UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Development of a Targeted Chromatin Associated Protein Purification Method Identifies Novel Histone Gene Regulators

Permalink

https://escholarship.org/uc/item/7841n4zc

Author

Tsui, Chiahao Kevin

Publication Date

2017

Peer reviewed|Thesis/dissertation

**Development of a Targeted Chromatin Associated Protein Purification Method
Identifies Novel Histone Gene Regulators**

by

Chiahao Kevin Tsui

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge
Professor Robert Tjian, Chair
Professor Donald Rio
Assistant Professor Dirk Hockemeyer
Associate Professor Danica Chen

Spring 2017

Abstract

Development of a Targeted Chromatin Associated Protein Purification Method Identifies Novel Histone Gene Regulators

Chiahao Kevin Tsui

Chair of Committee:
Professor Robert Tjian
Department of Molecular and Cell Biology

Regulation of gene expression is fundamental to many biological processes such as lineage specification during embryonic development or metabolic responses to environmental cues. Classic *in vitro* biochemical experiments have been successful at identifying various components that are necessary for controlling gene expression, such as transcription factors, and co-activators as well as components of the machinery that regulate RNA processing, stability and degradation. However, detailed mechanistic understanding of gene regulatory mechanisms for many specific genes and gene networks remains unknown. Recently, many labs have attempted to couple mass spectrometry with purification methods that enrich for a specific genomic locus to identify potential novel regulators associated with specific chromosomal targets of interest. Here, I describe the development of a novel dCAS9 mediated purification method for proteins associated with specific cis-regulatory chromatin elements. First, I established the proof of principle by identifying specific human telomere associated proteins. I then further advanced this "reverse-ChIP" strategy by applying it to the *de novo* identification of novel regulators linked to the *Drosophila* histone gene locus.

Chapter 1 provides an overview of what is currently known and unknown about gene regulation at the *Drosophila* histone cluster as well as a summary of recent advances in the field of "reverse-ChIP" techniques. Chapter 2 describes my development of the dCAS9 mediated purification method for proteins specifically associated with targeted chromatin elements, including the purification and identification of proteins bound to human telomeres. Chapter 3 details how I've adapted the dCAS9 mediated purification method to identify novel regulators of the *Drosophila* histone cluster.

Table of Contents

Acknowledgements

I have been very fortunate to have wonderful people help me through the 6 years of my graduate career. My life would have been exceedingly hard if it weren't for the joy and support from everyone around me.

I would like to thank my advisor, Professor Robert Tjian, for allowing me to try what seems to be impossible projects and encouraging me when the inevitable failures happened.

I would also like to thank the members of my thesis committee, Professors Donald Rio, Dirk Hockemeyer, Danica Chen, and former committee member, Professor Michael Levine, for their help and guidance throughout the years.

I have also received tremendous advice and help from the members of the Tjian/Darzacq labs. I am especially grateful to Professor Xavier Darzacq, Dr. Claudia Cattoglio, Dr. Carla Inouye, Dr. Benjamin Guglielmi, Dr. Jaclyn Ho, Dr. Nikki Kong, Dr. Elisa Zhang, Dr. Sheila Teves, Gina Dailey, Mallory Haggart, Janeen Lim, and Dr. Shuang Zheng. I would also like to thank my classmates of MCB 2011 for all the experiences and memories we shared throughout the years.

Last but not least, I would like to thank my family for helping me along the way. I can be quite stubborn so their patience with me is hard to overstate. I would also like to thank Carolyn Elya, Nora, and Bruce Banner Elya-Tsui for being the cheese to my macaroni. Without you guys, life would be flavorless.

Chapter One:

Description of the *Drosophila* Histone Cluster's Unique Gene Expression and a Brief History of the Reverse ChIP Technique

Regulation of the *Drosophila* histone cluster

Gene expression is critical for cellular function and animal development. By altering transcription or translation, cells are able to react to environmental changes or generate new cell types (Levine et al., 2014; Sonenberg and Hinnebusch, 2009). Perhaps not surprisingly, gene expression is a highly complex process with multiple points of regulation. In eukaryotes, the initial point of regulation is the packaging of the genome into chromatin, whose basic organizing unit consists of 147 base pairs of DNA wrapped around histone octamers, forming the nucleosomal structure. Chromatin is organized into euchromatin or heterochromatin states as an initial step to regulate DNA accessibility with the former being more accessible to DNA binding factors than the latter (Luger et al., 1997; Richmond et al., 1984). Different sequence-specific transcription factors then dictate the expression of a defined set of genes in response to environmental or differentiation cues (Golldack et al., 2011; Lonze and Ginty, 2002). Once the mRNA is transcribed from its genomic template, RNA binding proteins (RNPs) or miRNAs can regulate the protein level by promoting binding of the ribosome or destabilizing the mRNA, respectively (Hammond et al., 2001; Hutvágner and Zamore, 2002; Pestova and Kolupaeva, 2002).

An interesting example of a highly controlled gene expression program is the regulation of the canonical histone genes. As eukaryotic cells progress through the cell cycle, the doubling of the DNA content requires the production of histone proteins, H1, H2A, H2B, H3, and H4, to quickly bind and wrap the newly synthesized DNA back into chromatin. Mirroring DNA replication, histone protein synthesis is a tightly regulated process where histone mRNA levels increase by 35-fold as the cell enters S phase and quickly degrade once this cell cycle step has completed (Harris et al., 1991). The strict maintenance of low free histone levels when there is no active DNA replication is necessary for proper cell health as excess histones lead to abnormal chromosomes and possible interference of histone methyltransferases/deacetylases (Gunjan and Verreault, 2003; Singh et al., 2010). Some aspects of how histone mRNA expression is controlled are known, however there are still many questions regarding both the transcriptional and post-transcriptional regulation that remain unanswered. In this section, I will summarize our present understanding of gene regulation in the *Drosophila* histone cluster.

Organization of the histone cluster

Histone genes are an interesting subset of eukaryotic genes: in every species there are multiple copies of each gene and they are physically clustered within the genome. From the sea urchin to human, clustered histone genes are expressed in a coordinated fashion. In mice and humans, there are 6 to 15 copies of each histone gene located on two different chromosomes, whereas in *Drosophila* the histone genes are in a 5 kilobase sequence that itself is repeated approximately 100 times (Lifton et al., 1976; Marzluff and Duronio, 2002).

In vivo, the histone cluster is visualized as a distinct entity within the nucleus. Initially the histone body was identified as being associated with the cajal body in mammalian cells using DNA hybridization probes. However, it was not until later that experiments in *Drosophila* clearly showed the histone gene cluster as a separate nuclear body with distinct protein compositions (Frey and Matera, 1995; Liu et al., 2006). Not much is understood about why the histone cluster forms a distinct nuclear body, but one leading hypothesis postulates that the histone cluster functions as a highly condensed area in order to increase the local concentration of regulatory factors to facilitate macromolecular interactions and reactions (Sawyer and Dundr, 2016). In line with this hypothesis, the proper recruitment of many histone cluster associated proteins in high concentrations to the histone nuclear body is necessary for 100% efficient histone mRNA processing and transcription termination (Tatomer et al., 2016).

Histone mRNA structure and function

Histone mRNAs themselves are structurally unique. Unlike other mRNAs in the metazoan genome, histone mRNAs do not contain a poly-A tail. Instead they contain two sequences encoding a stem loop structure and the histone downstream element (HDE), both of which are important for the post processing of histone pre-mRNA (Dominski and Marzluff, 1999). The HDE was discovered when researchers noticed the conservation of a purine rich region at the end of the pre-mRNA which exhibits significant complementation to the U7 small nuclear RNA (snRNA)(Mowry and Steitz, 1987). Subsequently, the function of the HDE was discovered through a series of elegant *in vitro* and *in vivo* experiments where Bond et al. found that mutations in the HDE disrupted proper histone pre-mRNA processing but could be rescued with complementary mutations in the 5' end of the U7 snRNA (Bond et al., 1991). Similarly, the stem loop was first discovered through sequence conservation and later found to be important for the proper 3' processing of the H2A mRNA through mutational analysis in sea urchins (Birchmeier et al., 1983; Hentschel and Birnstiel, 1981).

Through a series of biochemical fractionation and affinity purification studies, researchers found that the HDE recruits a specific complex of U7 small nuclear ribonucleoprotein (snRNP), containing LSM10 and LSM11 proteins (Pillai et al., 2001). U7 snRNP belongs to a family of snRNPs that participate in pre-mRNA splicing, however, it is a minor snRNP of approximately 1% of the major spliceosomal snRNPs (Dominski and Marzluff, 1999). The association of U7 snRNP to the HDE is capable of recruiting all other factors needed for proper 3' processing of histone mRNA (Spycher et al., 1994). However, stabilization of the U7 snRNP to the pre-mRNA requires the binding of the stem loop binding protein (SLBP) to the stem loop structure. SLBP was identified in 1997 and is crucial for multiple steps of histone mRNA metabolism (Fig 1) (Marzluff and Duronio, 2002). In addition to stabilizing the U7 snRNP for proper 3' end processing, SLBP remains bound to the mature mRNA and is involved in subsequent mRNA translation (Sànchez et al., 2002). Interestingly, the protein level of SLBP is tightly regulated in a manner similar to canonical histone mRNA expression. SLBP mRNA levels remains largely the same throughout the cell cycle, but protein levels increase dramatically right before and quickly

falls at the end of S phase (Whitfield et al., 2000). That SLBP expression is inversely correlated with histone mRNA expression is one reason to suspect that SLBP regulates histone mRNA degradation. Interestingly histone mRNAs are seemingly SLBP's only *in vivo* targets suggesting a very specific regulatory function (Townley-Tilson et al., 2006). How SLBP is mechanistically involved in histone mRNA degradation remains unknown. One current model proposes that SLBP protects histone mRNA by binding it in a stoichiometric manner, but it is also possible that SLBP is necessary for the proper recruitment of ribonucleases at the end of S phase to facilitate degradation.

