification of the endogenous UTRs and the exogenous NEO4 cassette for *chd3Δ* (1541bp and 2383bp for the 5' and 3' UTRs respectively), *chd7Δ* (1541bp and 2383bp for 5' and 3' respectively), and *miz1Δ* (1324bp and 1790bp for 5' and 3' UTRs respectively).

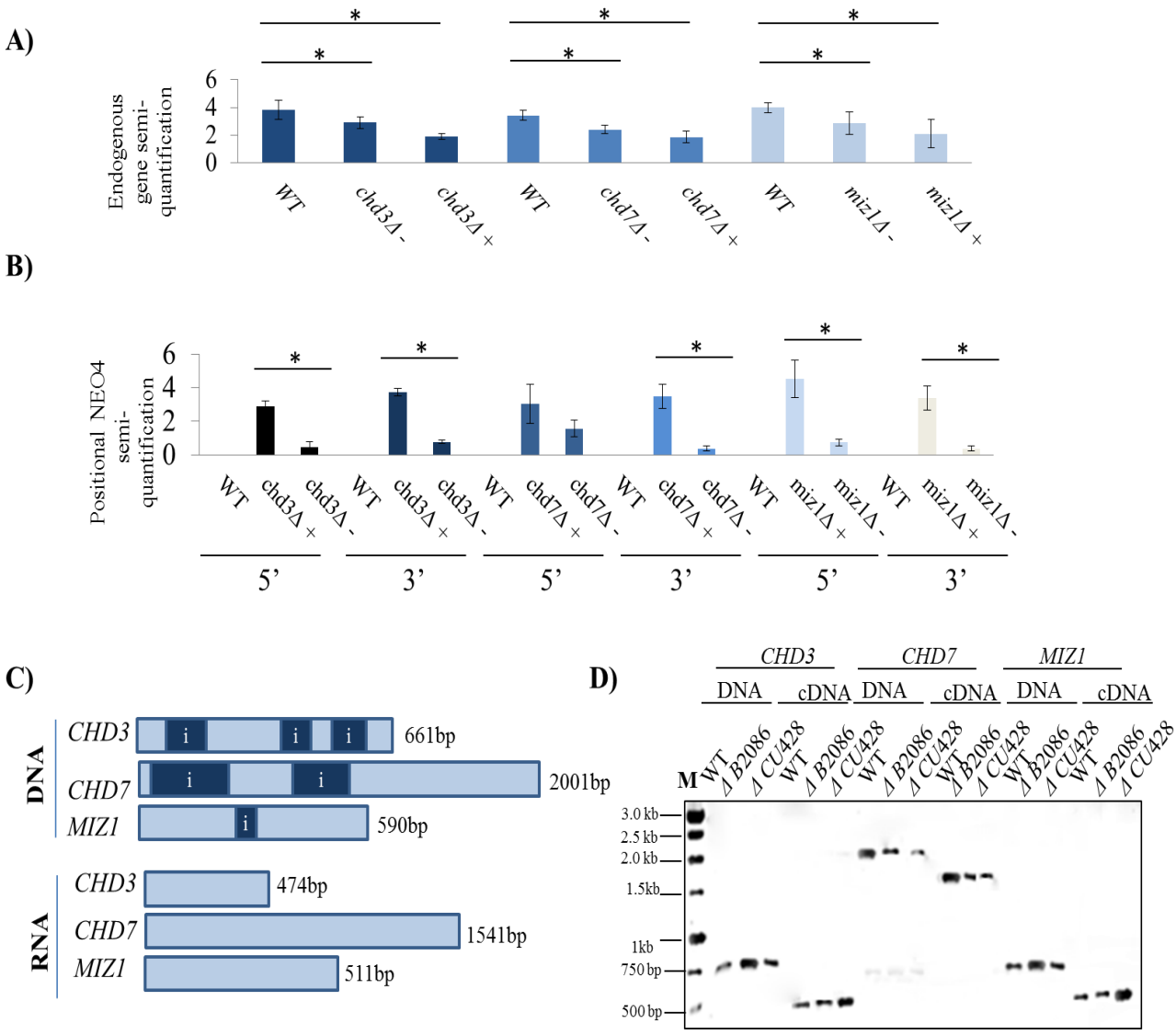


Figure 29. *Expression of endogenous genes and the NEO4 cassette upon removal of selective pressure indicates essentiality.* Knockout strains were grown for 200 generations with (+) or without (-) paromomycin as a selective pressure. All wildtype cells were grown without paromomycin. Upon removal of the drug, *Δchd3, Δchd7,* and *Δmiz1* knockout strains indicated greater **A)** endogenous gene copies, and fewer **B)** NEO4 copies. The opposite trend was observed for strains grown with drug (i.e. fewer endogenous gene copies and more NEO4 copies). Semi-quantification of PCR products was done using ImageJ software. The 95% confidence intervals (CI) are provided. All DNA samples used were roughly equilibrated for concentration by OD spectrophotometer reading, and quantified PCR values for genes of interest were divided by PCRs for β-tubulin. *p<0.05 (Student's t-test). Excluding semi-quantification of 5'/NEO4 region in *Δchd7* strains, change in paromomycin drug pressures resulted in significant differences in gene copy number. **C)** Both RNA and DNA were extracted from all knockout strains. "i" indicates the presence of an intron. **D)** Reverse transcription RT-PCR of endogenous genes was conducted on RNA, as well as PCR on genomic DNA. In both instances, there was incomplete replacement of the endogenous genes, and subsequent expression of the genes of interest within knockout strains, indicated gene essentiality.

KO cassette (data not shown). At the RNA level, the expression of endogenous genes in knockouts were also observed in cells grown with and without drug, supporting the observation of gene essentiality on the DNA level (**Figure 29C**).

Complete developmental immunofluoresence profiles for *CHD3*, *CHD7*, and *MIZ1* knockouts were conducted by DAPI-staining. Normal *Tetrahymena* growth and development was observed for all three knockout strains (data not shown). No difference between wildtype and knockouts was observed in terms of cell morphology and/or nuclear differentiation at each step of development. While *Tetrahymena* knockouts did complete sexual reproduction normally, all three knockout strains exhibited a slight growth defect. During vegetative growth, cells had a slight defect to the effect of an approximately ~15 minute delay per doubling, as compared to wildtype cells under equivalent growth conditions.

5.0 Discussion: Implications, Future Plans, and Current Model

5.10 Chds and Miz1 are chromatin remodelers which regulate transcriptional activity within the macronucleus

The Chd family of chromatin remodelers are SNF2 ATPases which mediate key developmentally-regulated processes in eukaryotes (Platt et al., 2013). In *Tetrahymena*, the peak protein expression levels of Chd3, Chd7 and the Chd3 protein interactor Miz1 all occurred between 6-8 hours of conjugation. At this time point, there was also an observed switch in protein localization patterns identified via indirect immunofluorescence. The localization of these proteins shifted from the old macronucleus, to Anlagen, the new, developing macronuclei. Anlagen are the sites of RNA-directed, irreversible genome silencing events in *T. thermophila*, and these data suggest the chromatin remodeling proteins (i.e. the Chd3:Miz1 complex, and Chd7) are essential for *Tetrahymena* development during zygotic macronuclear expression (Mochizuki & Gorovsky, 2004).

During vegetative growth, AP-MS and SAINT analyses did not identify Chd7 interactors indicative of a potential physiological function. AP-MS on conjugating cells suggests both Chd3 and Chd7 bound histone proteins *in vivo.* These DNA-binding proteins are therefore likely transcriptional regulators during development. Many histones and their variants have already been tagged in the Pearlman Lab, as a part of ongoing research into epigenetic chromatin modifications and chromatin dynamics. To validate Chd:Histone interactions, the reciprocal affinity purification of histone proteins and mass spectrometry during conjugation would be essential. Moreover, AP-MS of histone proteins during conjugation would provide a global view of other novel transcriptional regulators recruited to chromatin during this molecularly complex

time in the *Tetrahymena* lifecycle. AP-MS of Chd3 and Chd7 during conjugation also identified many ribosomal proteins as putative interactors during conjugation. Ribosomal proteins are highly abundant and are "sticky" proteins which are often "frequent flyers" in AP-MS data. Both mammalian Chd7 and yeast Chd1 (the orthologs to tChd7) are known to be essential regulators of ribosomal biogenesis, and have been identified to bind rDNA genes (Lee, Park, & Iyer, 2017; Gabriel E. Zentner et al., 2010). If ribosomal proteins are bona fide interactors, it is possible that the Chd family proteins are also transcriptional regulators of ribosomal protein genes, and may bind these proteins in a negative or positive feedback loop.

5.20 The Chd3:Miz1 complex functions during vegetative growth and conjugation, as a potential SUMO-ligase chromatin remodeling complex

AP-MS data identified Chd3:Miz1 as bona fide protein interactors during vegetative growth and conjugation. These data indicated that the essential proteins must function as an active chromatin remodeling complex throughout the life cycle. Miz1 is a previously uncharacterized MIZ-SP/RING finger protein. First discovered in 1990 after *RING1*, a gene named "Really Interesting New Gene", the cysteine-rich zinc finger family of proteins have diverse roles within eukaryotic cells (Freemont & Trowsdalet, 1990). Similar to PHD fingers, these motifs are coordinated by two zinc ions in a cross-brace topology which occurs in a conserved stretch of Cys3-His-Cys4 (or C3HC4) (Freemont & Trowsdalet, 1990). Unlike the other domains discussed above which are largely chromatin-associated, RING-finger containing proteins can bind histones and other diverse substrates. In fact, the largest class of RING finger proteins are the SUMO and Ubiquitin E3 ligases.

Protein sequence homology of tMiz1 suggests the zinc finger proteins is a candidate SUMO-E3 ligase. SUMOylation of a target protein occurs in three main steps (reviewed in detail by Gareau and Lima, 2010). **Figure 30** illustrates the steps of the eukaryotic SUMO pathway. The process is initiated by the E1 activating enzyme tUba2, which adenylates mature SUMO and forms a thioester bond using a conserved E1 cysteine residue (Nasir et al., 2014; Gareau and Lima, 2010). SUMO is then directly conjugated to another conserved cysteine residue on the SUMO E2 enzyme Ubc9 via a thioester bond (Gareau and Lima, 2010). SUMOylation of a target protein occurs on lysines, and is facilitated but not contingent on the presence of an E3 SUMO ligase (Gareau and Lima, 2010). E3 SUMO ligases enhance the SUMOylation of target substrates. E3s can directly bind a SUMO target, and bring it into close proximity with the SUMO-containing E2 enzyme (Gareau and Lima, 2010). Alternatively, E3s can bind the SUMO-containing E2 enzyme, and promote the release of SUMO from the E2 to the target (Gareau and Lima, 2010). In *Tetrahymena*, the final step of SUMOylation is deSUMOylation of targets, and is likely mediated by Ubiquitin-like protease 2 (Ulp2) (Nasir et al., 2014). It is important to note that target proteins can be multiply SUMOylated at different acceptor lysine residues (Gareau and Lima, 2010). SUMO itself has an acceptor lysine residue, which enables the formation of SUMO chains that can be conjugated to target proteins (Gareau and Lima, 2010).

As mentioned, Western blot analysis of Miz1-FZZ shows that it is 70 kDa larger than expected. A monoSUMOylated protein increases in size by 15-17 kDa (Park-Sarge & Sarge, 2008). This larger Miz1 isoform was almost exclusively observed on Western blots of whole cell extracts, suggesting the multiply SUMOylated form of Miz1 is most abundant (Xiao et al., 2015). However, anti-SUMO AP-WB results did not confirm Miz1:SUMO binding, or its function as an E3 SUMO-ligase. Potentially, the weak anti-SUMO signal from AP-WB data resulted from the