Transcription regulation of the histone cluster

To rapidly synthesize histone proteins for S phase, there must be an equally rapid synthesis of histone mRNA at the onset of S phase. And although the other aspects of canonical histone mRNA regulation have been described at the mechanistic level, the transcriptional regulation of the histone genes remains unclear. By studying total mRNA levels, it was previously believed that the transcription of all canonical histone genes is coordinated throughout S phase (Harris et al., 1991). However, a recent study from our lab showed that there is a differential pattern of H1 transcription compared to the rest of the core histone genes (Guglielmi et al., 2013). Using a precise BrdU/ Edu pulse chase system coupled with RNA fluorescent *in situ* hybridization, Guglielmi et al. showed that H1 gene is transcribed throughout the entire S phase whereas H2A is only actively transcribed for the first two out of six hours. What is the specific mechanism that dictates this temporal transcriptional difference remains an unsolved question.

One difference between H1 and the core histone genes is the usage of alternate PIC components. Previous studies found that core histone genes use the prototypical TATA binding protein (TBP) for transcription whereas H1 promoter is bound by a related factor called TBP related factor 2 (TRF2) (Fig 1) (Isogai et al., 2007). However, this difference is not likely to be responsible for the temporal transcription difference between H1 and H2A as TBP is bound at the histone cluster even outside of S phase (Guglielmi et al., 2013). In addition, the precise transcriptional activation at the start of S phase hints at the involvement of a specific as opposed to a broad transcription activator that can be regulated in accordance with the cell cycle. Interestingly, immunofluorescence experiments also suggest that the PIC at the histone cluster lacks classic subunits such as TFIIB. These results bring up the possibility that a distinct PIC is used for the transcription of the canonical histone genes.

In mammalian cells OCT1 is proposed to be the DNA binding transcription factor that is responsible for initiating S phase specific transcription of the H2B gene (Fletcher et al., 1987; Zheng et al., 2003). This specificity is driven by a canonical octamer sequence element found within the mammalian H2B gene promoter. However, characterization of the *Drosophila* histone genes did not reveal any canonical octamer elements, making a direct connection between mammals and *Drosophila* difficult (Kremer and Hennig, 1990). Despite the lack of potential octamer sequences, Lee et al. were able to identify PDM-1, a protein with a homologous DNA binding domain to OCT1, as a

possible binding and regulator of core histone genes (Fig 1) (Lee et al., 2010). However, the hypothesis that PDM-1 acts as the main regulator of S phase specific histone gene transcription is an imperfect explanation since PDM-1 is a ubiquitous transcription factor that has no discernible cell cycle based regulation.

Although not a DNA binding protein, NPAT gained prominence as a transcription regulator of canonical histone genes in the last 17 years. NPAT's role in histone gene transcription was initially discovered in 2000 when overexpression and mutant analysis demonstrated its ability to affect histone gene activation (Zhao et al., 2000). Precise mutations of NPAT also showed that histone gene transcription is diminished when it can no longer be phosphorylated by cyclin E/ Cdk2 (Ma et al., 2000). As cyclin E/ Cdk2 are well known regulators for entry into S phase, the activity of NPAT being dependent on its phosphorylation conveniently places it as an activator that can respond to cell cycle signals (Hindley and Philpott, 2012). The identity of the NPAT ortholog in *Drosophila* remained a mystery for over a decade until a double stranded RNA library screen for histone gene regulators identified Multi Sex Combs (MXC). The homology of the MXC's LisH domain to NPAT suggested that MXC is the *Drosophila* ortholog to NPAT (White et al., 2011). However, as hinted by the non-DNA binding characteristic of NPAT and MXC, further studies into the role of MXC using knockdowns and immunofluorescence assays for other histone cluster components suggest that it is key for the recruitment of pre-mRNA processing subunits but its presence at the histone cluster does not dictate active transcription (Fig 1) (Guglielmi et al., 2013; White et al., 2011).

While the mechanism for activating transcription at the *Drosophila* histone cluster remains murky, the process of repression outside of S phase is slightly clearer. During a genetic screen for transcription coregulators, Ito et al. discovered that the HERS protein binds to the histone cluster on polytene chromosomes (Ito et al., 2012). By performing double immunofluorescence against HERS and a marker for active histone gene transcription, they were able to show that HERS binds specifically to the histone cluster in late S phase and shuts down the expression of the entire histone cluster through its interaction with the Su(var)3-9/ HP1 repressor complex (Fig 1) (Ito et al., 2012).

Reverse-ChIP

The development of genome-wide techniques such as Chromatin Immunoprecipitation coupled with DNA sequencing (ChIP-seq) has been extremely useful for biologists. Using a high affinity antibody against a protein of interest, ChIP-seq allows for the identification of all potential DNA binding sites for a given protein within the genome. With a list of all genomic locations where the protein of interest is bound, researchers can generate and test hypotheses regarding regulatory functions at enriched sequences and in doing so begin to understand the protein's overall role within the cell (Chen et al., 2008). However, eukaryotic gene expression results from the coordination of tens if not hundreds of proteins working in concert to ensure proper gene regulation (Taatjes et al., 2004). Finding available and highly specific antibodies for each individual protein has been challenging and remains a roadblock to our

understanding of many genomic factors. Furthermore, the ChIP-seq approach requires prior knowledge that the protein of interest has regulatory functions within the nucleus. These challenges have made the discovery of the complete protein composition at specific genomic locus a difficult and slowly advancing process.

To address this issue, there have been multiple attempts over the last 30 years to isolate specific regions of chromatin and to characterize the proteins associated with them (Griesenbeck et al., 2003; Jasinskas and Hamkalo, 1999; Vincenz et al., 1991; Workman and Langmore, 1985). However, it was not until the last ten years that mass spectrometry analysis became sensitive enough to detect and obtain a whole proteome view of what binds to a specific purified genomic region (Walther and Mann, 2010; Wierer and Mann, 2016). We refer to this general strategy as a "reverse-ChIP" since the method is designed to fish out a specific DNA sequence in order to identify all the proteins associated with the locus which is, essentially the opposite approach to the conventional ChIP-seq.

PICh, TChP, and other published methods

The first successful reverse-ChIP technique was described in 2009 (Déjardin and Kingston, 2009). In this study, the authors used a nucleic acid hybridization method to specifically enrich for their target of interest. To maximize their signal to noise ratio, the authors chose to purify telomeric chromatin in mammalian cells as the repetitive telomeric sequence represents ~0.01 to 0.07% of the genome. By coupling a biotin analog to a Locked Nucleic Acid DNA hybridization probe and introducing it to crosslinked chromatin, they were able to enrich for telomeric sequences and perform mass spectrometry analysis. This technique was successful in purifying and identifying proteins that have been previously associated with telomeres *in vivo*, such as Shelterin components. In addition, the authors were able to identify new telomere associated proteins including Homeobox containing protein 1, a putative transcription factor. Despite its initial success, the authors acknowledge that they were unable to identify many known telomere associated proteins such as Rad51D and Tankyrase 1.

After Dejardin & Kingston's report in 2009, two other groups published a reverse-ChIP technique using similar set ups. Fujita and Fujii generated a transgenic cell line harboring 16 repeats of a LexA binding sequence next to their target of interest, the HS4 insulator sequence (Fujita and Fujii, 2011; Kim and Little, 1992). After overexpressing a LexA DNA sequence binding protein with a FLAG peptide tag, the authors crosslinked the chromatin and were able to specifically isolate proteins binding to the HS4 insulator sequence using an anti-FLAG antibody. Although they were unable to find proteins that were previously identified components of insulator complexes such as CTCF, they managed to detect p68, a reported component of the *gypsy* insulator in *D. melanogaster* (Lei and Corces, 2006; Yao et al., 2010). Pourfarzad et al. took a similar approach of introducing exogenous DNA sequences and corresponding DNA binding proteins as a method to isolate a specific chromatin of interest. After inserting TetO sites next to their target region, the authors overexpressed a modified TetR protein and performed a two-step purification to enrich for the γ-globin gene. Pourfarzad et al. were successful in identifying a number of proteins that bind to the γ-globin locus

via ChIP and found a transcriptional repressor, ZBP-89, that leads to the reactivation of γ-globin upon knockdown (Pourfarzad et al., 2013).

Owing to the various shortcomings of these methods and the herculean effort required to attempt the aforementioned reverse-ChIP experiments, experimental approaches were constantly being revamped in order to develop methods that are less expensive, more robust, and easily modifiable to target new regions of interest. The recent biochemical characterization of the RNA reprogrammable DNA binding protein, CAS9, has demonstrated that it is a versatile protein that could fit this bill (Jinek et al., 2012). However, published methods using CAS9 as the basis for specific chromatin purification has not yet yielded any promising results. For example, even though targeting of the *GAL1* locus in yeast revealed multiple proteins related to transcription, proteins known to be essential for the activation of *GAL1* such as GAL4 and RNA Polymerase II was not found (Waldrip et al., 2014). Similarly, the purification of the *IRF1* promoter by a CAS9 mediated method was able to identify a number of nuclear associated proteins, but again no obvious candidates for the activation of *IRF1* was found (Fujita and Fujii, 2013).