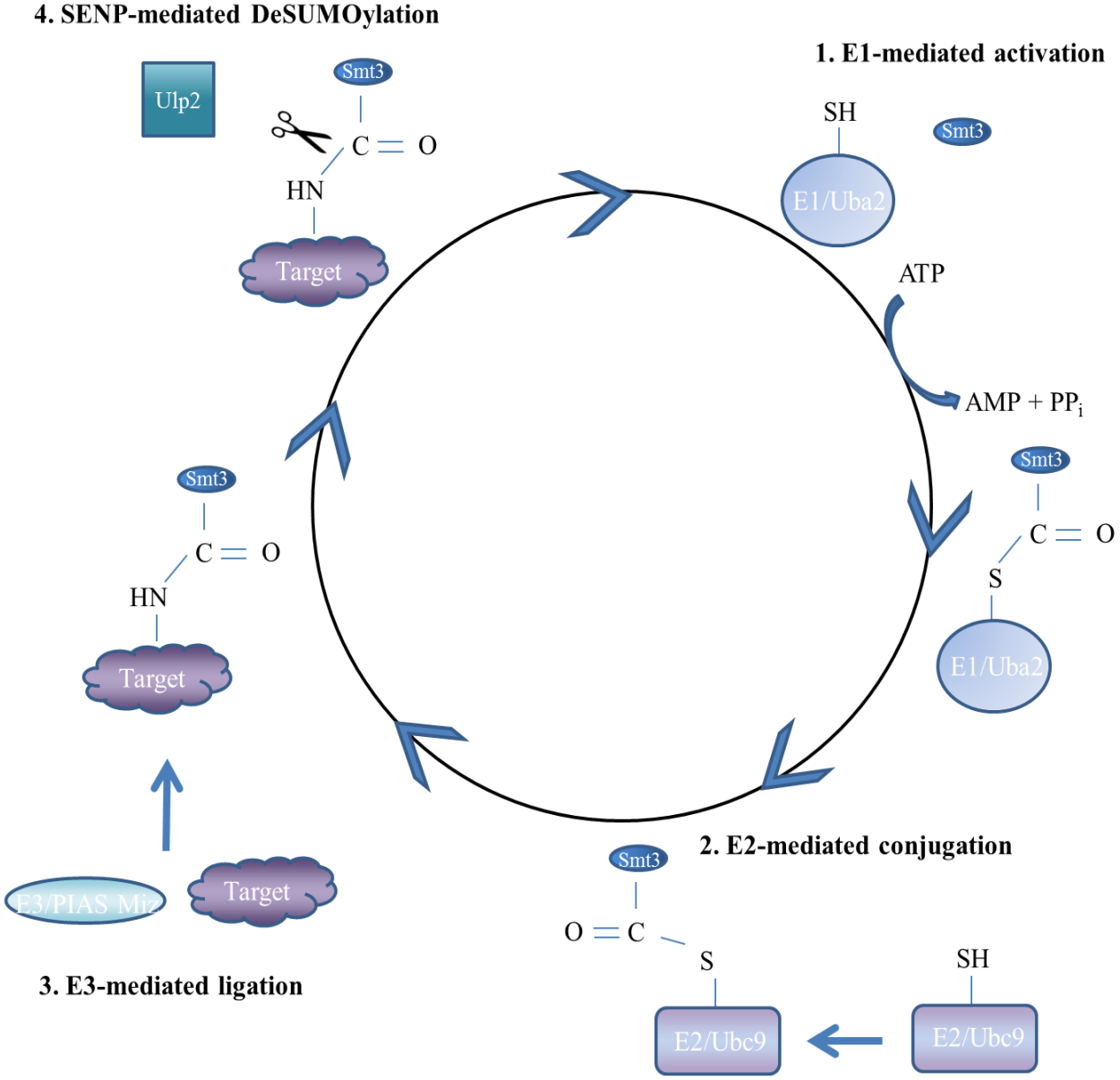


Figure 30. *SUMOylation cycle. Tetrahymena* SUMOylation starts with the processing of SUMO to its mature form, and subsequent formation of a thioester bond to a conserved cysteine residue on an E1 activating enzyme. SUMO is then transferred to the cysteine residue of the SUMO E2 conjugating enzyme Ubc9. Target SUMOylation is enhanced by E3 SUMO ligases, which promote target specificity and E2-mediated SUMOylation efficiency. The deSUMOylation of targets occurs by Ulp proteases. Image adapted from Gareau and Lima (2010).

immunoprecipitation of Miz1 at a developmental stage that did not correspond to Miz1's SUMO-ligase activity. In such a case, the experimental technique BioID would be ideal for the analysis of a developmentally-regulated, post-translational modifying protein, such as Miz1 which may undergo many transient protein interactions. BioID generates proximity-dependent biotinylation of prey proteins, when the bait is fused to a biotin protein ligase (Roux et al., 2012). The subsequent purification of biotinylated proteins is coupled to mass spectrometry (Roux et al., 2012). In *Tetrahymena*, BioID could provide the global view of Miz1's protein interactions throughout the life cycle, instead of the interaction "snapshot" generated by standard affinity purification-mass spectrometry. Following validation of Miz1 as an E3-SUMO ligase, it would be necessary to identify whether tChd3 is a SUMO substrate. *In vitro* SUMOylation assays could also be used to identify if Miz1 facilitates SUMOylation as an E3 ligase, and/or if Chd3 is a potential Miz1-mediated target.

There is strong evidence for Chd:SUMO interactions. *In vitro*, human Chd2 binds one of the three forms of SUMO present in mammalian cells (Vertegaal et al., 2006). dMi-2 (Chd3 in *Drosophila*) is a core component of the NuRD (Nucleosome Remodeling Deacetylase) complex (Fasulo et al., 2012). dMi-2 binds SUMO, and the SUMOylated transcription factors Sp3 and Ttk69, *in vitro* (Murawska & Brehm, 2011). In fact, dMi-2 is an essential factor for mediating SUMO-dependent transcriptional repression by Sp3 (Murawska & Brehm, 2011). This binding interaction is hypothesized to modulate transcriptional repression, by promoting recruitment of Mi-2 containing complexes to these SUMOylated transcription factors (Murawska & Brehm, 2011).

This interaction is conserved among mammals, where the SUMOylated form of the KAP1 transcriptional co-repressor is also known to bind Chd3 in NuRD-mediated repression of gene

targets (Goodarzi, Kurka, & Jeggo, 2011). While BLASTp sequence analysis of KAP1 and Miz1 does not indicate orthology, KAP1 is also an E3 SUMO-ligase with an N-terminal MIZ domain and C-terminal PHD domain. The PHD finger of KAP1 has E3 SUMO-ligase activity (Ivanov et al., 2007). When the E2 enzymeUbc9 SUMOylates KAP1, the Chd3-NuRD complex can bind target gene loci and promote heterochromatinization (reviewed by Peng & Wysocka, 2008).

5.30 Chds bind RNA in vivo throughout the lifecycle, and may guide chromatin silencing events via long, G-rich RNAs

RNA immunoprecipitation of Chd3 and Chd7 suggest the chromatin remodelers largely bind longer RNA species, many over 300nt. These RNAs could be mRNA transcripts, or lncRNAs. If Chds bind mRNAs, it is possible these chromatin remodelers are DRBPs that function as mRNA post-transcriptional regulators. Chds may be regulated by, or recruited to chromatin (as identified by histone binding affinity in APMS data) via these lncRNAs. The longer RNA species may help guide these chromatin remodelers to specific macronuclear loci, where they would alter chromatin structure by physically manipulating nucleosomes. The Chds did not bind small IES RNAs which are characterized as ~28nt species, nor did they bind species smaller than 50nt. As mentioned, mammalian Chd7 is active in the nucleolus, where it functions as a positive regulator of the ribosomal biogenesis pathway (Zentner et al., 2010).

RNAcompete analysis of the triplicate PHD finger region of Chd7 identified a clear RNA-binding motif. Successful RNAcompete analysis of the PHD-PHD-PHD region suggested the protein binds poly(G) sequences strongly *in vitro.* Previous RNAcompete analyses have suggested many non-canonical RBPs bind G-rich RNAs, and these data have been corroborated by other proteomic approaches (Debashish Ray, personal communication). EMSA analyses on

the triplicate PHD finger suggested a preferential binding to G-rich sequences over other RNA probes. However, EMSA quantification, and the subsequent generation of a dissociation constant (K_d) as a measure of the protein's affinity for poly(G) RNAs was not possible. Future studies will aim to re-purify the PHD finger, and run the eluate through a heparin column. This would prevent non-specific binding of the chromatin remodeing protein to nucleic acids, that may occur during purification and contribute to the aggregation of protein:RNA complexes into wells during EMSAs.

While the propensity of tChd7 to potentially bind these G-rich sequences within mRNAs and/or lncRNAs should not be overlooked, this G-binding function may be related to binding higher order RNA structures. G-rich nucleic acid sequences have a propensity to form G-quartets and/or G-quadruplexes. Four guanine bases interact in a planar, square formation via hydrogen bonding interaction, in what is known as a G-quartet (**Figure 31A**). Two or more G-quartets can stack vertically to form a G-quadruplex (**Figure 31B**). The nucleic acid structure is stabilized by a single monovalent cation, usually potassium, which lies in the center of each G-quartet. At the DNA level, highly structured G-quadruplexes play a protective role, as they are normally found at telomeres, but also at UTRs and 5' ends of the of the first intron which implicates this structure in transcriptional regulation (Cammas & Millevoi, 2016). At the RNA level, G-quadruplexes can form on pre-mRNAs to promote post-transcriptional processing, as well as lncRNAs and miRNAs in diverse RNA-guided regulatory mechanisms (Cammas & Millevoi, 2016).

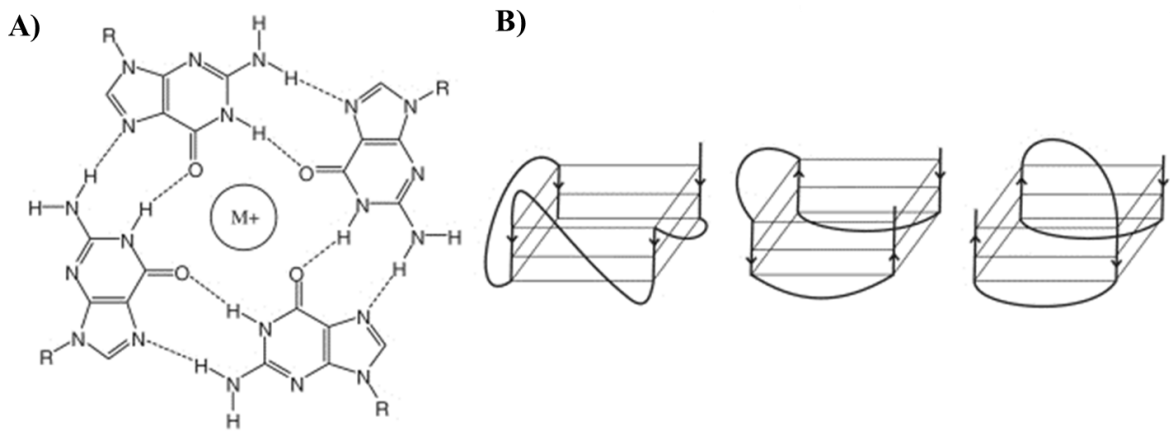


Figure 31. *RNA G-quartets and G-quadruplexes.* **A)** Four guanine residues in a G-rich sequence interact via hydrogen bonding. The carbonyl oxygens within the planar structure are stabilized by a monovalent cation that sits inside the guanine quartet ($K^+ > Na^+ > Li^+$). **B)** Multiple G-quartets can stack on top of each other, in a higher order structure known as a G-quadruplex. RNA G-quadruplexes have a parallel conformation, in that the bottom of one RNA strand connects to the bottom of the next. Adopted from Mullen et al. (2010).

The small RNAs produced during *Tetrahymena* developmentally-regulated genome rearrangements do not follow a strong consensus sequence (Xiong et al., 2012). Rather, the ~28nt non-coding transcripts show sequence heterogeneity (Xiong et al., 2012). However, the small RNAs have a bias for poly(A) tracts, as they are transcribed from A+T-rich intergenic or intronic regions (Xiong et al., 2012). The RNAcompete data thus suggested tChd7 does not bind these small, A-rich transcripts. Previous RIP studies on tChd7 also indicated the protein does not bind the ~28nt small RNAs, but rather longer RNA species (i.e. >80nt). Recent studies in *Tetrahymena* suggest the presence of numerous lncRNAs which co-localize with tChd7 to macronuclei (Kurth & Mochizuki, 2009). Sequencing analysis of long non-coding RNA species present during *Tetrahymena* conjugation remains uncharacterized. However, it is hypothesized that a secondary round of bidirectional transcription within parental macronuclei, enables scanRNA:lncRNA interactions (Schoeberl & Mochizuki, 2011). In a sequence-based, selective degradation process, scanRNA complexes unbound to lncRNAs (i.e. those without homology) move to the developing MACs (Schoeberl & Mochizuki, 2011). In the new MACs, there is also

bi-directional transcription of the genome to generate lncRNAs, which may also interact with scan RNA (Schoeberl & Mochizuki, 2011).

While poly(G) sequences are not associated with small RNA sequences that guide DNA elimination during *Tetrahymena* sexual development, G-rich sequences flanking certain IESs are transcribed in their lncRNA precursors. Recent studies have indicated that transcription of 3' C5 strands in developing macronuclei leads to G5-containing lncRNAs that form G-rich quadruplexes with the unwound 5' G5 DNA (Carle et al., 2016). The excision of IESs is contingent on the formation of this DNA:RNA hybrid, as the protein Lia3 specifically binds the G-quadruplex and recruits the domesticated transposase Tpb2 which ultimately catalyzes the final step of IES excision (Carle et al., 2016; Vogt & Mochizuki, 2013). A hybrid DNA/RNA quadruplex is hypothesized to form during transcription of IESs prior to their excision in the developing MACs (Carle et al., 2016). As such, the transcription of IES and flanking G-rich borders, would enable unwinding of G-tracts which could interact as a G-quadruplex (Carle et al., 2016). If the Chd7-PPP does in fact bind G-rich sequences, G-quadruplexes and G-quartets are also potential targets for future studies, especially in the context of IES excision. Future studies such as RNA-seq on RNA associated with Chd7 would characterize the long RNA species bound to the chromatin remodeler, potentially such as the G-rich lncRNAs described above.