Concluding remarks

How gene expression is regulated is a long standing question that dates back to the 1960s when Monod and his colleagues provided the first clear mechanism for the activation of the Lac operon (Jacob and Monod, 1961). Since then, molecular and biochemical techniques have yielded tremendous insight into both the components necessary for transcription and how accurate gene expression patterns can be set up using a combination of regulatory sequences and sequence-specific transcription factors (Kadonaga et al., 1987; Small et al., 1992). Despite our best efforts, there are still many interesting gene regulation questions not yet solved, such as the control of the *Drosophila* Histone Cluster gene expression. In the next two chapters I describe my efforts in developing a new reverse-ChIP method and then adapting it to discover novel regulators of *Drosophila* histone gene expression.

Figures
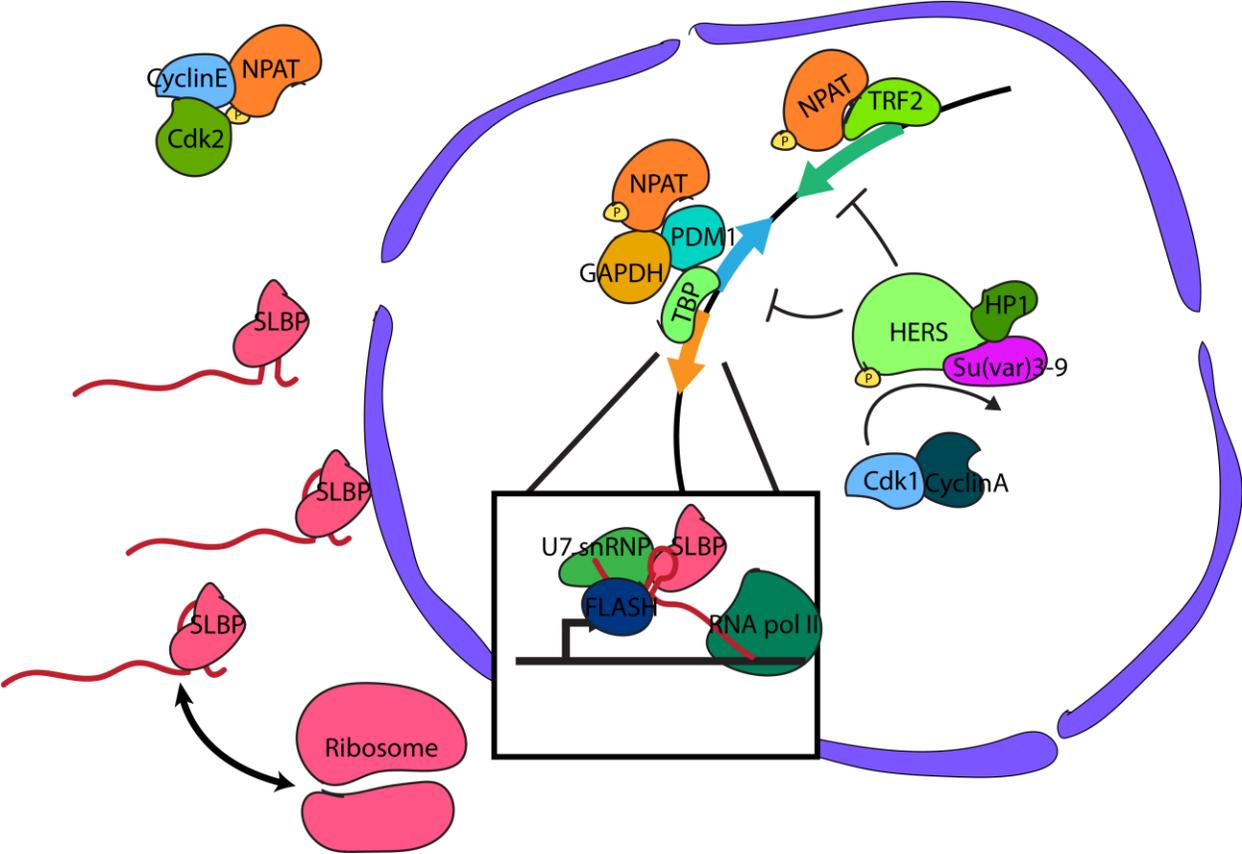
Figure 1. A graphical representation of the *Drosophila* Histone Cluster regulations.

Chapter Two:

Development of an *in vitro* dCAS9 Mediated Chromatin Associated Protein Purification Method

Abstract

Eukaryotic gene regulation is a complex process necessitating tens to hundreds of coordinating gene products. Although past research has been able to identify basal machinery and other broad essential factors involved in gene regulation, numerous gene specific co-factors and activators remain undiscovered. To uncover gene specific regulators, researchers have been working to purify specific chromatin fragments and identify proteins associated with them. Here I describe my work to develop a sequence-specific *in vitro* chromatin purification method using dCAS9 that is inexpensive and amenable to quickly switching chromatin targets. Using this method, I have successfully enriched and identified proteins previously found to be associated with telomere sequences. Interestingly, I have consistently identified TEL2 at telomere sequences. This finding suggests TEL2 has a role for telomere function similar to the yeast despite previous TEL2 knockdown studies in mouse and human cells having shown no discernible telomere maintenance effects.

Introduction

Chromatin associated proteins and DNA binding factors play a critical role in all aspects of biological responses. However, our ability to analyze these factors and their *in vivo* functions on chromatin typically require two criteria: a) *a priori* knowledge of the protein and b) a highly specific and suitable antibody. With whole genome sequencing and homology identification of proteins across species, the former is often not an issue. Still, it relies on initial protein identification through biochemical or genetic screens. Even after initial characterization, trying to ascertain novel function of a known protein is akin to making an educated guess. As for the second criterion, obtaining the proper antibody for an experiment such as ChIP-seq is difficult enough for the majority of proteins that researchers have dedicated entire methods to work around this issue (van Steensel et al., 2001; Vogel et al., 2007).

In contrast, studying a particular genomic locus only requires knowledge of the DNA sequence of interest. However, identifying the proteins associated with the locus requires highly challenging purification and protein identification. While multiple labs have been trying to purify specific chromatin sequences attached to associated proteins for more than 30 years, the majority of the efforts to identify novel locus-specific proteins have not been successful due to minimal yields and the lack of robust and sensitive protein identification technologies (Griesenbeck et al., 2003; Jasinskas and Hamkalo, 1999; Vincenz et al., 1991; Workman and Langmore, 1985). Recently, advances in mass spectrometry and the continued development of purification methods have allowed researchers to begin identifying proteins from isolated chromatin (Wierer and Mann, 2016). For example, Dejardin and Kingston successfully used a DNA hybridization probe to isolate and enrich for proteins associated to telomere sequences in HeLa cells (Déjardin and Kingston, 2009). Using a combination of mass spectrometry and immunofluorescence, they were able to link Homeobox containing protein 1 (HMBOX1) as a novel telomere associating protein even though the functional role of HMBOX1 was unknown at the time. Other successful reverse-ChIP publications also include an example where the researchers introduced multiple TetO sites into the γ-globin locus, and they successfully used the overexpression of tagged TetR protein to specifically pull down the chromatin of interest and identified associated proteins (Pourfarzad et al., 2013).

While reverse-ChIP experiments are beginning to be technologically feasible, the few methods that have been published suffers from various flaws such as inflexibility, sensitivity and cost. Thus, we set out to develop a novel chromatin purification method using the RNA programmable DNA binding protein, CAS9, that will allow us to easily enrich various genomic targets and identify the associated proteins.

Results

Development of the *in vitro* dCas9 pull down system

When it was first published in 2012 as a RNA guided DNA binding protein, CAS9 endonuclease's potential in applications such as genome editing was immediately recognized (Jinek et al., 2012, 2013; Mali et al., 2013). In addition, the D10A/ H840A mutant of CAS9 that abolishes its nuclease activity (dCAS9) also proved to be a versatile tool in imaging and chromatin modification when fused with a proper subunit (Chen et al., 2013; Deng et al., 2015; Gilbert et al., 2013; Hilton et al., 2015). So when the extremely stable *in vitro* binding of dCAS9 to its DNA target was described we believed that it might be possible to expand dCAS9's increasing versatility by turning it into a DNA targeting protein for an *in vitro* chromatin purification scheme (Sternberg et al., 2014). We set out to develop a reverse ChIP method using dCAS9 in a similar fashion as the PICh protocol in order to take advantage of the ease of switching dCAS9 targeting via a single guide RNA (sgRNA) sequence (Déjardin and Kingston, 2009). A key difference that proved critical for our approach was the use of in vitro purified RNPs bearing specific guide RNAs as the primary instrument for efficient and robust site specific targeting of selected chromatin sequences.

The overall scheme of the purification is to: 1. Crosslink the chromatin in the cells of interest with crosslinkers such as formaldehyde. 2. Isolate and shear the chromatin to the appropriate size. 3. Add purified recombinant dCAS9/sgRNA fused to a tag. 4. Enrich for the RNP and the targeted chromatin site by pulling down the tag and wash away non-specific chromatin fragments. 5. Isolate DNA and/or protein for identification (FIG 1). Totest the selectivity and effectiveness of our strategy, we performed a proof of principle experiment by targeting telomere sequences in HeLa cells. These target sites had been well characterized for binding proteins such as the shelterin complex and these relatively abundant sites represent a significant portion (0.01% to 0.07%) of the genome (Déjardin and Kingston, 2009; Grolimund et al., 2013).