5.40 Chd7 binds H3K36 via its Chromo-Bromo-Chromo region, which may indicate a role in regulating constitutively expressed and/or ribosomal genes

Though the CBC region of tChd7 bound many unmodified and epigenetically-modified chromatin signatures via peptide array analysis, the interaction with H3K36 was the most

supported and was prioritized. This was because the CBC was observed to bind the unmodified state and all four epigenetic states of this histone residue with the most specificity over all other chromatin signatures. This interaction is also of interest when taking into consideration the domain architecture of this protein. Chromodomain proteins classically bind methylated lysine residues, and bromodomain proteins are known to bind acetylated lysine residues, correlating to the methylated and acetylated H3K36-binding of the CBC (Dhalluin et al., 1999; Rea et al., 2000). While it is rare for chromatin-binding proteins to interact with multiple epigenetic states on the same residue, the CBC domain architecture of tChd7 is also rare, and unique to Apicomplexans species.

A close ortholog of tChd7 is yeast yChd1, which is intimately associated with the acetylation and methylation states of H3K36. In yeast, the methyltransferase Set2 interacts with active RNA polymerase II to dimethylate H3K36, to prevent intragenic transcription and reset chromatin structure in the wake of RNA polymerase (Butler & Dent, 2012). yChd1 binds Set2 and RNAPII, where it prevents the exchange of preacetylated histones over ORFs, further suppressing the expression of transcripts from cryptic promoters (Butler & Dent, 2012). It is possible that in *Tetrahymena*, H3K36 is an acetyl/methyl switch which requires tChd7 to use its unique domain architecture to "read" H3K36, and ultimately regulate transcription throughout the lifecycle.

To corroborate the array data, AP-WB analysis was conducted to provide direct, *in vivo* evidence of Chd7:H3K36 interactions. Results suggested tChd7 binds H3K36Ac strongly in vegetative growth, starvation, and conjugation suggesting this chromatin remodeler uses it's bromodomain to read this mark of active transcription throughout the life cycle. The AP-WB signal observed for H3K36me3 was present but somewhat weak, which may indicate that human H3K36me3 antibodies are not suited for *Tetrahymena.* Previous studies on acid-purified *Tetrahymena*

histone proteins suggested H3K36 epigenetic marks are difficult to analyze with mammalian antibodies, due to poor sequence similarity (Morris et al., 2007). Specifically, the *Tetrahymena* H3 sequence contains an isoleucine which flanks H3K36, instead of valine. This different, bulkier isoleucine residue may inhibit antibody recognition (Morris et al., 2007). Alternatively, the H3K36me3 mark itself could be quite rare in *Tetrahymena.* Chd7 may bind this intrinsically low abundance signature, and Western blot analyses reflected H3K36me3's low global expression.

It is also important to note that the protein microarray analyses also suggested the CBC region bound H3K27me1/2/3 minimally. While H3K27me3 was not one of the top ten CBC-binders, this mark is highly enriched *in vivo* during *Tetrahymena* development, and co-expresses and co-localizes with Chd7 to developing macronuclei (Chung & Yao, 2012). The IESs of developing macronuclei are epigenetically marked by H3K9me3 and H3K27me3, which is essential for their irreversible elimination (Liu et al., 2007; Taverna, Coyne, & Allis, 2002).

Chd7's close ortholog, yeast Chd1, is also a member of the SAGA/SLIK histone acetyltransferase and coactivator complex. ChIP analyses indicated Chd7 bound *HMGB1, BTU2, PDD1,* and *HSP70* minimally. It is possible Chd7 is a transcriptional co-activator for *HMGB1, BTU2,* and *HSP70,* using epigenetically modified H3K36 as a docking site to physically manipulate the underlying chromatin structure of these constitutively expressed genes.

5.50 The Chd3 and Miz1 complex functions as a negative co-regulator of PDD1 gene expression during vegetative growth

As protein interactors, Chd3 and Miz1 were expected to bind similar gene loci following ChIP-qPCR analyses. Data indicated the proteins bind *BTU2* and *CIT2* strongly, indicating a role in

transcriptional activation of these genes during vegetative growth. Interestingly, both Chd3 and Miz1 bound the developmentally regulated gene *PDD1*. This gene encodes a chromodomain protein involved in targeting loci destined for elimination during conjugation. The strength of *Tetrahymena* as a nuclearly dimorphic model organism for chromatin studies is exemplified in ChIP analyses. Since the localization patterns of Chd3 and Miz1 are exclusive to the macronucleus, ChIP data indicate gene binding may be a product of transcriptional repression of the *PDD1* gene within the somatic genome. Negative controls can also be tailored to each protein if their nuclear localization patterns are known. For example, vegetative ChIP analyses on MAC-localized proteins can include MIC-localized gene bodies (e.g. IESs) as a negative control.

Ultimately, Next-Generation Sequencing (NGS) of immunoprecipitated chromatin for Chd7, Chd3 and Miz1 would ensure an unbiased, global view of all remodeler:chromatin intreactions. However, these initial qPCR studies provided data that can be used for future comparative analyses to ChIP-seq data on Chd3, Chd7 and Miz1. Furthermore, ChIP-Seq on acetylated or tri-methylated H3K36 could be compared to ChIP-seq data on Chd7 providing co-localization data of the two proteins, and *in vivo* support of Chd7 as an H3K36 reader.

5.60 Crystallization trials conducted on the CBC region of Chd7 suggest re-cloning and/or ligand co-incubation for structural resolution

Crystallization trials on the Chromo-Bromo-Chromo region of *Tetrahymena* Chd7 were conducted. Screens indicated one plate/well combination produced small microcrystals, in a buffer composed of 0.2M $MgCl_2$ Hexahydrate, 0.1M HEPES pH 7.5, 30% isopropanol. These buffer conditions were used for a subsequent additive screen, but unfortunately, larger crystals did not form. The limiting factor to these crystallization trial screens was the stability and

abundance of the protein of interest, which can be resolved in two ways. First, a specific molecular ligand identified to bind the CBC region can be used to promote crystallization. The *in vitro* histone peptide array analysis suggests H3K36 (unmodified, Me1, Me2, Me3, Ac) is an ideal CBC-binding candidate. Not only would co-incubation of H3K36 with the CBC promote protein crystallization via stabilization, but X-ray crystallography of the complex would illustrate the exact mechanism of their potential *in vivo* binding. The second option for successful protein crystalography, is to generate a novel clone of the protein. This can be done by either selecting different domains for expression (e.g. just two tandem domains), or by adding/removing a few amino acids from the CBC sequence, to potentially increase stability of neighboring residues and ultimately the final protein fold. Since the entire CBC region of tChd7 is of interest, due to its unique triplicate domain architecture, the latter option should be pursued.

5.70 Summary

Mass spectrometry suggested that *in vivo,* Chds bind histones, ribosomal proteins, and other chromatin-associated factors. Chd3 specifically and abundantly binds an uncharacterized MIZ/SP-RING protein, here referred to as Miz1. Miz1 orthologs, and AP-WB analyses on Miz1, indicate this protein is a tentative SUMO E3-ligase. RNA-binding affinities of Chds indicate the proteins exhibited similar RNA binding throughout the lifecycle. There were many long RNAs (> 300nt) associated with both Chd3 and Chd7. Chromatin immunoprecipitation indicates Chd7 binds to the coding regions of β-tubulin, rDNA and *PDD1* genes, as well as the 5' upstream region of a citrate synthase gene. Both Chd3 and Miz1 bind coding regions of *HMGB* and *HSP70*, and strongly to the coding region of the developmentally-regulated gene *PDD1*.

In vitro analysis of Chd7 indicated that the protein's CBC region is able to bind unmodified H3K36, as well as mono, di-, tri-methylated, and acetylated states via the unique CBC domain

architecture. Furthermore, the *Tetrahymena* Chd7 is able to bind H3K36Ac strongly *in vivo*, and H3K36me3 minimally based on AP-WB studies. RNA binding analyses via RNAcompete indicate the triple PHD domain region of Chd7 has a strong and selective preference for poly-G sequences.

Complete replacement of the Chd family genes was not possible in *Tetrahymena,* indicating these genes are essential for growth and development. Instead, a gene knockdown genotype was possible. The *Tetrahymena* knockdowns display normal nuclear localization and morphology during development, and have a slight growth defect in the form of developmental delays of ~15 minutes per 2 1/2 hour doubling time.

5.80 Working Model and Final Conclusions

It is hypothesized that Chd3 and Miz1 function as a protein complex throughout the *Tetrahymena* lifecycle. During vegetative growth, the E3-ligase activity of Miz1 may act to target SUMOylation of Chd3 and/or other nuclear proteins, potentially including histones. The epigenetic reader domains of both Chd3 and Miz1 (i.e. PHD and MIZ fingers) suggest they have the capacity to read SUMO epigenetic marks, and the SUMOylation of histones may promote the repression of specific target genes by these proteins such as the vegetatively repressed *PDD1* gene (**Figure 32**).

In the case of Chd7, the protein binds (un)modified H3K36 potentially acting as an epigenetic reader protein via its unique domain architecture. Ultimately, Chd7 could regulate transcription of highly transcribed genes during vegetative growth (e.g. *BTU2, HMGB1, HSP70* and/or ribosomal protein genes), by reading the acetyl/methyl status of this residue. RNAcompete data also suggest Chd7 has the ability to bind poly(G) sequences via its triplicate PHD domains.

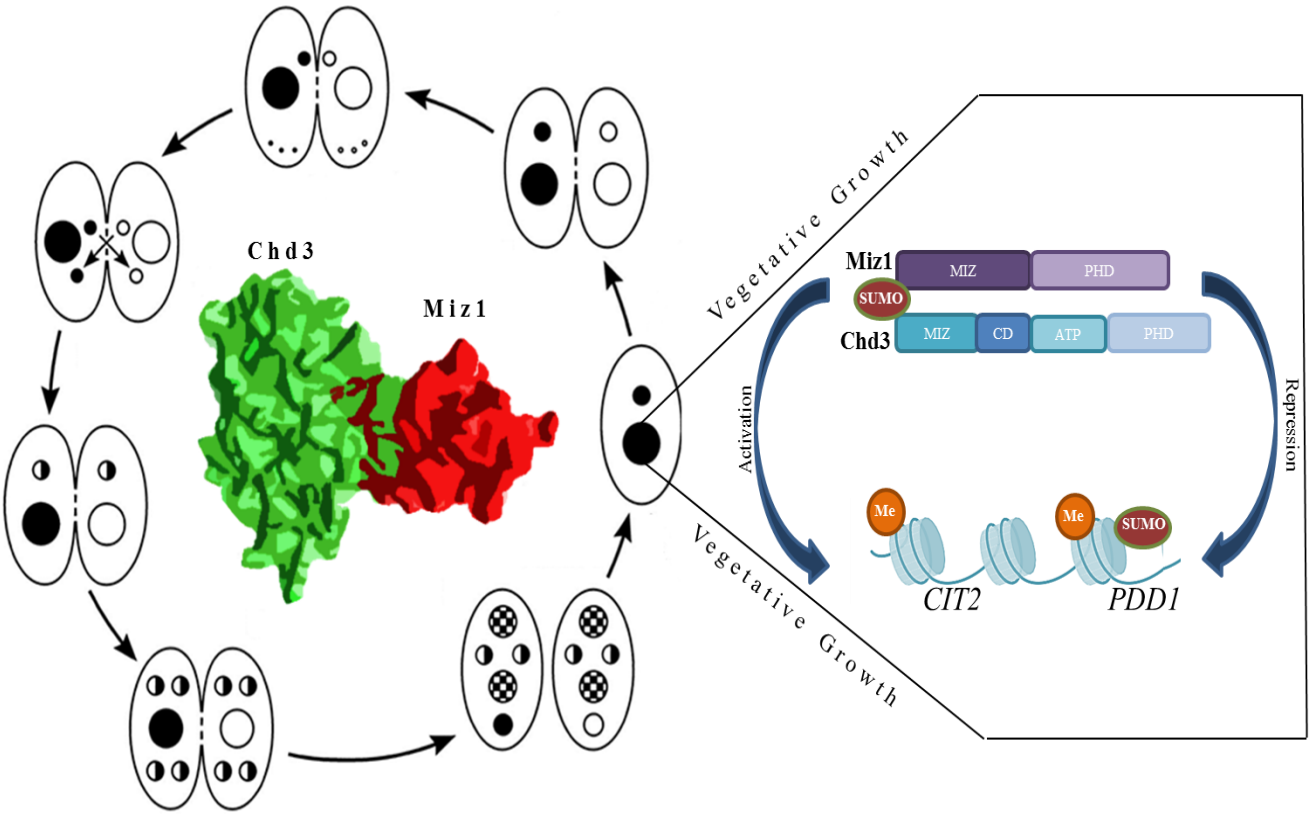


Figure 32. *Chd3 and Miz1 function as a chromatin-associated SUMO ligase protein complex throughout the lifecycle.* During vegetative growth, the E3-ligase activity of Miz1 may promote SUMOylation of target substrates such as nuclear proteins, histones and even itself. Similar to the KAP1:CHD3 protein interaction in mammalian cells, it is possible that the SUMOylation of Miz1 promotes Chd3 binding. Upon binding, heterodimerization of MIZ or PHD domains may contribute to stabilization of the protein interaction (illustrated by arrows). The chromodomain epigenetic reader domain of Chd3 may then be recruited to methylated lysine residues on histone proteins within chromatin, and the Miz1:Chd3 may promote repression of target genes (e.g. *PDD1*) via heterochromatinization, or by the SUMOylation of histones. Alternatively, the Miz1:Chd3 complex could be recruited to active chromatin (e.g. *BTU2, CIT2* genes) via methylated lysine residues of other target histones, to promote gene expression. During development, Miz1:Chd3 may function to activate transcription of zygotically expressed genes or developmentally regulated genes, as well as repress expression of vegetative target genes in an analogous mechanism.