In order to isolate and enrich for the dCAS9 RNP and any chromatin it might be binding, we initially chose the HALO protein domain as our tag as it provides a covalent bond that would allow for stringent washes (Los et al., 2008). 6x His-dCAS9-HALO (hereby referred to as dCas9-HALO) fusion protein was inserted into a pET vector for E. coli overexpression and the recombinant protein is induced at 18C overnight at 0.3mM IPTG. The recombinant protein is isolated and purified with Ni-NTA and then a POROS HS20 strong cation exchange chromatography (FIG 2a, 2b). Once purified to approximately 90% homogeneity, we tested the function of the dCAS9-HALO fusion protein by assessing its ability to bind to a short target DNA oligo labeled with Cy5 or nonspecifically to an off target DNA oligo labeled with AF488 fluorephore. As expected, when a dCAS9-HALO fusion protein binding to the on target sgRNA is added to a mixture of specific and non-specific DNA oligos and then sequestered by a HALO-ligand resin the target DNA is bound and depleted from the solution quickly as assessed by fluorescence while the off target DNA oligo remains largely unchanged (FIG 3A). This result suggests that the dCAS9-HALO fusion protein remains functional.

Although dCAS9 has been shown to bind quickly and specifically to a double stranded DNA oligo, what was unknown is whether or not it is capable of doing the same to a chromatin fragment that has been crosslinked with formaldehyde (Sternberg et al., 2014). To test this possibility, we generated HeLa cell pellets that are fixed with formaldehyde at 1% for thirty minutes, sheared with a sonicator, and incubated the soluble chromatin with dCAS9-HALO RNP with either telomere sequence-specific sgRNA or a nonspecific sgRNA targeting a random Lambda phage sequence. After isolating the RNP and the associated DNA, we used a telomere sequence oligo radiolabeled with p32 to gauge the amount of target DNA isolated. Surprisingly, dCAS9-HALO, when loaded with a telomere targeting sgRNA, is capable of enriching for telomere DNA even after the DNA has been extensively crosslinked with formaldehyde (FIG 3B). And using a radiolabeled oligo to probe for ALU SINE sequences as background contamination, we can see that the telomere DNA enrichment is specific and comparable to other methods that isolate telomere sequences by utilizing antibodies against various shelterin components (Grolimund et al., 2013).

To see if telomere associated proteins are indeed enriched we performed a telomere specific and non-specific pull down and used a silver stain to look for distinct bands that might suggest enriched proteins. Perhaps not entirely surprising, there were no discernible unique protein bands in the telomere specific pull down compared to the non-specific control despite the high amount of telomere DNA enrichment (FIG 4a). One possible reason for this is that most of the protein bands visible in the bulk silver stain gels represent non-specific background (likely highly abundant) proteins brought down by the resin itself. However, blocking the resin with known purified proteins such as BSA or insulin did not significantly diminish the resin's ability to non-specifically bind to various proteins in the HeLa chromatin sample. In order to decrease the non-specific background, we decided to try a desthiobiotin/ biotin elution system utilized in the PICh protocol (Déjardin and Kingston, 2009). By using a HALO ligand covalently linked to a desthiobiotin molecule, we should be able to pull down the chromatin/RNP complex with streptavidin linked resin with high affinity and specifically elute the chromatin/RNP complex using biotin as a competitor (Hirsch et al., 2002). Surprisingly, once the desthiobiotin/ HALO ligand molecule binds to the dCAS9-HALO it no longer binds to streptavidin on a membrane or in solution despite previously published literature utilizing a similar setup (FIG 4B, data not shown) (So et al., 2008). This result is even more surprising as it does not seem to be a result of the desthiobiotin ligand being hidden away in the dCAS9-HALO fusion protein as it is still capable of being recognized by the anti-biotin antibody (FIG 4B).

As the desthiobiotin elution strategy proved to be ineffective, we next turned TEV proteolytic cleavage, which is another elution method commonly used in tandem affinity purification systems (Kaiser et al 2008). We introduced a TEV cleavage site between dCas9 and HALO and again purified it to ~90% homogeneity using Ni-NTA followed by strong cation exchange chromatography (FIG 5A). TEV protease cleavage occurs efficiently in solution, going to completion in less than one hour at 37C (FIG 5B). However, the HALO ligand resin used to bind dCAS9-HALO seems to impede TEV cleavage as it does not go to completion on resin even after overnight incubation. Efficient elution of the cleaved dCAS9 protein from resin also requires some form of detergent otherwise it

retains nonspecifically (FIG 5C). After optimizing TEV elution, we verified that the dCAS9-TEV-HALO protein is capable of enriching for telomere DNA from fixed and sheared HeLa chromatin and that TEV elution also elutes the DNA (FIG 6A).

At this point we prepared a large sample of 250 million HeLa cells and performed the dCAS9 pull down using telomere targeting sgRNAs and non-targeting sgRNAs, isolated both DNA and protein, and assessed for enrichment. As expected, the dCAS9 mediated pull down highly enriched for the telomere DNA sequence (FIG 6B). To assess whether we're specifically enriching for telomere associated proteins, we used an antibody raised against TPP1, a component of the Shelterin complex. To our delight, the dCAS9 pull down loaded with telomere sgRNA was indeed able to specifically enrich for TPP1 protein when compared to the non-specific control (FIG 6B). However, there are still some issues that needed to be resolved including a low TPP1 signal that also comes down with the non-specific sample, suggesting the high background despite TEV elution. And although we were isolating approximately 4-7% of all available telomere DNA as assessed by radioactive DNA dot blots, we were only enriching approximately 0.1% of all available TPP1. This result suggests that most of the telomere DNA fragments we are isolating might have been stripped of TPP1 and were naked double stranded DNA since previous studies have suggested that as much as 50% of all TPP1 proteins within the cell are bound to telomeres (Takai et al., 2010). A previous chromatin purification study suggested that purification schemes could be flawed in enriching for large chromatin structures (Griesenbeck et al., 2003). To test whether our dCAS9-TEV-HALO-resin pull down setup suffers from a similar issue, we built a six kilobase plasmid with a 9x telomere TTAGGG repeat to see if our pull down method can efficiently isolate this large molecular weight structure. As expected, the dCAS9-HALO RNP loaded with telomere specific sgRNA was able to deplete the telomere sequence containing plasmid but not the non-specific sgRNA dCAS9-HALO RNP (FIG 6C). However, to our surprise the dCAS9-HALO RNP elution sample did not show an enrichment of the telomere sequence plasmid suggesting that while it can bind the target sequence, the plasmid is lost during subsequent washing steps (FIG 6C).

If dCAS9-TEV-HALO cannot sustain binding to large molecular weight structures, such as a plasmid, it might also indicate that dCAS9-TEV-HALO is preferentially binding to small, possibly protein free, DNA fragments that are generated when the chromatin is mechanically sheared. This might explain why there is such a high percentage of total telomere DNA enrichment but a much smaller percentage of TPP1 isolated despite being part of the shelterin complex that is proposed to bind along the entire telomere end (Palm and de Lange, 2008).

When Griesenbeck et al. found the retention issue regarding size of the fragment, they solved it by using a longer adaptor for interaction to the resin. To see if something similar can help, we moved to using a 6x His-dCAS9-3x FLAG (dCAS9-FLAG) system coupled with the anti-FLAG agrose resin. Again, dCAS9-FLAG was purified to ~90% homogeneity using Ni-NTA followed by POROS HS20 strong cation exchange chromatography (FIG 7A). And when used to pull down telomere sequence containing plasmid, the dCAS9-FLAG method was able to successfully enrich and elute the plasmid (FIG 7B). As expected, the dCAS9-FLAG setup is also able to successfully enrich for

TPP1 protein when the dCAS9 RNP is loaded with telomere specific sgRNA (FIG 8A). In addition, by titrating the amount of dCAS9-FLAG used we were able to specifically enrich up to 1% of total TPP1 in the lysate without increasing background, which is a significant improvement over the dCAS9-TEV-HALO method (FIG 8A). In addition to TPP1, five other proteins make the shelterin complex (Palm and de Lange, 2008). And after obtaining antibodies for two more shelterin components, TRF2 and POT1, we were able to show that they are also specifically enriched in the telomere targeted dCAS9-FLAG pull down.

Proof of principle with HeLa telomere sequence purification

Encouraged by the enrichment of specific shelterin components, we prepared telomere specific and non-specific pull down samples using 500 million HeLa cells each. We then collaborated with the Washburn group at the Stowers Institute to perform MudPIT mass spectrometry to identify the proteins within each sample. Multiple replicates were submitted for mass spectrometry analysis and a representative diagram of the number of proteins found and sample overlap from one preparation is shown in FIG 8C. Overall there are approximately 1300 proteins identified within each sample, with the majority of them overlapping between the specific and non-specific samples (FIG 8C). We categorized proteins as being enriched in the telomere specific pull down by grouping proteins that are only found within the telomere specific sample with those that are found within both samples but have a distributed normalized spectral abundance factor (dNSAF) ratio of greater than 1.5. Gene ontology (GO) term analysis with the DAVID bioinformatics database of the telomere enriched protein list identified a number of proteins associated with telomere function, including shelterin complex proteins such as TRF2 and RAP1 (FIG 8D) (Huang et al., 2009a, 2009b). Surprisingly, TPP1 was not identified during the initial mass spectrometry analysis despite being highly enriched by western blot analysis (FIG 8B). One possible reason for this discrepancy could be the particular protease pair used for the mass spectrometry proteolytic cleavage do not generate suitable peptide sequences for peptide ionization and detection. Indeed, when a different protease pair was used for proteolytic cleavage, TPP1 was readily identified.