Poly(G) sequences in the form of G-quadruplexes are known to be transcribed at IES boundaries during *Tetrahymena* development of zygotic macronuclei. *In vivo*, it is possible Chd7 binds these G-rich boundaries, and/or these higher order G-rich RNA structures during development (see **Figure 33**).

In sum, the goal of this research project was to conduct functional analyses on *Tetrahymena* Chd family proteins, to identify their *in vivo* mechanisms and developmental functions. Specifically, *in vivo* and *in vitro* models for characterizing novel interactions were performed in the context of DRBP interactions and the *Tetrahymena* scanRNA model. For *in vivo* analyses, a three-way immunoprecipitation protocol was used; protein immunoprecipitation, chromatin immunoprecipitation, and RNA immunoprecipitation. *In vitro* functional analyses on the CBC and PPP regions guided the characterization of Chd7's function via its unique domain structure. The *Tetrahymena* Chd proteins are essential chromatin remodelers with diverse binding-affinities within the cell, especially during RNA-guided, irreversible genome silencing events. These analyses illustrate how RNA and epigenetics play essential and complex roles in the orchestration of chromatin dynamics within eukaryotes.

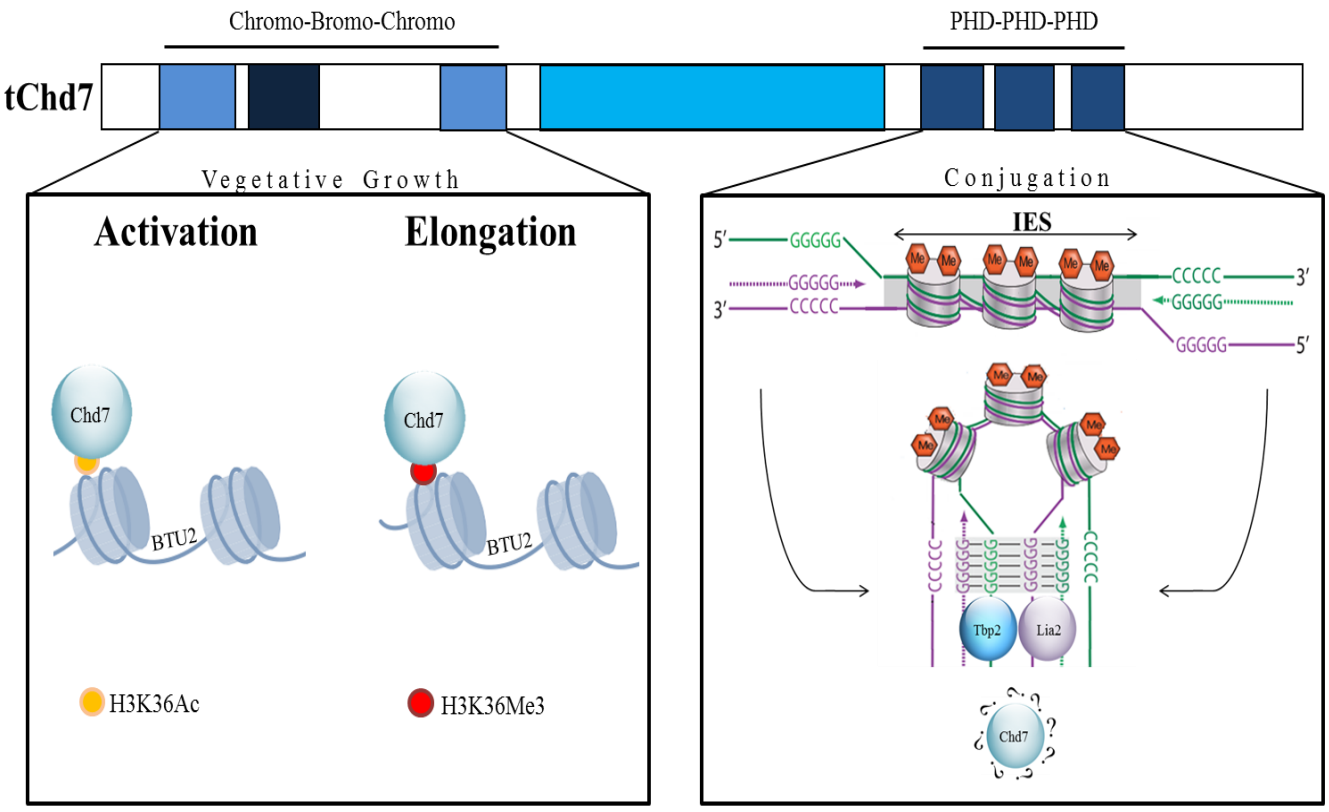


Figure 33. *Chd7 binding affinities are indicative of developmentally-specific DRBP interactions.* Vegetatively (left panel), Chd7 binds H3K36 and may act as a key epigenetic reader protein of this mark, via it's unique domain architecture. Chd7 has the potential to bind the acetylated and methylated forms of H3K36 via it's Chromo-Bromo-Chromo region. Ultimately, Chd7 could regulate transcription of highly transcribed genes during vegetative growth (e.g. *BTU2, HMGB1, HSP70* and/or ribosomal protein genes), by reading the acetyl/methyl status of this residue. Chd7 also has the ability to bind poly(G) sequences via it's triplicate PHD domains. Developmentally (right panel) poly(G) sequences can form G-quadruplexes, and are known to be transcribed at IES boundaries during formation of zygotic macronuclei. The transposase Tbp2 and RNA helicase Lia2 are known to bind G-quadruplexes formed around IESs during their excision. Image of G-quadruplex adapted from Carle *et al.* (2016).

References

Aalfs, J. D., & Kingston, R. E. (2000). What does "chromatin remodeling" mean? *Trends in Biochemical Sciences*, *25*(11), 548–555. http://doi.org/10.1016/S0968-0004(00)01689-3

Aichinger, E., Villar, C. B. R., Farrona, S., Reyes, J. C., Hennig, L., & Köhler, C. (2009). CHD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. *PLoS Genetics*, *5*(8). http://doi.org/10.1371/journal.pgen.1000605

Akematsu, T., Fukuda, Y., Garg, J., Fillingham, J. S., Pearlman, R. E., & Loidl, J. (2017). Post-meiotic DNA double-strand breaks occur in Tetrahymena, and require Topoisomerase II and Spo11. *eLife*, *6*, 1–26. http://doi.org/10.7554/eLife.26176

Akematsu, T., Pearlman, R. E., & Endoh, H. (2010). Gigantic macroautophagy in programmed nuclear death of Tetrahymena thermophila. *Autophagy*, *6*(7), 901–911. http://doi.org/10.4161/auto.6.7.13287

Akhtar, A., Zink, D., & Becker, P. B. (2000). Chromodomains are protein – RNA interaction modules. *Nature*, *732*(1992), 8031–8035.

Arents, G., & Moudrianakis, E. N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(24), 11170–4. http://doi.org/10.1073/pnas.92.24.11170

Aronica, L., Bednenko, J., Noto, T., DeSouza, L. V, Siu, K. W., Loidl, J., … Mochizuki, K. (2008). Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena. *Genes & Development*, *22*(16), 2228–2241.

http://doi.org/10.1101/gad.481908; 10.1101/gad.481908

Bajpai, R., Chen, D. a, Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., … Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, *463*(7283), 958–62. http://doi.org/10.1038/nature08733

Ball, L. J., Murzina, N. V., Broadhurst, R. W., Raine, A. R. C., Archer, S. J., Stott, F. J., … Laue, E. D. (1997). Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. *EMBO Journal*, *16*(9), 2473–2481. http://doi.org/10.1093/emboj/16.9.2473

Baltz, A. G., Vasile, A., Murakawa, Y., Munschauer, M., Schueler, M., Youngs, N., … Landthaler, M. (2012). Resource The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol Cell*, *46*, 674–690. http://doi.org/10.1016/j.molcel.2012.05.021

Bao, Y., & Shen, X. (2007). INO80 subfamily of chromatin remodeling complexes. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *618*(1–2), 18–29. http://doi.org/10.1016/j.mrfmmm.2006.10.006

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., … Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, *129*(4), 823–837. http://doi.org/10.1016/j.cell.2007.05.009

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M., & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, *143*(3), 470–484. http://doi.org/10.1016/j.cell.2010.10.012

Bernstein, E., & Allis, C. D. (2005). RNA meets chromatin. *Genes Dev*, (212), 1635–1655. http://doi.org/10.1101/gad.1324305.GENES

Brownell, J. E., & Allis, C. D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Current Opinion in Genetics & Development*, *6*(2), 176–184.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., & Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, *84*(6), 843–851. http://doi.org/10.1016/S0092-8674(00)81063-6

Butler, J. S., & Dent, S. Y. R. (2012a). Chromatin 'resetting' during transcription elongation: a central role for methylated H3K36. *Nature Structural & Molecular Biology*, *19*(9), 863–864. http://doi.org/10.1038/nsmb.2370

Butler, J. S., & Dent, S. Y. R. (2012b). Chromatin resetting during transcription elongation: a central role for methylated H3K36. *Nature Structural & Molecular Biology*, *19*(9), 863–864. http://doi.org/10.1038/nsmb.2370

Callen, J. P., & Wortmann, R. L. (2006). Dermatomyositis. *Clinics in Dermatology*, *24*(5), 363–373. http://doi.org/10.1016/j.clindermatol.2006.07.001

Cammas, A., & Millevoi, S. (2016). RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Research*, *45*(18), gkw1280. http://doi.org/10.1093/nar/gkw1280

Carle, C. M., Zaher, H. S., & Chalker, D. L. (2016). A Parallel G Quadruplex-Binding Protein

Regulates the Boundaries of DNA Elimination Events of Tetrahymena thermophila. *PLoS Genetics, 12*(3), 1–22. http://doi.org/10.1371/journal.pgen.1005842

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., … Versteeg, R. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science (New York, N.Y.), 291*(5507), 1289–92. http://doi.org/10.1126/science.1056794

Cassiday, L., & James, M. (2002). Having it both ways : transcription factors that bind DNA and RNA. *Nucleic Acids Research*, *30*(19), 4118–4126.

Cassidy-Hanley, D. M. (2012). Tetrahymena in the Laboratory: Strain Resources, Methods for Culture, Maintenance, and Storage. *Methods in Cell Biology*, *109*, 237–276. http://doi.org/10.1016/B978-0-12-385967-9.00008-6

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., … Hentze, M. W. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, *149*(6), 1393–1406. http://doi.org/10.1016/j.cell.2012.04.031

Cech, T. R. (1990). Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Bioscience Reports*, *10*(3), 239–261.

Cech, T. R. (2015). RNA World research — still evolving. *Rna*, 474–475. http://doi.org/10.1261/rna.049965.115.

Cerase, A., Pintacuda, G., Tattermusch, A., & Avner, P. (2015). Xist localization and function: new insights from multiple levels. *Genome Biology*, *16*(1), 166.

http://doi.org/10.1186/s13059-015-0733-y

Cervantes, M. D., Hamilton, E. P., Xiong, J., Lawson, M. J., Yuan, D., Hadjithomas, M., … Orias, E. (2013). Selecting One of Several Mating Types through Gene Segment Joining and Deletion in Tetrahymena thermophila. *PLoS Biology*, *11*(3). http://doi.org/10.1371/journal.pbio.1001518

Cherry, J. M., & Blackburn, E. H. (1985). The internally located telomeric sequences in the germ-line chromosomes of Tetrahymena are at the ends of transposon-like elements. *Cell*, *43*(3 Pt 2), 747–758.

Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., … Nesvizhskii, A. I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature Methods*, *8*(1), 70–3. http://doi.org/10.1038/nmeth.1541

Chung, P. H., & Yao, M. C. (2012). Tetrahymena thermophila JMJD3 homolog regulates H3K27 methylation and nuclear differentiation. *Eukaryotic Cell*, *11*(5), 601–614. http://doi.org/10.1128/EC.05290-11

Coyne, R. S., Nikiforov, M. A., Smothers, J. F., Allis, C. D., & Yao, M. C. (1999). Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. *Molecular Cell*, *4*(5), 865–872.

Coyne, R. S., Thiagarajan, M., Jones, K. M., Wortman, J. R., Tallon, L. J., Haas, B. J., … Methe, B. A. (2008). Refined annotation and assembly of the Tetrahymena thermophila genome sequence through EST analysis, comparative genomic hybridization, and targeted gap closure. *BMC Genomics*, *9*, 562. http://doi.org/10.1186/1471-2164-9-562

D.J., A., & J.D., F. (2000). *Tetrahymena thermophila (2000) Methods in Cell Biology*.

Dalakas, M. C., & Hohlfeld, R. (2003). Polymyositis and dermatomyositis. *The Lancet*, *362*(9388), 971–982. http://doi.org/10.1016/S0140-6736(03)14368-1

de Napoles, M., Jaqueline, M., Wakao, R., Tang, T., Endoh, M., Appanah, R., … Brockdorff, N. (2004). Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Dev Cell*, *7*, 663–676.

de Ruijter, A. J. M., van Gennip, A. H., Caron, H. N., Kemp, S., & van Kuilenburg, A. B. P. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. *The Biochemical Journal*, *370*(Pt 3), 737–49. http://doi.org/10.1042/BJ20021321

Delmas, V., Stokes, D. G., & Perry, R. P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(6), 2414–2418.