Discussion

A key step to understanding the molecular mechanism of transcription and post-transcription regulation for a specific gene is to identify all the major players involved in the process. And while incredible progress has been made in identifying basal machinery, overall important transcription factors/ co-factors, and major chromatin remodelers, there is still a large gap of knowledge on the molecular mechanism of regulation at a majority of genes (Thomas and Chiang, 2006). Multiple publications have tried to tackle this problem by coupling some form of chromatin purification to mass spectrometry (Byrum et al., 2012; Hamperl et al., 2014; Pourfarzad et al., 2013; Unnikrishnan et al., 2010; Waldrip et al., 2014). However, these methods often require substantial investment in building transgenic cell lines that are tricky to perfect and most still do not result in the identification of specific regulators. The most successful of these methods seems to be PICh, which was originally developed in the Kingston lab (Déjardin and Kingston, 2009). But the protocol suffers from the high cost of essential materials and the need to substantially change the protocol when targeting different loci (Antao et al., 2012; Ide and Dejardin, 2015). With no cost efficient and effective method available, we set out to develop a chromatin purification protocol that is easy to perform, practical for targeting a large number of sequences, and easily adaptable to different cell types.

CAS9 has turned into the Swiss Army knife of molecular biology since its initial biochemical characterization in 2012, and we have added to the list of uses for CAS9 by successfully adopting it into an *in vitro* chromatin purification method (FIG 1) (Jinek et al., 2012). As a proof of principle, we showed that dCAS9-FLAG is capable of enriching for the human telomere region for both the DNA and the proteins associated with it (FIG 6B, 8B). Combined with MudPIT mass spectrometry to identify the proteome of the enriched sample, we were able to identify various proteins previously described to associate to telomere sequences (FIG 8D).

By being able to verify which proteins are physically associated with telomere sequences *in vivo*, we can better elucidate the mechanism of how they might function. For example, DNA-dependent protein kinase catalytic subunit (DNA-PKcs) was identified to be enriched in our telomere pull down samples and it was previously only indirectly associated with telomere function (Chai et al., 2002; Gauthier et al., 2012; Sui et al., 2015; Zhou et al., 2013). By knowing that it physically associates with telomere sequences, researchers can start to ask mechanistic questions such as where on the telomere sequence does it bind, if the recruitment is mediated by protein-protein, protein-DNA, and/or protein-RNA interactions, and what proteins besides hnRNP-A1, a previously described target, can it phosphorylate at the telomere? (Davis et al., 2014; Ting et al., 2009). PURα is another protein identified to be enriched in our telomere pull down (FIG 8D). Best known as a nucleic acid binding protein with DNA-unwinding activity, PURα's interaction serves as additional evidence that it could be serving an *in vivo* functional role at telomeres as previous publications have only shown PURα affinity to telomere sequences *in vitro* (Gallia et al., 2000; Im and Lee, 2005; Wortman et al., 2005).

An especially interesting protein found to be enriched in our telomere pull down is TEL2 (FIG 8D). Although it was first identified in yeast as a telomere length maintenance protein, recent studies have uncovered an additional role for TEL2 as an essential stabilizer for phosphatidylinositol 3-kinase-related kinases (PIKKs) (Hořejší et al., 2010; Lustig and Petes, 1986; Takai et al., 2007). The work that led to the discovery of TEL2 stabilizing PIKKs started with an attempt to understand its telomere related function in human and mouse cells with assays such as immunofluorescence, ChIP, and knockdown. However, they did not find any significant effect on telomere length homeostasis (Takai et al., 2007). Their results bring into question whether mammalian TEL2 has a functional role at telomeres at all, but our observation that TEL2 is physically associated to telomere sequences *in vivo* suggests that, like in S. cerevisiae, TEL2 has a role at mammalian telomeres. It is possible that it is harder to uncover TEL2's role in mammalian telomere maintenance due to redundant pathways that have evolved since the divergence from the budding yeast. Perhaps one way to solve this issue is to sensitize the system by generating telomere maintenance specific TEL2 mutations and couple that to knockdowns of other telomere maintenance proteins to look for synthetically enhanced telomere dysfunction (Rozario and Siede, 2012).

In summary, we developed a cost effective method of purifying chromatin from a specific region of interest. We were successful in purifying the telomere sequence from HeLa cells as a proof of concept, and we were able to identify multiple telomere associated proteins, including TEL2, suggesting that it serves a functional role in human telomeres much like the yeast homolog.

Figure 1.
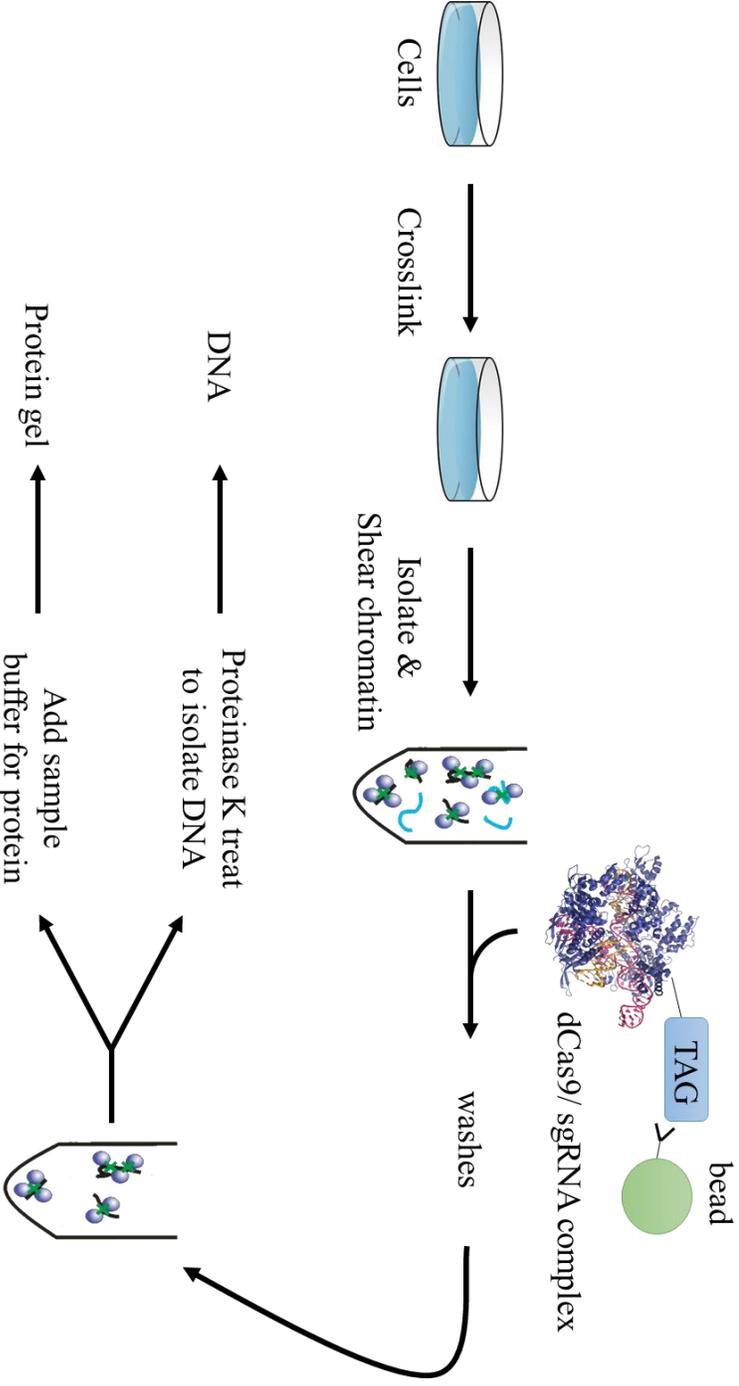
Figure 1. Conceptual layout of the dCas9 based chromatin purification method.
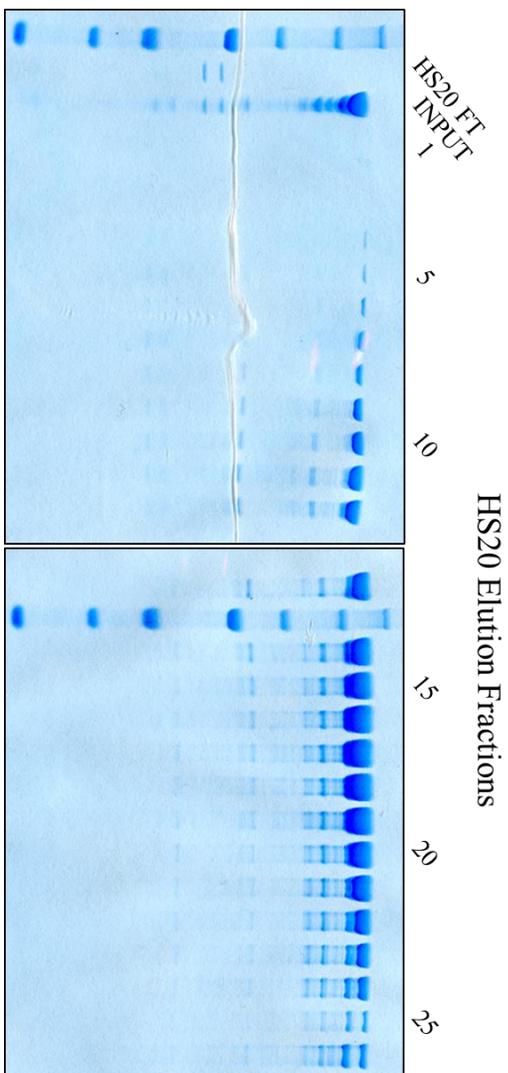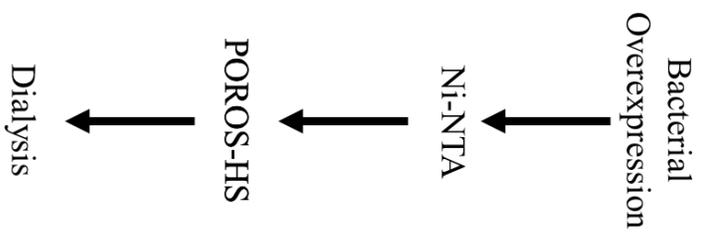
Figure 2.