Denslow, S. A., & Wade, P. A. (2007). The human Mi-2/NuRD complex and gene regulation. *Oncogene*, *26*(37), 5433–5438. http://doi.org/10.1038/sj.onc.1210611

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., & Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, *399*(6735), 491–496. http://doi.org/10.1038/20974

Ebbert, R., Birkmann, A., & Schüller, H. J. (1999). The product of the SNF2/SWI2 paralogue INO80 of Saccharomyces cerevisiae required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. *Molecular*

Microbiology, *32*(4), 741–751. http://doi.org/10.1046/j.1365-2958.1999.01390.x

Eberharter, A., Sterner, D. E., Schieltz, D., Hassan, A., Yates, J. R., Berger, S. L., & Workman, J. L. (1999). The ADA complex is a distinct histone acetyltransferase complex in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *19*(10), 6621–31.

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., … Orias, E. (2006). Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. *PLoS Biology*, *4*(9), 1620–1642. http://doi.org/10.1371/journal.pbio.0040286

Fasulo, B., Deuring, R., Murawska, M., Gause, M., Dorighi, K. M., Schaaf, C. A., … Tamkun, J. W. (2012). The Drosophila Mi-2 Chromatin-Remodeling Factor Regulates Higher-Order Chromatin Structure and Cohesin Dynamics In Vivo. *Plos Genetics*, *8*(8). http://doi.org/10.1371/journal.pgen.1002878

Feng, W., Kawauchi, D., Körkel-Qu, H., Deng, H., Serger, E., Sieber, L., … Liu, H.-K. (2017). Chd7 is indispensable for mammalian brain development through activation of a neuronal differentiation programme. *Nature Communications*, *8*, 14758. http://doi.org/10.1038/ncomms14758

Fillingham, J. S., Garg, J., Tsao, N., Vythilingum, N., Nishikawa, T., & Pearlman, R. E. (2006). Molecular genetic analysis of an SNF2/brahma-related gene in Tetrahymena thermophila suggests roles in growth and nuclear development. *Eukaryotic Cell*, *5*(8), 1347–1359. http://doi.org/10.1128/EC.00149-06

Fire, A., Albertson, D., Harrison, S. W., & Moerman, D. G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle.

Development (Cambridge, England), *113*(2), 503–514.

Fischle, W., Wang, Y., & Allis, C. D. (2003). Histone and chromatin cross-talk. *Current Opinion in Cell Biology*, *15*(2), 172–183. http://doi.org/10.1016/S0955-0674(03)00013-9

Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., … Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, *438*(7071), 1181–1185. http://doi.org/10.1038/nature04290

Fnu, S., Williamson, E. A., De Haro, L. P., Brenneman, M., Wray, J., Shaheen, M., … Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proceedings of the National Academy of Sciences*, *108*(2), 540–545. http://doi.org/10.1073/pnas.1013571108

Frankel, J. (2000). Cell biology of Tetrahymena thermophila. *Methods in Cell Biology*, *62*, 27–125.

Freemont, P. S., & Trowsdalet, J. (1990). A novel cysteine-rich motif. *Cell*, *265*(63), 6513–6516. http://doi.org/10.1016/j.pain.2014.03.007

G Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B., & Rinn, J. L. (2016). Widespread RNA binding by chromatin-associated proteins. *Genome Biology*, *17*(1), 28. http://doi.org/10.1186/s13059-016-0878-3

Gaspar-Maia, A., Alajem, A., Meshorer, E., & Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. *Nature Reviews Molecular Cell Biology*, *12*(1), 36–47. http://doi.org/10.1038/nrm3036

Goodarzi, A., Kurka, T., & Jeggo, P. (2011). KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. *Nature*, *18*(7), 831-U112.

Gorovsky, M. A. (1973). Macro- and micronuclei of Tetrahymena pyriformis: a model system for studying the structure and function of eukaryotic nuclei. *The Journal of Protozoology*, *20*(1), 19–25.

Grant, P. A., Duggan, L., C, J., Roberts, S. M., Brownell, J. E., Candau, R., … Workman, J. L. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: Characterization of an ada complex and the saga (spt/ada) complex. *Genes and Development*, *11*(13), 1640–1650. http://doi.org/10.1101/GAD.11.13.1640

Greider, C. W., & Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell*, *43*(2 Pt 1), 405–413.

Gu, B., Lee, M., Takahashi, K., Yamanaka, S., Meissner, A., Hawkins, R., … Weissman, I. (2013). Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells. *Cell & Bioscience*, *3*(1), 39. http://doi.org/10.1186/2045-3701-3-39

Haberland, M., Montgomery, R. L., & Olson, E. N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature Reviews. Genetics*, *10*(1), 32–42. http://doi.org/10.1038/nrg2485

Hamilton, E., Bruns, P., Lin, C., Merriam, V., Orias, E., Vong, L., & Cassidy-Hanley, D. (2005). Genome-wide characterization of tetrahymena thermophila chromosome breakage sites. I.

Cloning and identification of functional sites. *Genetics*, *170*(4), 1611–1621. http://doi.org/10.1534/genetics.104.031401

Hamilton, E. P., Kapusta, A., Huvos, P. E., Bidwell, S. L., Zafar, N., Tang, H., … Coyne, R. S. (2016). Structure of the germline genome of Tetrahymena thermophila and relationship to the massively rearranged somatic genome. *eLife*, *5*. http://doi.org/10.7554/eLife.19090

Hanna, M. M., & Liu, K. (1998). Nascent RNA in transcription complexes interacts with CspE, a small protein in E. coli implicated in chromatin condensation. *Journal of Molecular Biology*, *282*(2), 227–39. http://doi.org/10.1006/jmbi.1998.2005

Helbig, R., & Fackelmayer, F. O. (2003). Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosoe territories through its RGG domain. *Chromosoma*, *112*(4), 173–182. http://doi.org/10.1007/s00412-003-0258-0

Hennig, B. P., Bendrin, K., Zhou, Y., & Fischer, T. (2012). Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. *EMBO Reports*, *13*(11), 997–1003. http://doi.org/10.1038/embor.2012.146

Ho, L., & Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature*, *463*(7280), 474–84. http://doi.org/10.1038/nature08911

Horita, D. A., Ivanova, A. V, Altieri, A. S., Klar, A. J. ., & Byrd, R. a. (2001). Solution structure, domain features, and structural implications of mutants of the chromo domain from the fission yeast histone methyltransferase Clr4. *Journal of Molecular Biology*, *307*(3), 861–870. http://doi.org/10.1006/jmbi.2001.4515

Hudson, W. H., & Ortlund, E. A. (2015). The structure, function and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol*, *15*(11), 749–760. http://doi.org/10.1038/nrm3884.The

Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R., & Kadonaga, J. T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell*, *90*(1), 145–155. http://doi.org/10.1016/S0092-8674(00)80321-9

Ivanov, A. V., Peng, H., Yurchenko, V., Yap, K. L., Negorev, D. G., Schultz, D. C., … Rauscher, F. J. (2007). PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain Required for Gene Silencing. *Molecular Cell*, *28*(5), 823–837. http://doi.org/10.1016/j.molcel.2007.11.012

J, G., & Lima, C. (2010). The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol*, *11*(12), 861–871. http://doi.org/10.1038/nrm3011.The

Jacobs, S. A., & Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science (New York, N.Y.)*, *295*(5562), 2080–2083. http://doi.org/10.1126/science.1069473

Jacobs, S. A., Taverna, S. D., Zhang, Y., Briggs, S. D., Li, J., Eissenberg, J. C., … Khorasanizadeh, S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *The EMBO Journal*, *20*(18), 5232–5241. http://doi.org/10.1093/emboj/20.18.5232

James, T. C., & Elgin, S. C. (1986). Identification of a nonhistone chromosomal protein

associated with heterochromatin in *Drosophila melanogaster* and its gene. *Molecular and Cellular Biology*, *6*(11), 3862–3872. http://doi.org/10.1128/MCB.6.11.3862.Updated

Johnsson, P., Lipovich, L., Grandér, D., & Morris, K. V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochimica et Biophysica Acta*, *1840*(3), 1063–71. http://doi.org/10.1016/j.bbagen.2013.10.035

Josling, G. A., Selvarajah, S. A., Petter, M., & Duffy, M. F. (2012). The role of bromodomain proteins in regulating gene expression. *Genes*, *3*(2), 320–343. http://doi.org/10.3390/genes3020320

Kac-rich, Z., Savitsky, P., Krojer, T., Fujisawa, T., Strahl, B. D., Gingras, A., … Wang, C. (2016). Multivalent Histone and DNA Engagement by a PHD / BRD / PWWP Triple Reader Cassette Recruits Article Multivalent Histone and DNA Engagement by a PHD / BRD / PWWP Triple Reader Cassette Recruits ZMYND8 to K14ac-Rich Chromatin. *CellReports*, *17*(10), 2724–2737. http://doi.org/10.1016/j.celrep.2016.11.014

Ketting, R. F., Fischer, S. E. J., Bernstein, E., Sijen, T., Hannon, G. J., & Plasterk, R. H. A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes & Development*, (516), 2654–2659. http://doi.org/10.1101/gad.927801.2654

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., & Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biology*, *2*(7). http://doi.org/10.1371/journal.pbio.0020171

Korber, P., & Barbaric, S. (2014). The yeast PHO5 promoter: From single locus to systems

biology of a paradigm for gene regulation through chromatin. *Nucleic Acids Research*, *42*(17), 10888–10902. http://doi.org/10.1093/nar/gku784

Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., … Greenblatt, J. F. (2002). RNA Polymerase II Elongation Factors of Saccharomyces cerevisiae : a Targeted Proteomics Approach. *Society*, *22*(20), 6979–6992. http://doi.org/10.1128/MCB.22.20.6979

Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., … Greenblatt, J. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Molecular and Cellular Biology*, *23*(12), 4207–18. http://doi.org/10.1128/MCB.23.12.4207

Kurth, H. M., & Mochizuki, K. (2009). Non-coding RNA: A bridge between small RNA and DNA. *RNA Biology*, *6*(2), 138–140. http://doi.org/10.4161/rna.6.2.7792

Kwan, A. H. Y., Gell, D. A., Verger, A., Crossley, M., Matthews, J. M., & Mackay, J. P. (2003). Engineering a protein scaffold from a PHD finger. *Structure*, *11*(7), 803–813. http://doi.org/10.1016/S0969-2126(03)00122-9

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., & Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, *410*(6824), 116–120. http://doi.org/10.1038/35065132

Larson, M. E., Sherman, M. A., Greimel, S., Kuskowski, M., Schneider, A., Bennett, D. A., & Lesné, S. E. (2010). Identification of Histone H3 Lysine 36 Acetylation as a highly conserved Histone Modification, *32*(30), 10253–10266. http://doi.org/10.1523/JNEUROSCI.0581-12.2012.Soluble

Lee, Y., Park, D., & Iyer, V. R. (2017). The ATP-dependent chromatin remodeler Chd1 is recruited by transcription elongation factors and maintains H3K4me3/H3K36me3 domains at actively transcribed and spliced genes. *Nucleic Acids Research*, *45*(12), 1–11. http://doi.org/10.1093/nar/gkx321

Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., & Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*, *442*(7098), 91–95. http://doi.org/10.1038/nature04802

Liu, J., Carmell, M. A., Rivas, F. V, Marsden, C. G., Thomson, J. M., Song, J.-J., … Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science (New York, N.Y.)*, *305*(5689), 1437–41. http://doi.org/10.1126/science.1102513

Liu, Y., Mochizuki, K., & Gorovsky, M. A. (2004). Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(6), 1679–1684. http://doi.org/10.1073/pnas.0305421101

Liu, Y., Taverna, S. D., Muratore, T. L., Shabanowitz, J., Hunt, D. F., & Allis, C. D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes and Development*, *21*(12), 1530–1545. http://doi.org/10.1101/gad.1544207

Lovén, J., Hoke, H. a, Lin, C. Y., Lau, A., David, A., Vakoc, C. R., … Richard, A. (2014). Selective Inhibition of Tumor Oncogenes by Disruption of SuperEnhancers. *Cell*, *153*(2), 320–334. http://doi.org/10.1016/j.cell.2013.03.036.Selective

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. http://doi.org/10.1038/38444

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature Publishing Group*, *389*, 251–260.

Madireddi, M. T., Coyne, R. S., Smothers, J. F., Mickey, K. M., Yao, M. C., & Allis, C. D. (1996). Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. *Cell*, *87*(1), 75–84.

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., … Branlant, C. (2010). 2-D structure of the a region of Xist RNA and its implication for PRC2 association. *PLoS Biology*, *8*(1), 1–16. http://doi.org/10.1371/journal.pbio.1000276

Marfella, Concetta G.A., and A. N. I. (2007). The Chd Family of Chromatin Remodelers. *Mutation Research*, *618*(1–2), 30–40.

Marmorstein, R., & Roth, S. Y. (2001). Histone acetyltransferases: Function, structure, and catalysis. *Current Opinion in Genetics and Development*, *11*(2), 155–161. http://doi.org/10.1016/S0959-437X(00)00173-8

Martens, J. a, & Winston, F. (2002). Evidence that Swi / Snf directly represses transcription in S . cerevisiae. *Genes & Development*, (617), 2231–2236. http://doi.org/10.1101/gad.1009902.GENES

McHugh, C. A., Chen, C.-K., Chow, A., Surka, C. F., Tran, C., McDonel, P., … Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, *521*(7551), 232–6. http://doi.org/10.1038/nature14443

Miao, W., Xiong, J., Bowen, J., Wang, W., Liu, Y., Braguinets, O., … Gorovsky, M. A. (2009). Microarray analyses of gene expression during the Tetrahymena thermophila life cycle. *PLoS ONE*, *4*(2), e4429. http://doi.org/10.1371/journal.pone.0004429

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., & Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science (New York, N.Y.)*, *303*(5656), 343–8. http://doi.org/10.1126/science.1090701

Mochizuki, K., Fine, N. A., Fujisawa, T., & Gorovsky, M. A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell*, *110*(6), 689–699.

Mochizuki, K., & Gorovsky, M. A. (2004a). Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes and Development*, *18*(17), 2068–2073. http://doi.org/10.1101/gad.1219904

Mochizuki, K., & Gorovsky, M. A. (2004b). Small RNAs in genome rearrangement in Tetrahymena. *Current Opinion in Genetics and Development*, *14*(2), 181–187. http://doi.org/10.1016/j.gde.2004.01.004

Mochizuki, K., & Gorovsky, M. A. (2005). A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes & Development*, *19*(1), 77–89. http://doi.org/10.1101/gad.1265105

Morris, S. A., Rao, B., Garcia, B. A., Hake, S. B., Diaz, R. L., Shabanowitz, J., … Strahl, B. D. (2007). Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. *Journal of Biological Chemistry*, *282*(10), 7632–7640. http://doi.org/10.1074/jbc.M607909200

Mujtaba, S., He, Y., Zeng, L., Yan, S., Plotnikova, O., Sachchidanand, … Zhou, M. M. (2004). Structural Mechanism of the Bromodomain of the Coactivator CBP in p53 Transcriptional Activation. *Molecular Cell*, *13*(2), 251–263. http://doi.org/10.1016/S1097-2765(03)00528-8

Muller, S., Filippakopoulos, P., & Knapp, S. (2011). Bromodomains as therapeutic targets. *Expert Reviews in Molecular Medicine*, *13*(September), e29. http://doi.org/10.1017/S1462399411001992

Murawska, M., & Brehm, A. (2011). CHD chromatin remodelers and the transcription cycle. *Transcription*, *2*(6), 244–253. http://doi.org/10.4161/trns.2.6.17840

Nady, N., Min, J., Kareta, M. S., Chedin, F., & Arrowsmith, C. H. (2008). A SPOT on the chromatin landscape? Histone peptide arrays as a tool for epigenetic research. *Trends in Biochemical Sciences*, *33*(7), 305–313. http://doi.org/10.1016/j.tibs.2008.04.014

Nasir, A. M., Yang, Q., Chalker, D. L., & Forney, J. D. (2014). SUMOylation is Developmentally Regulated and Required for Cell Pairing during Conjugation in Tetrahymena. *Eukaryotic Cell*, *14*(2), 170–81. http://doi.org/10.1128/EC.00252-14

Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., … Laue, E. D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature*, *416*(6876), 103-107Nielsen, P.R. et 2002. Structure of th.

http://doi.org/10.1038/nature722

Nikiforov, M. A., Gorovsky, M. A., & Allis, C. D. (2000). A novel chromodomain protein, pdd3p, associates with internal eliminated sequences during macronuclear development in Tetrahymena thermophila. *Molecular and Cellular Biology*, *20*(11), 4128–4134.

Nitarska, J., Smith, J. G., Sherlock, W. T., Hillege, M. M. G., Nott, A., Barshop, W. D., … Riccio, A. (2016). A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell Reports*, *17*(6), 1683–1698. http://doi.org/10.1016/j.celrep.2016.10.022

Noto, T., Kataoka, K., Suhren, J. H., Hayashi, A., Woolcock, K. J., Gorovsky, M. A., & Mochizuki, K. (2015). Small-RNA-Mediated Genome-wide trans-Recognition Network in Tetrahymena DNA Elimination. *Molecular Cell*, *59*(2), 229–42. http://doi.org/10.1016/j.molcel.2015.05.024

Obbard, D. J., Gordon, K. H. J., Buck, A. H., & Jiggins, F. M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1513), 99–115. http://doi.org/10.1098/rstb.2008.0168

Ogas, J., Kaufmann, S., Henderson, J., & Somerville, C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13839–13844.

Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., & O'Shea, C. C. (2017).

ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, *357*(6349). http://doi.org/10.1126/science.aag0025

Owen, D. J., Ornaghi, P., Yang, J. C., Lowe, N., Evans, P. R., Ballario, P., … Travers, A. A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *The EMBO Journal*, *19*(22), 6141–6149. http://doi.org/10.1093/emboj/19.22.6141

Park-Sarge, O. K., & Sarge, K. D. (2008). Detection of sumoylated proteins. *Methods in Molecular Biology*, *464*(859), 255–265. http://doi.org/10.1007/978-1-60327-461-6-14

Paro, R., & Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(1), 263–267.

Peng, J., & Wysocka, J. (2008). It takes a PHD to SUMO. *Trends in Biochemical Sciences*, *33*(5), 191–194. http://doi.org/10.1016/j.tibs.2008.02.003

Plath, K. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science*, *300*(April), 131–135. http://doi.org/10.1126/science.1084274

Platt, J. L., Rogers, B. J., Rogers, K. C., Harwood, A. J., & Kimmel, A. R. (2013). Different CHD chromatin remodelers are required for expression of distinct gene sets and specific stages during development of Dictyostelium discoideum. *Development (Cambridge, England)*, *140*(24), 4926–36. http://doi.org/10.1242/dev.099879

Pray-grant, M. G., Schieltz, D., Mcmahon, S. J., Wood, J. M., Kennedy, E. L., Cook, R. G., …

Workman, J. L. (2002). The Novel SLIK Histone Acetyltransferase Complex Functions in the Yeast Retrograde Response Pathway The Novel SLIK Histone Acetyltransferase Complex Functions in the Yeast Retrograde Response Pathway. *Society*, *22*(24), 8774–8786. http://doi.org/10.1128/MCB.22.24.8774

Quan, T. K., & Hartzog, G. A. (2010). Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of Saccharomyces cerevisiae Spt4-Spt5 in transcription. *Genetics*, *184*(2), 321–334. http://doi.org/10.1534/genetics.109.111526

Radman-Livaja, M., Quan, T. K., Valenzuela, L., Armstrong, J. A., van Welsem, T., Kim, T., … Hartzog, G. A. (2012). A key role for Chd1 in histone H3 dynamics at the 3??? ends of long genes in yeast. *PLoS Genetics*, *8*(7). http://doi.org/10.1371/journal.pgen.1002811

Ragvin, A., Valvatne, H., Erdal, S., Årskog, V., Tufteland, K. R., Breen, K., … Aasland, R. (2004). Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *Journal of Molecular Biology*, *337*(4), 773–788. http://doi.org/10.1016/j.jmb.2004.01.051

Ray, D., Chaudhry, S., Kazan, H., Chan, E. T., Pen, L., Talukder, S., … Hughes, T. R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins, *27*(7). http://doi.org/10.1038/nbt.1550

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., … Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, *406*(6796), 593–599. http://doi.org/10.1038/35020506

Robinson, P. J., & Rhodes, D. (2006). Structure of the "30 nm" chromatin fibre: A key role for

the linker histone. *Current Opinion in Structural Biology*, *16*(3), 336–343. http://doi.org/10.1016/j.sbi.2006.05.007

Rogaia, D., Grignani, F., Carbone, R., Riganelli, D., LoCoco, F., Nakamura, T., … Pelicci, P. G. (1997). The localization of the HRX/ALL1 protein to specific nuclear subdomains is altered by fusion with its eps15 translocation partner. *Cancer Research*, *57*(5), 799–802.

Ropero, S., & Esteller, M. (2007). The role of histone deacetylases (HDACs) in human cancer. *Molecular Oncology*, *1*(1), 19–25. http://doi.org/10.1016/j.molonc.2007.01.001

Rougeulle, C., Chaumeil, J., Sarma, K., David, C., Reinberg, D., Avner, P., … Allis, C. D. (2004). Differential Histone H3 Lys-9 and Lys-27 Methylation Profiles on the X Chromosome Differential Histone H3 Lys-9 and Lys-27 Methylation Profiles on the X Chromosome †. *Molecular and Cellular Biology*, *24*(12), 5475–5484. http://doi.org/10.1128/MCB.24.12.5475

Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology*, *196*(6), 801–810. http://doi.org/10.1083/jcb.201112098

Sanchez, R., & Zhou, M.-M. (2009). The role of human bromodomains in chromatin biology and gene transcription. *Current Opinion in Drug Discovery & Development*, *12*(5), 659–65. http://doi.org/10.1016/j.bbi.2008.05.010

Sanchez, R., & Zhou, M. M. (2011). The PHD finger: A versatile epigenome reader. *Trends in Biochemical Sciences*. http://doi.org/10.1016/j.tibs.2011.03.005

Schindler, U., Beckmann, H., & Cashmore, A. R. (1993). HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *The Plant Journal : For Cell and Molecular Biology*, *4*(1), 137–150.

Schlissel, M. (2004). The spreading influence of chromatin modification. *Nature Genetics*, *36*(5), 438–440. http://doi.org/10.1038/ng0504-438

Schnetz, M. P., Bartels, C. F., Shastri, K., Balasubramanian, D., Zentner, G. E., Balaji, R., … Scacheri, P. C. (2009). Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Research*, *19*(4), 590–601. http://doi.org/10.1101/gr.086983.108

Schoeberl, U. E., & Mochizuki, K. (2011). Keeping the soma free of transposons: Programmed DNA elimination in ciliates. *Journal of Biological Chemistry*, *286*(43), 37045–37052. http://doi.org/10.1074/jbc.R111.276964

Shahbazian, M. D., & Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annual Review of Biochemistry*, *76*, 75–100. http://doi.org/10.1146/annurev.biochem.76.052705.162114

Sharma, A., Jenkins, K. R., Héroux, A., & Bowman, G. D. (2011). Crystal structure of the chromodomain helicase DNA-binding protein 1 (Chd1) DNA-binding domain in complex with DNA. *Journal of Biological Chemistry*, *286*(49), 42099–42104. http://doi.org/10.1074/jbc.C111.294462

Shen, X., & Gorovsky, M. A. (1996). Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell*, *86*(3), 475–483.

Shen, X., Mizuguchi, G., Hamiche, a, & Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature*, *406*(6795), 541–544. http://doi.org/10.1038/35020123

Siggens, L., Cordeddu, L., Rönnerblad, M., Lennartsson, A., & Ekwall, K. (2015). Transcription-coupled recruitment of human CHD1 and CHD2 influences chromatin accessibility and histone H3 and H3.3 occupancy at active chromatin regions. *Epigenetics & Chromatin*, *8*(1), 4. http://doi.org/10.1186/1756-8935-8-4

Silencing, R. T., Wierzbicki, A. T., Zhu, Y., Rowley, M. J., & Bo, G. (2013). A SWI / SNF Chromatin-Remodeling Complex Acts. *Cell*, *49*, 298–309. http://doi.org/10.1016/j.molcel.2012.11.011

Sims, R. J., & Reinberg, D. (2006). Histone H3 Lys 4 methylation: Caught in a bind? *Genes and Development*, *20*(20), 2779–2786. http://doi.org/10.1101/gad.1468206

Sims 3rd, R. J., Chen, C. F., Santos-Rosa, H., Kouzarides, T., Patel, S. S., & Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *The Journal of Biological Chemistry*, *280*(51), 41789–41792. http://doi.org/10.1074/jbc.C500395200

Skourti-Stathaki, K., & Proudfoot, N. J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes and Development*, *28*, 1384–1396. http://doi.org/10.1101/gad.242990.114.Freely

Stokes, D. G., & Perry, R. P. (1995). DNA-binding and chromatin localization properties of CHD1. *Molecular and Cellular Biology*, *15*(5), 2745–2753.

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41–45. http://doi.org/10.1038/47412

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes & Development*, *12*(5), 599–606. http://doi.org/10.1101/gad.12.5.599

Sudarsanam, P., & Winston, F. (2000). The Swi/Snf family. *Trends in Genetics*, *16*(8), 345–351. http://doi.org/10.1016/S0168-9525(00)02060-6

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., & Grewal, S. I. S. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(1), 152–7. http://doi.org/10.1073/pnas.0407641102

Tajul-Arifin, K., Teasdale, R., Ravasi, T., Hume, D. A., Mattick, J. S., Group, R. G. E. R., & Members, G. S. L. (2003). Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. *Genome Research*, *13*(6B), 1416–1429. http://doi.org/10.1101/gr.1015703

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., & Kennison, J. A. (1992). brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2 SWI2. *Cell*, *68*(3), 561–572. http://doi.org/10.1016/0092-8674(92)90191-E

Tauton, J., Hassig, C., & Schrediber, S. (1996). A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p. *Science*, *272*(5260), 408–511.