Figure 2. Purification scheme and dCAS9-HALO elution fractions from the POROS HS20 column. A. 6x His-dCAS9-HALO was driven in a pET302 vector and transformed into BL21-Codon Plus RIPL competent cells (Agilent). Cultures are induced at 18°C overnight at 0.3mM IPTG, lysed, purified with Ni-NTA, and select fractions are further purified with POROS HS20 column. 6x His-dCAS9-HALO enriched fractions are collected, dialyzed, and flash frozen. **B.** Elution fractions of the major products from the POROS HS20 purification visualized by coomassie stain.

Figure 3.

A

B
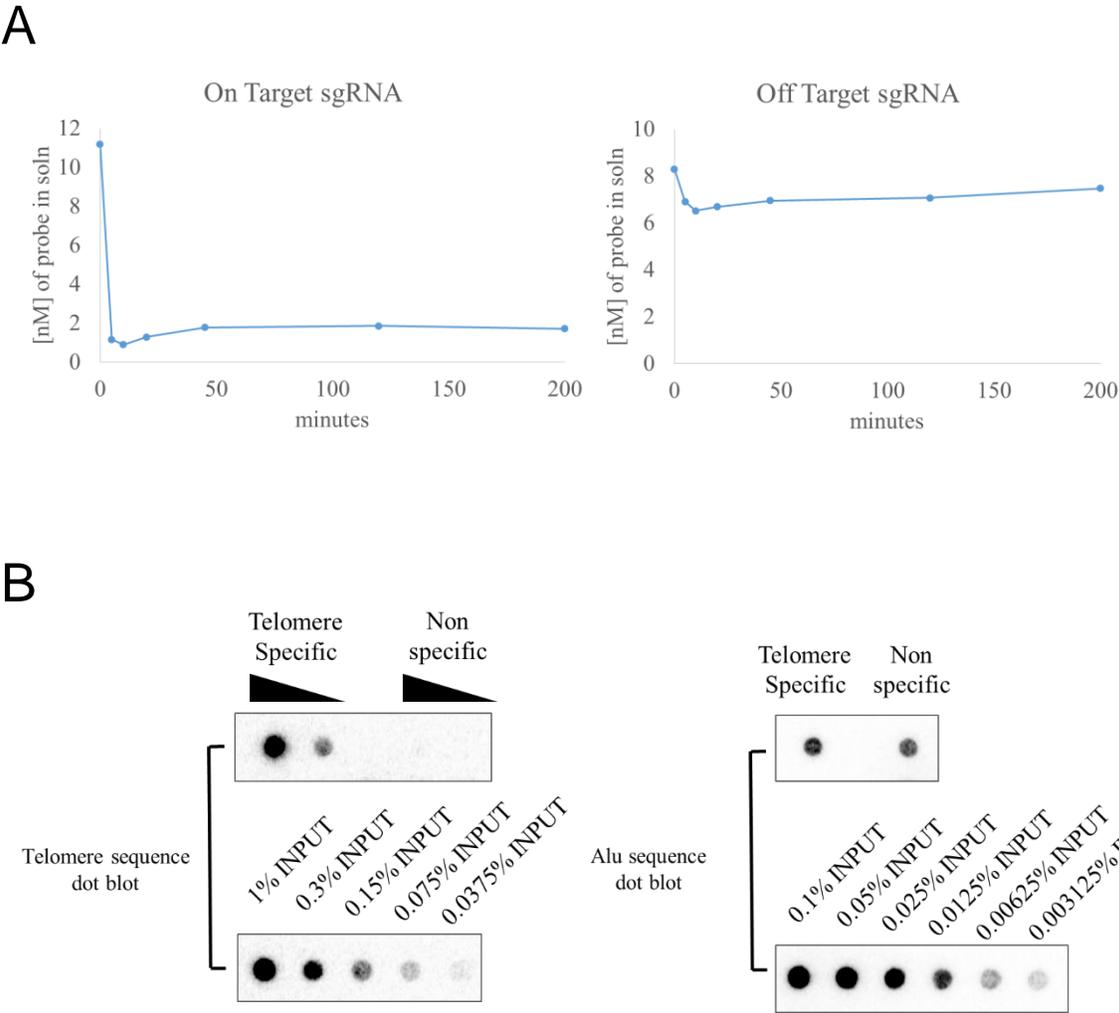
Figure 3. Specificity of dCAS9-HALO and it is able to enrich for crosslinked chromatin. A. dCAS9-HALO bound to a specific sgRNA is incubated with an targeted dsDNA labeled with Cy5 and off target dsDNA labeled with AlexaFluor488. dCAS9-HALO is precipitated from solution with HALO-ligand resin and the supernatant is measured for Cy5 and AF488 signal over time. **B.** HeLa chromatin crosslinked with formaldehyde is sheared and incubated with dCAS9-HALO bound to sgRNA targeting telomere sequences overnight. dCAS9-HALO is isolated via HALO-ligand resin and DNA is isolated and detected with radioactive probes complementary to telomere sequences or ALU DNA sequences.

Figure 4.

A

Telomere Specific / Non-specific / Telomere Specific / Non-specific

15min incubation

30min incubation

B

Blocking agent

none / BSA / Insulin

C

HALO-desthiobiotin / HALO-desthiobiotin / DMSO / DMSO / Biotin-Cys4 / Promega HALO ligand / HALO-desthiobiotin + HALO protein / Promega HALO ligand + HALO protein

Amount: + ++ + ++

HALO-desthiobiotin / HALO-desthiobiotin / DMSO / DMSO / Biotin-Cys4 / Promega HALO ligand / HALO-desthiobiotin + HALO protein / Promega HALO ligand + HALO protein

Amount: + ++ + ++

Streptavidin-HRP

Goat α biotin antibody

Figure 4. HALO-ligand resin has large non-specific interaction with proteins and blocking with purified proteins do not help. dCAS9-HALO does not interact with streptavidin when a HALO-ligand coupled to desthiobiotin reagent is used. A. dCAS9-HALO bound to telomere targeting or non-specific sgRNA was incubated with sheared HeLa chromatin and isolated by HALO-ligand resin. The isolated mixture was boiled in 1x sample buffer and a silver stain was performed to look for overall protein composition. **B.** Purified BSA and insulin was used as blocking reagents for the HALO-ligand resin. Silver stains were performed to gauge resin's binding to non-specific proteins in sheared HeLa chromatin. **C.** HALO ligand-desthiobiotin conjugate, commercial HALO ligand-biotin, and DMSO was incubated with dCAS9-HALO or commercial HALO protein. Labeling of the various proteins with desthiobiotin or biotin was tested with a western blot assay using either streptavidin-HRP or a goat α biotin antibody. Cys4 nonspecifically labeled with biotin was used as a positive control.

Figure 5.
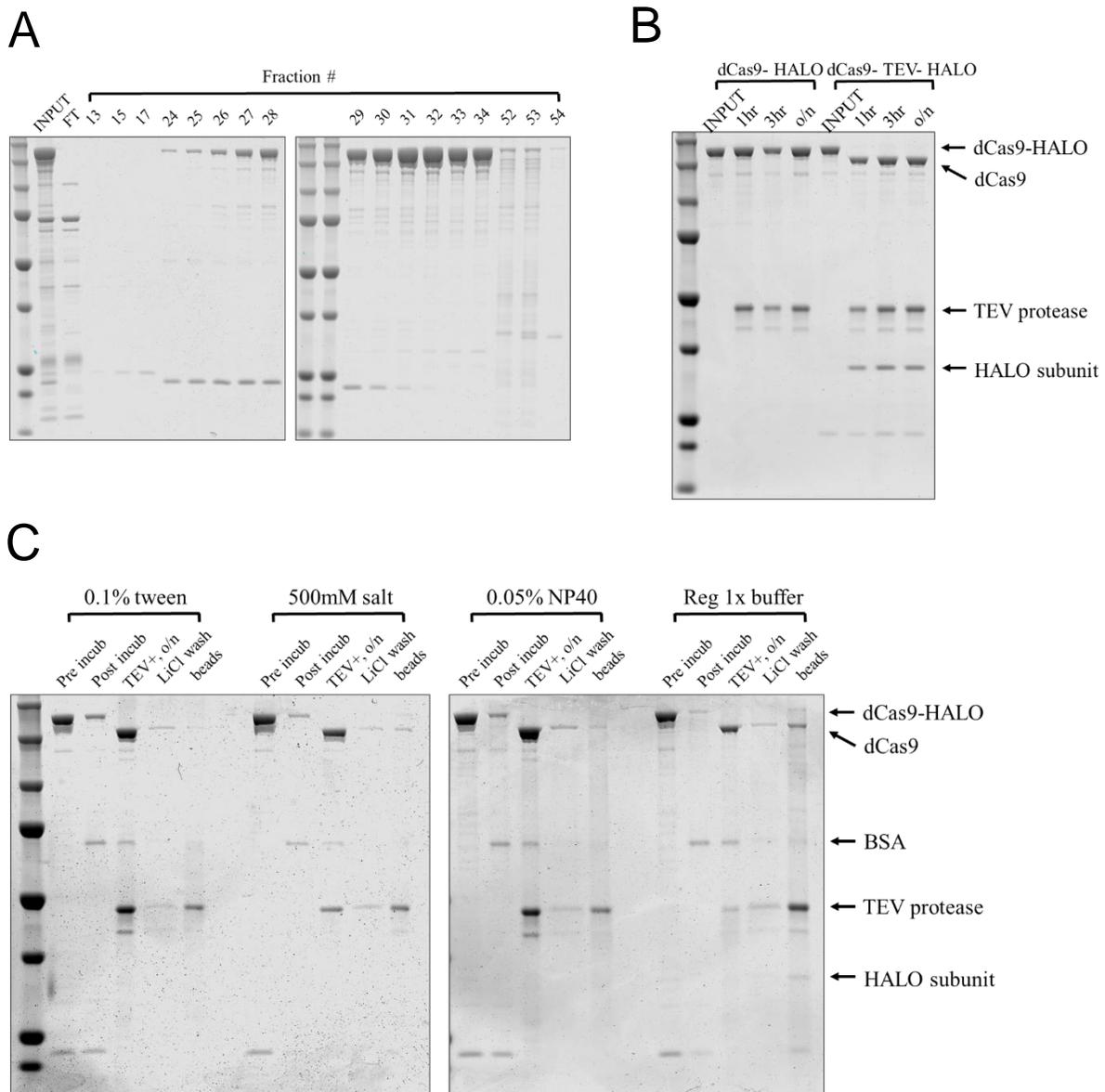
Figure 5. Purification of dCAS9-TEV-HALO, verification of TEV cleavage, and optimizing the cleavage elution of dCAS9 from HALO ligand resin. A. POROS HS20 elution fractions of dCAS9-TEV-HALO visualized by coomassie stain. **B.** dCAS9-TEV-HALO cleavage by TEV protease is tested in 1x CAS9 buffer. **C.** dCAS9-TEV-HALO is bound to HALO ligand resin and then incubated in 1x CAS9 buffer with TEV protease for cleavage overnight. Different buffer conditions such as the addition of 0.1% tween 20, 500mM NaCl, or 0.05% NP40 is tested for effect on dCAS9 release. Cleavage and release of the dCAS9 is assessed via coomassie stain.

Figure 6.

A
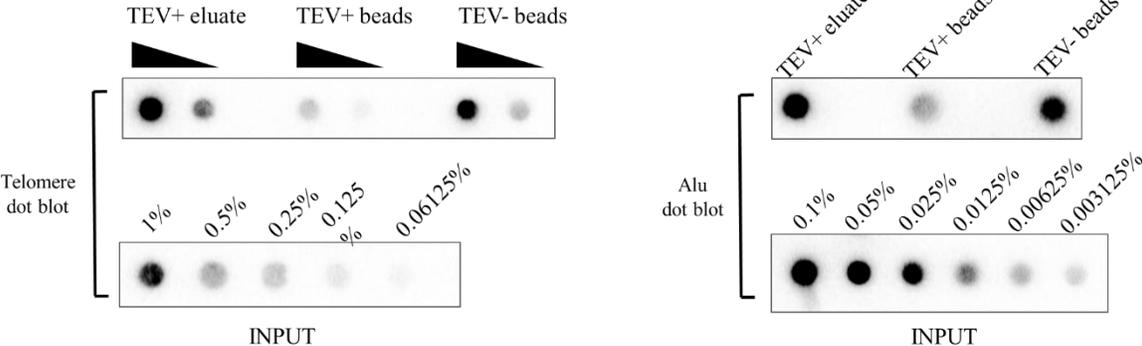
B
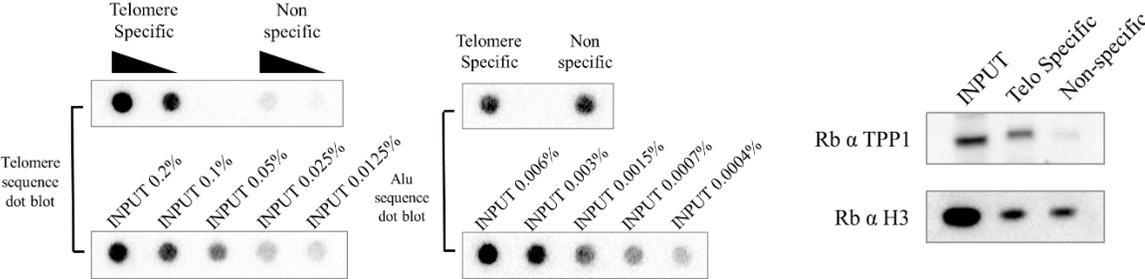
C
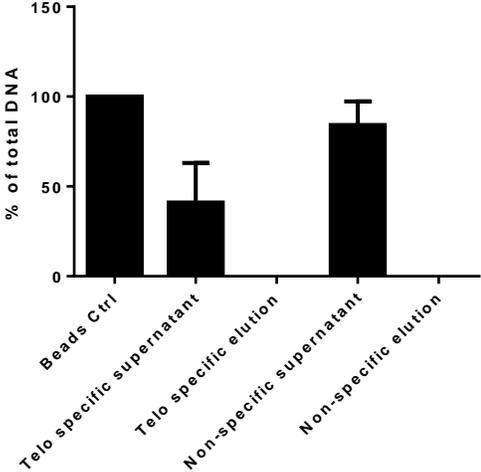
**Figure 6. dCAS-TEV-HALO can efficiently elute off beads and enrich for telomere DNA and TPP1, however, it cannot efficiently enrich for large molecular weight structures. A.** dCAS9-TEV-HALO bound to telomere specific sgRNA is incubated with formaldehyde crosslinked and sheared HeLa chromatin and eluted off beads via TEV cleavage. Enrichment of telomere DNA is tested with radioactive probes complementary to telomere DNA sequence. Background is assessed with probes complementary to ALU sequence. **B.** dCAS9-TEV-HALO is bound to either telomere specific or non-specific sgRNA and incubated with formaldehyde crosslinked and sheared HeLa chromatin. dCAS9 and associated chromatin is eluted off beads with TEV protease and DNA enrichment is assessed with telomere DNA specific radioactive probes. The same sample is tested for TPP1 protein enrichment via western blot analysis. **C.** Plasmid containing telomere repeats are incubated with dCAS-TEV-HALO and assessed for enrichment and depletion with qPCR.

Figure 7.

A
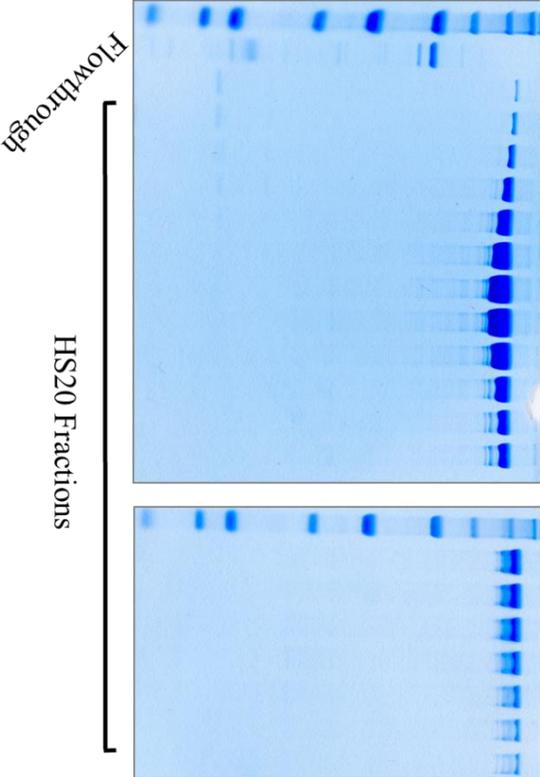
B
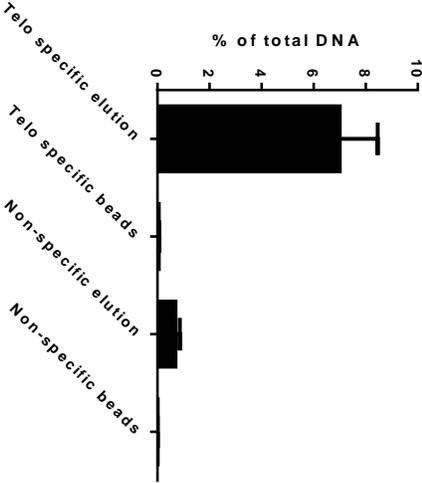
Figure 7. Purification of 6x His-dCAS-3x FLAG and assessing functionality for enriching large molecular structures. A. POROS HS20 elution fractions of 6x His-dCAS9-3x FLAG visualized by coomassie stain. **B.** Plasmid containing telomere repeats are incubated with dCAS-3x FLAG and precipitated with anti FLAG resin. Enrichment is assessed after 3x FLAG elution with qPCR.