Taverna, S. D., Coyne, R. S., & Allis, C. D. (2002a). Methylation of histone H3 at lysine 9 targets programmed DNA elimination in Tetrahymena. *Cell*, *110*(6), 701–711. http://doi.org/10.1016/S0092-8674(02)00941-8

Taverna, S. D., Coyne, R. S., & Allis, C. D. (2002b). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. *Cell*, *110*(6), 701–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12297044

Tsukiyama, T., Daniel, C., Tamkun, J., & Wu, C. (1995). ISWI, a member of the SWl2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell*, *83*(6), 1021–1026. http://doi.org/10.1016/0092-8674(95)90217-1

Vakoc, C. R., Sachdeva, M. M., Wang, H., & Blobel, G. A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. *Molecular and Cellular Biology*, *26*(24), 9185–95. http://doi.org/10.1128/MCB.01529-06

Varga-Weisz, P. D., Wilm, M., Bonte, E., Dumas, K., Mann, M., & Becker, P. B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature*, *388*(October), 598–602. http://doi.org/10.1038/41587

Vertegaal, A. C. O., Andersen, J. S., Ogg, S. C., Hay, R. T., Mann, M., & Lamond, A. I. (2006). Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Molecular & Cellular Proteomics : MCP*, *5*(12), 2298–2310. http://doi.org/10.1074/mcp.M600212-MCP200

Vogt, A., & Mochizuki, K. (2013). A domesticated PiggyBac transposase interacts with heterochromatin and catalyzes reproducible DNA elimination in Tetrahymena. *PLoS*

Genetics, *9*(12), e1004032. http://doi.org/10.1371/journal.pgen.1004032; 10.1371/journal.pgen.1004032

Wade, P. a, Jones, P. L., Vermaak, D., & Wolffe, a P. (1998). A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. *Current Biology : CB*, *8*, 843–846. http://doi.org/10.1016/S0960-9822(98)70328-8

Winkler, D. D., & Luger, K. (2011). The histone chaperone FACT: Structural insights and mechanisms for nucleosome reorganization. *Journal of Biological Chemistry*, *286*(21), 18369–18374. http://doi.org/10.1074/jbc.R110.180778

Winston, F., & Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends in Genetics*, *8*(11), 387–391. http://doi.org/10.1016/0168-9525(92)90300-S

Woo, T., Chao, J., & Yao, C. (2016). Dynamic distributions of long double-stranded RNA in Tetrahymena during nuclear development and genome rearrangements. *JCS*, *21*(January).

Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P., & Collins, F. S. (1997). Characterization of the CHD family of proteins. *Genetics*, *94*(October), 11472–11477. http://doi.org/10.1073/pnas.94.21.11472

Wu, H., Min, J., Lunin, V. V., Antoshenko, T., Dombrovski, L., Zeng, H., … Schapira, M. (2010). Structural biology of human H3K9 methyltransferases. *PLoS ONE*, *5*(1). http://doi.org/10.1371/journal.pone.0008570

Xiao, Y., Pollack, D., Nieves, E., Winchell, A., Callaway, M., & Vigodner, M. (2015). Can your

protein be sumoylated? A quick summary and important tips to study SUMO-modified proteins. *Anal Biochem*, *8*(5), 583–592. http://doi.org/10.1002/aur.1474.Replication

Xiong, J., Gao, S., Dui, W., Yang, W., Chen, X., Taverna, S. D., … Liu, Y. (2016). Dissecting relative contributions of cis-and trans-determinants to nucleosome distribution by comparing Tetrahymena macronuclear and micronuclear chromatin. *Nucleic Acids Research*, *44*(21), 10091–10105. http://doi.org/10.1093/nar/gkw684

Xiong, J., Lu, X., Zhou, Z., Chang, Y., Yuan, D., Tian, M., … Miao, W. (2012). Transcriptome analysis of the model protozoan, tetrahymena thermophila, using deep RNA sequencing. *PLoS ONE*, *7*(2). http://doi.org/10.1371/journal.pone.0030630

Xiong, J., Lu, X., Zhou, Z., Chang, Y., Yuan, D., Tian, M., … Miao, W. (2012). Transcriptome analysis of the model protozoan, Tetrahymena thermophila, using Deep RNA sequencing. *PloS One*, *7*(2), e30630. http://doi.org/10.1371/journal.pone.0030630; 10.1371/journal.pone.0030630

Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Cote, J., & Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Molecular Cell*, *2*(6), 851–861.

Yan, Z., Cui, K., Murray, D. M., Ling, C., Xue, Y., Gerstein, A., … Wang, W. (2005). PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes. *Genes and Development*, *19*(14), 1662–1667. http://doi.org/10.1101/gad.1323805

Yao, M.-C., Yao, C.-H., & Monks, B. (1990). The controlling sequence for site-specific

chromosome breakage in tetrahymena. *Cell*, *63*(4), 763–772. http://doi.org/10.1016/0092-8674(90)90142-2

Yao, M. C., Choi, J., Yokoyama, S., Austerberry, C. F., & Yao, C. H. (1984). DNA elimination in tetrahymena: A developmental process involving extensive breakage and rejoining of DNA at defined sites. *Cell*, *36*(2), 433–440. http://doi.org/10.1016/0092-8674(84)90236-8

Yap, K. L., Li, S., Muñoz-cabello, A. M., Raguz, S., Zeng, L., Gil, J., … Zhou, M. (2011). Molecular Interplay of the Non-coding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Mol Cell*, *38*(5), 662–674. http://doi.org/10.1016/j.molcel.2010.03.021.Molecular

Yoo, K. H., & Hennighausen, L. (2011). EZH2 methyltransferase and H3K27 methylation in breast cancer. *International Journal of Biological Sciences*, *8*(1), 59–65. http://doi.org/10.7150/ijbs.8.59

Yu, G. L., Hasson, M., & Blackburn, E. H. (1988). Circular ribosomal DNA plasmids transform Tetrahymena thermophila by homologous recombination with endogenous macronuclear ribosomal DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(14), 5151–5155.

Yu, T., Meiners, L. C., Danielsen, K., Wong, M. T. Y., Bowler, T., Reinberg, D., … Albert Basson, M. (2013). Deregulated FGF and homeotic gene expression underlies cerebellar vermis hypoplasia in CHARGE syndrome. *eLife*, *2013*(2), 1–15. http://doi.org/10.7554/eLife.01305

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., & Zhu, B. (2011). H3K36 methylation

antagonizes PRC2-mediated H3K27 methylation. *Journal of Biological Chemistry*, *286*(10), 7983–7989. http://doi.org/10.1074/jbc.M110.194027

Zeng, L., & Zhou, M. M. (2002). Bromodomain: an acetyl-lysine binding domain. *FEBS Letters*, *513*(1), 124–128.

Zentner, G. E., Hurd, E. A., Schnetz, M. P., Handoko, L., Wang, C., Wang, Z., … Scacheri, P. C. (2010). CHD7 functions in the nucleolus as a positive regulator of ribosomal RNA biogenesis. *Human Molecular Genetics*, *19*(18), 3491–3501. http://doi.org/10.1093/hmg/ddq265

Zentner, G. E., Hurd, E. A., Schnetz, M. P., Handoko, L., Wang, C., Wang, Z., … Scacheri, P. C. (2010). CHD7 functions in the nucleolus as a positive regulator of ribosomal RNA biogenesis. *Human Molecular Genetics*, *19*(18), 3491–3501. http://doi.org/10.1093/hmg/ddq265

Zentner, G., Layman, W., Martin, D., & Scacheri, P. (2011). Molecular and phenotypic aspects of CHD7 mutation in CHARGE syndrome. *American Journal of Medical Genetics*, (3), 674–686. http://doi.org/10.1002/ajmg.a.33323.Molecular

Zhang, K., Mosch, K., Fischle, W., & Grewal, S. I. S. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nature Structural & Molecular Biology*, *15*(4), 381–8. http://doi.org/10.1038/nsmb.1406

Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S., & Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell*, *95*(2), 279–289. http://doi.org/10.1016/S0092-

8674(00)81758-4

Zhang, Y., & Reinberg, D. (2001). Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes and Development*, *15*(18), 2343–2360. http://doi.org/10.1101/gad.927301

Zhu, B., Mandal, S. S., Pham, A. D., Zheng, Y., Erdjument-Bromage, H., Batra, S. K., … Reinberg, D. (2005). The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes and Development*, *19*(14), 1668–1673. http://doi.org/10.1101/gad.1292105

Zhu, Y., Rowley, M. J., Böhmdorfer, G., & Wierzbicki, A. T. (2013). A SWI/SNF Chromatin-Remodeling Complex Acts in Noncoding RNA-Mediated Transcriptional Silencing. *Molecular Cell*, *49*(2), 298–309. http://doi.org/10.1016/j.molcel.2012.11.011

Appendices

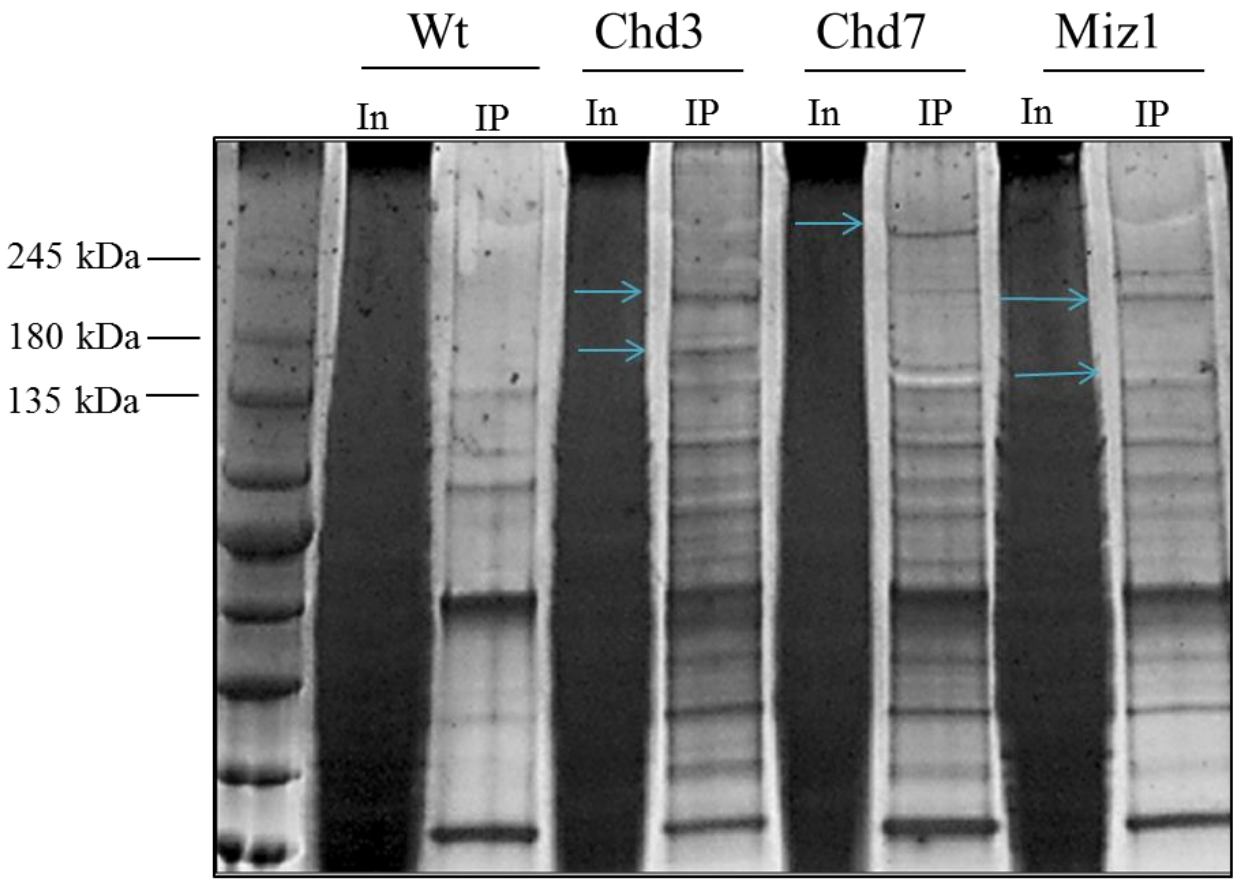
Appendix A. *AP-MS protein interactors and SAINT for Chd7-FZZ, in duplicate.* Chd7-FZZ and interacting complexes during **A)** vegetative growth and **B)** 8 hours conjugation were immunopurified, and identified by Mass Spectrometry, with SAINT filtering. TTHERM_ numbers (left column) are annotated (right column). Peptide hits for each interactor, or "SAINT HITS", are listed in the central column. Although all these interactors are validated by SAINT, no functionally interesting proteins were selected for downstream applications (e.g. reciprocal tagging).

A)

TTHERM Number	SAINT HITS	Protein Annotation
TTHERM_00193800	1691	CHD7 **BAIT**
TTHERM_00691520	4	KRS1 (lysyl-tRNA synthetase)

B)

TTHERM Number	SAINT HITS	Protein Annotation
TTHERM_00193800	162	CHD7 **BAIT**
TTHERM_00348510	1362	BTU1 (Beta TUbulin)
TTHERM_00558620	381	ATU1 (Alpha-TUbulin)
TTHERM_01151500	336	RPS3 (Ribosomal Protein of the Small subunit 3)
TTHERM_01053000	270	RPS3 (Ribosomal Protein of the Small subunit 3)
TTHERM_00633360	227	HTB1 (Histone H Two B)
TTHERM_00333210	224	RPL4 (Ribosomal Protein of the large subunit)
TTHERM_00736480	205	RPL5 (Ribosomal protein of the large subunit)
TTHERM_00339810	163	RPL17 (Ribosomal Protein of the Large subunit #17)
TTHERM_00454080	142	RPS22 (Ribosomal Protein of the Small subunit 22)
TTHERM_00467660	106	RPS24 (Ribosomal Protein of the Small subunit 24)
TTHERM_00636970	77	RPP0 (Ribosomal Protein P0 of the large subunit)
TTHERM_00266390	75	RPS15 (Ribosomal Protein of the Small subunit 15)
TTHERM_00697490	73	RPL14 (Ribosomal protein of the large subunit)
TTHERM_00992760	45	RPS20 (Ribosomal Protein of the Small subunit 20)
TTHERM_00338210	39	RPL18 (Ribosomal Protein of the Large subunit #18)



Appendix B. *Silver Stain Analysis of Chd3:Miz1 Protein Interactions.* Affinity purification of FZZ-tagged strains was conducted on Chd3 and its interaction partner Miz1, as well as Chd7. Immunopurified (IP) protein samples were electrophoresed alongside total soluble input protein fractions (In), through 8% SDS-PAGE. Gels were stained with ProteoSilver Stain. As a qualitative assay of protein interactions, this supports Chd3:Miz1 AP-MS interaction data. Within IP lanes, tagged proteins and their binding partners are marked by blue arrows. Chd3-FZZ is 203 kDa, Chd7-FZZ is 274 kDa, and Miz1-FZZ is 167 kDa. For Miz1-FZZ IPs, the Chd3 protein (without the 18kDa FZZ tag) is visible at 185 kDa, and for Chd3-FZZ IPs, the Miz1 protein is present at 149 kDa.

Appendix C. *Protein Interactors identified by mass spectrometry and SAINT for Chd3-FZZ, in duplicate.* Following affinity purification of Chd3-FZZ proteins during **A)** vegetative growth and **B)** 7 hours of conjugation, immunoprecipitated protein complexes were analyzed by Mass Spectrometry, and filtered by SAINT. TTHERM_ numbers (left column) were annotated (right column) from data on www.ciliate.org. SAINT filtered peptide hits that were identified by mass spectrometry are listed as 'SAINT HITS' (central column).

A)

TTHERM Number	SAINT HITS	Protein Annotation
TTHERM_00049310	268	CHD3 **BAIT**
TTHERM_00442270	277	MIZ zinc finger family protein
TTHERM_00105110	86	HSP70 (Heat Shock Protein)
TTHERM_00363210	32	Mitochondrial carrier protein
TTHERM_00051730	19	Hypothetical RNA recognition motif protein

B)

TTHERM Number	SAINT HITS	Protein Annotation
TTHERM_00049310	297	CHD3 **BAIT**
TTHERM_00442270	256	MIZ zinc finger family protein
TTHERM_00191800	21	Aconitate hydratase
TTHERM_01128560	17	RPS10 (Ribosomal Protein of the Small subunit 10)
TTHERM_00926980	12	Acetyl-CoA acyltransferases family protein
TTHERM_00773160	9	Hypothetical - acyl-CoA dehydrogenase activity
TTHERM_00344030	8	Dehydrogenase, isocitrate/isopropylmalate family protein
TTHERM_00318570	6	Citrate synthase
TTHERM_00529790	6	CIT1 (CITrate synthase)
TTHERM_00051730	5	Hypothetical protein
TTHERM_00691520	5	KRS1 (lysyl-tRNA synthetase)
TTHERM_01048110	5	Glutamate/Leucine/Phenylalanine/Valine dehydrogenase family protein
TTHERM_00056140	4	ERS3 (glutaminyl-tRNA synthetase 3)
TTHERM_00787250	4	Glutamate/Leucine/Phenylalanine/Valine dehydrogenase family protein
TTHERM_00812780	4	RPL38 (Ribosomal Protein of the Large subunit #38)
TTHERM_00878160	4	Malate synthase A family protein
TTHERM_00899460	4	Acetyl-CoA acyltransferases family protein
TTHERM_00016360	3	BBC53 (chromasome regulation)
TTHERM_00133500	3	Peroxidase family protein
TTHERM_00437600	3	Succinyl-CoA synthetase, beta subunit family protein
TTHERM_00442210	3	RPN2 (26S proteasome regulatory subunit N2)
TTHERM_00446090	3	RPN3 (26S proteasome regulatory subunit N3)
TTHERM_01276420	3	DYH3 (DYnein Heavy chain)
TTHERM_00069210	2	Glycosyl transferase, group 2 family protein
TTHERM_00114360	2	SRS1 (seryl-tRNA synthetase 1)
TTHERM_00551090	2	RPT6 (26S proteasome regulatory subunit T6)
TTHERM_00554190	2	TDH2 (glyceraldehyde-3-phosphate dehydrogenase)
TTHERM_00649110	2	RPN12 (26S proteasome regulatory subunit N12)
TTHERM_00659010	2	hypothetical protein oxidoreductase activity
TTHERM_00661650	2	GST34 (Glutathione S-transferase mu 34, TtGSTm34, GSTm34)
TTHERM_00713350	2	hypothetical protein; putative transmembrane
TTHERM_00723390	2	Dehydrogenase, isocitrate/isopropylmalate family protein
TTHERM_01043090	2	RPS27 (Ribosomal Protein of the Small subunit 27)

Appendix D. Complete SAINT analysis of replicate AP-MS data for Miz1-FZZ in vegetative growth.

	Miz1 Vegetative	
TTHERM Number	**Spectral Counts**	**Protein Annotation**
TTHERM_00442270	235	MIZ zinc finger family protein **BAIT**
TTHERM_00049310	352	CHD3
TTHERM_00348510	124	BTU1 (Beta TUbulin)
TTHERM_00047550	111	RPA1 (RNA Polymerase I largest subunit)
TTHERM_00558620	75	ATU1 (Alpha-TUbulin)
TTHERM_01075780	46	RPA2 (RNA polymerase I subunit 2)
TTHERM_00529860	36	VMA2 (Vacuolar Membrane Atpase)
TTHERM_00551160	32	TDH1 (glyceraldehyde-3-phosphate dehydrogenase 1)
TTHERM_00384910	28	CEN1 (CENtrin)
TTHERM_00938820	26	EEF2 (Eukaryotic translation Elongation Factor)
TTHERM_00193740	24	RPS2 (Ribosomal Protein of the Small subunit 2)
TTHERM_00190950	22	ACT1 (ACTin)
TTHERM_00105110	21	HSP70 (Heat Shock Protein)
TTHERM_00339640	19	VMA1 (Vacuolar Membrane Atpase)
TTHERM_00275740	17	CHC1 (Clathrin Heavy Chain-like protein)
TTHERM_00189170	13	HHF2 (Histone H Four)
TTHERM_00052160	12	TTNB (Tetrin B)
TTHERM_00633360	3	HTB1 (Histone h Two B)

Appendix E. Complete SAINT analysis of replicate AP-MS data for Miz1-FZZ at 8 hours of conjugation.

	Miz1 8HC	
Gene Name	**Annotation**	**SAINT HITS**
TTHERM_00348510	BTU1 (Beta TUbulin)	1373
TTHERM_00497880	RPL3 (Ribosomal Protein of the Large subunit)	1257
TTHERM_00558620	ATU1 (Alpha-TUbulin)	1233
TTHERM_00333210	RPL4 (Ribosomal Protein of the large subunit)	1171
TTHERM_01151500	RPS3 (Ribosomal Protein of the Small subunit 3)	1087
TTHERM_00578640	RPL13 (Ribosomal protein of the large subunit)	859
TTHERM_00499420	TGP1 (Tetrahymena G-DNA binding Protein)	852
TTHERM_01053000	RPL10 (Ribosomal protein of the large subunit)	613
TTHERM_00655820	translation elongation factor EF-1, subunit alpha	607
TTHERM_00736480	RPL5 (Ribosomal protein of the large subunit)	602
TTHERM_00829380	RPL11 (Ribosomal protein of the large subunit)	573
TTHERM_00633360	HTB1 (Histone h Two B)	548
TTHERM_00666540	RPL21 (Ribosomal Protein of the Large subunit #21)	547
TTHERM_00577240	L1P family of ribosomal proteins containing protein	536
TTHERM_00339810	RPL17 (Ribosomal Protein of the Large subunit #17)	459
TTHERM_00075670	RPL43 (Ribosomal Protein of the Large subunit #43)	406
TTHERM_00149300	RPS4 (Ribosomal Protein of the Small subunit 4)	373
TTHERM_00467660	RPS24 (Ribosomal Protein of the Small subunit 24)	345
TTHERM_00136120	RPL6 (Ribosomal protein of the large subunit)	308
TTHERM_01049200	Glutamate/Leucine/Phenylalanine/Valine dehydrogenase family protein	301
TTHERM_00149798	RPS6 (Ribosomal Protein of the Small subunit 6)	298
TTHERM_00193740	RPS2 (Ribosomal Protein of the Smalle subunit 2)	266
TTHERM_01289110	RPL15 (Ribosomal protein of the large subunit)	252
TTHERM_00316500	HTA2 (Histone H Two A)	248
TTHERM_00790790	HTA1 (Histone h Two A)	246
TTHERM_00636970	RPP0 (Ribosomal Protein P0 of the large subunit)	230
TTHERM_00414440	enoyl-CoA hydratase/isomerase family protein	228
TTHERM_00189170	HHF2 (Histone H Four)	226
TTHERM_00227270	RPL22 (Ribosomal Protein of the Large subunit #22)	223
TTHERM_01151600	40s ribsomal protein	218
TTHERM_00499440	TGP3 (Tetrahymena G-DNA binding Protein)	213
TTHERM_00049310	CHD3 (SNF2-related protein; Chromo-Helicase-DNA-binding domain)	212
TTHERM_00773340	RPL8 (Ribosomal protein of the large subunit)	207
TTHERM_00145010	hypothetical protein	205
TTHERM_00444510	RPL19 (Ribosomal Protein of the Large subunit #19)	204
TTHERM_00382350	RPL26 (Ribosomal Protein of the Large Subunit #26)	203
TTHERM_00476670	RPL20 (Ribosomal Protein of the Large subunit #20)	200
TTHERM_00773330	RPS13 (Ribosomal Protein of the Small Subunit 13)	200
TTHERM_00550810	Hypothetical	196
TTHERM_00773350	NO ANNOTATION	196
TTHERM_00190950	ACT1 (ACTin)	190
TTHERM_00992730	RPL33 (Ribosomal Protein of the Large subunit #33)	190
TTHERM_00143660	HTA3 (Histone h Two A)	187
TTHERM_00721150	ribosomal protein L30p/L7e containing protein	186
TTHERM_00434070	RPS9 (Ribosomal Protein of the Small subunit 9)	163
TTHERM_00047480	RPS1 (Ribosomal Protein of the Small subunit 1)	162
TTHERM_00134940	RPL25 (Ribosomal Protein of the Large subunit #25)	161
TTHERM_00442270	Miz RING **BAIT**	159
TTHERM_00716240	RPL2 (Ribosomal Protein of the Large subunit 2)	152
TTHERM_00926980	acetyl-CoA acyltransferases family protein	151
TTHERM_00193800	CHD1 (Chromo-Helicase-DNA-binding domain)	149
TTHERM_00085180	RPL32 (Ribosomal Protein of the Large subunit #32)	148
TTHERM_00780960	RPL27 (Ribosomal Protein of the Large subunit #27)	144
TTHERM_01161040	TWI1 (Tetrahymena piWI)	144
TTHERM_00266390	RPS15 (Ribosomal Protein of the Small subunit 15)	138
TTHERM_00561490	hypothetical protein	136
TTHERM_00454080	RPS22 (Ribosomal Protein of the Small subunit 22)	135
TTHERM_00338210	RPL18 (Ribosomal Protein of the Large subunit #18)	134
TTHERM_00476640	Ribosomal protein L24e containing protein	130
TTHERM_00765300	RPS14 (Ribosomal Protein of the Small subunit 14)	129
TTHERM_00647339	hypothetical	124
TTHERM_01207720	hypothetical	123
TTHERM_00095540	CBF5 (Nucleolar protein CBF5)	122
TTHERM_00434040	RPS12 (Ribosomal Protein of the Small subunit)	121
TTHERM_00474970	60s Acidic ribosomal protein	113
TTHERM_00059330	AAA domain protein	112
TTHERM_00051840	hypothetical	111
TTHERM_00686220	RPS25 (Ribosomal Protein of the Small subunit 25)	109
TTHERM_01109770	RPS11 (Ribosomal Protein of the Small subunit 11)	107
TTHERM_00670170	ribosomal protein S9 containing protein	102
TTHERM_00929450	PGK1 (PhosphoGlycerate Kinase)	100