Figure 8.

A

B

C

Telo specific pull down   Non-specific pull down

254   1120   210
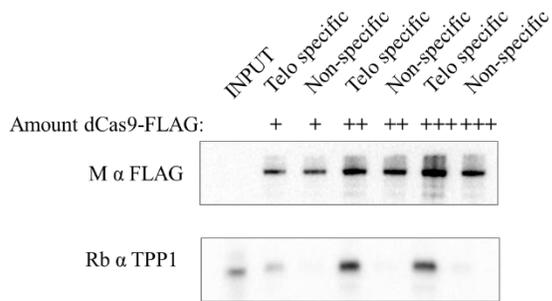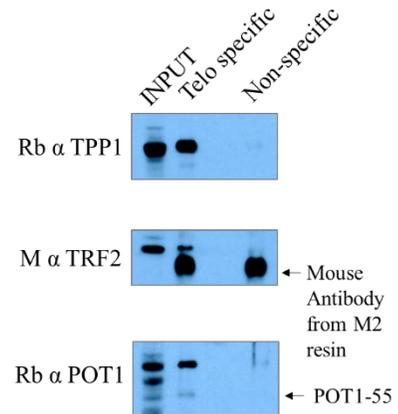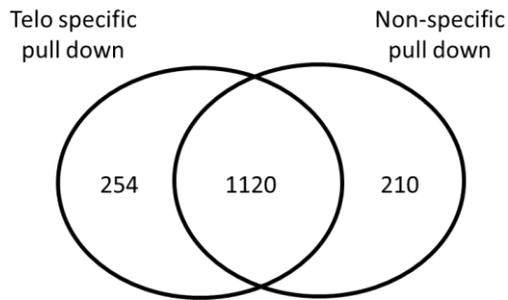
D

Protein Name	Telomere/ Non-specific dNSAF ratio
Tankyrase 1 Binding Protein	3.33
TRF2	>999
RAP1	3.78
PURα	>999
TELO2	5.69
TPP1	>999
PRKDC	>999

Figure 8. Increased ratio of dCAS9-3x FLAG to chromatin increases TPP1 enrichment. Other shelterin complex subunits can also be enriched and MudPIT mass spectrometry can identify other proteins previously associated to telomere DNA. A. Increasing amounts of dCAS-3x FLAG is incubated with the same amount of chromatin. Enrichment of TPP1 is assessed with western blot analysis. **B.** Western blot analysis of other shelterin subunits with telomere specific and non-specific pull down. **C.** Venn diagram of proteins found in MudPIT mass spectrometry results of telomere specific pull down compared to the non-specific pull down control. **D.** Table of proteins identified to associate to telomeres in (C) and their respective dNSAF ratio.

Materials and Methods

Purification of recombinant dCAS9 fusion proteins

dCAS9 fusion proteins were cloned into pET302 NT-His vectors (Thermo Fisher) and transformed into BL21-Codon Plus RIPL competent cells (Agilent). Bacterial cultures were induced at 0.6OD for incubation at 18C overnight with 0.3mM IPTG. Cell pellets were lysed in lysis buffer (500mM NaCl, 50mM HEPES, pH7.5, 5% Glycerol, 10mM 2-mercaptoethanol, 1% Triton X100, 10mM imidazole, and protease inhibitors). Lysates are frozen at -80C overnight and sonicated. Sonicated lysates were cleared by ultracentrifugation and incubated with Ni-NTA resin overnight at 4C. Resin is then washed with 20x resin volume of 250mM NaCl wash buffer (250mM NaCl, 50mM HEPES pH7.5, 5% glycerol, 10mM 2-mercaptoethanol, and 25mM imidazole) and eluted with 250mM NaCl wash buffer + 250mM imidazole. Peak elution fractions are pooled and applied to a POROS HS20 column (Applied Biosystems) and subjected to a linear gradient from 0.25M NaCl to 1M NaCl. Eluted fractions are analyzed by SDS-PAGE followed by PageBlue staining (Thermo Fisher). Peak fractions are pooled and dialyzed to 200mM NaCl, 50mM HEPES pH 7.5, 5% glycerol, and 1mM DTT. Samples are aliquoted and flash frozen for storage in -80C.

dsDNA pull down and detection

dsDNA is generated by Integrated DNA Technologies to have Cy5 or AF488 ligated at the 5' end. Equal molar concentrations of dCAS9 protein and dsDNA is incubated in 1x CAS9 buffer (100mM NaCl, 10mM Tris-HCl pH 7.9, 5mM MgCl2). The RNP is depleted from solution using the Magne HaloTag Beads (Promega) and the remaining solution is measured with Synergy H4 microplate reader. Concentration of dsDNA is measured by comparison to a standard curve of dsDNA fluorescence measurements at known concentrations.

Dot blot and hybridization

DNA samples are loaded onto Hybond-N+ membrane (Amersham Pharmacia) with a dot blot apparatus. Wells are washed with 2x SSC (300mM NaCl, 30mM NaCitrate, pH 7). The membrane is denatured with 1.5M NaCl/ 0.5N NaOH buffer for 10 minutes, neutralized with 1M NaCl/ 0.5M Tris-HCl pH 7 for 10 minutes, and the membrane is lightly dried and crosslinked using the optimum setting on the SpectroLinker XL1500 for 1 minute. The membrane is then incubated with Church buffer (1% BSA, 1mM EDTA, 500mM phosphate buffer, 7% SDS) for 30 minutes at 60C. Telomere sequence or ALU SINE sequence specific DNA oligo is labeled with radioactive p32 with T4 PNK (NEB), cleaned with Microspin G-25 columns (Fisher), and added to the membrane to incubate

overnight at 60C. The membrane is washed with 2x SSC and 0.2x SSC until non-specific background is gone. The membrane is dried, wrapped in saran wrap, and exposed to a phosphorimager screen (Kodak) and visualized with the PharosFX Plus (Bio-Rad).

Silver Stain

Protein samples are run on 10% acrylamide Bis-Tris SDS PAGE gel and then fixed with 50% MeOH and 10% acetic acid for 1 hour. The gel is then washed with 50% EtOH and then briefly treated with sodium thiosulfate before washing with water and treating with silver nitrate and 0.027% formaldehyde for 20 minutes. The gel is washed with water and the signal is developed with a sodium carbonate solution and stopped with a MeOH/ Acetic acid solution when the signal is saturated.

Western blotting

Cells are lysed in 1x Laemmli sample buffer and ran on a 10% Bis-Tris SDS PAGE gel with 1x MOPS-Tris running buffer. The proteins are transferred onto nitrocellulose membrane (GE Healthcare), blocked with 10% milk in 1x TBS + 0.1% Tween-20 (TBST), and incubated with either mouse anti-FLAG antibody (F3165, Sigma), rabbit anti-TPP1 (A303-069a, Bethyl), mouse anti-TRF2 (NB100-56506, Novusbio), rabbit anti-POT1 (AB21382, Abcam), overnight at 4C. The membrane is washed with 1x TBST, incubated with goat anti-mouse or anti-rabbit IgG coupled with HRP (PI31462, PI31430, Fisher) for one hour at room temperature, and treated with the Western Lightning ECL + detection system (Perkin Elmer).

RNA isolation, reverse transcription, and real time PCR analysis

Total RNA was extracted and purified using TRIzol reagent (Life Technologies), according to manufacturers' protocol. cDNA synthesis was performed with 1 µg of total RNA using iScript cDNA Synthesis Kit (Bio-Rad) and diluted 10-fold. Real time PCR analysis was carried out with SYBR Select Master Mix for CFX (Life Technologies) using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Gene specific primer sequences are provided in the appendix.

HeLa cell culture

HeLa cells were cultured in DMEM high glucose with GlutaMAX (Life Technologies) supplemented with 10% fetal bovine serum (HyClone).

In vitro sgRNA transcription and purification

The 19 base pair targeted DNA sequence is inserted into the middle of a 58 base pair primer behind a T7 promoter sequence (5'-TTAATACGACTCACTATAGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGC-3'). The custom primer is then used with a reverse template (5'-AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC- 3') in a DNA polymerase extension reaction to generate a dsDNA template. The dsDNA template is used with the HiScribe T7 High Yield RNA synthesis kit (NEB) to generate single stranded RNA of approximately 100 bases in length. The reaction is DNaseI treated and full length RNA is purified by isolating the correct length after running on a denaturing polyacrylamide gel with 8M urea.

MudPIT mass spectrometry and analysis

The TCA precipitated proteins were urea denatured, reduced, alkylated, and digested with recombinant endoproteinase Lys-C (Promega) and modified trypsin (Promega)(Florens and Washburn, 2006; Washburn et al., 2001). Peptides were loaded onto 100-µm fused silica (Polymicro Technologies) capillary column packed with 3 cm of 5-µm reverse phase (RP) C18 resin (Aqua, Phenomenx), 4 cm of 5-µm strong cation exchange resin (Partisphere SCX, Whatman), and 8 cm of RP C18 resin. The loaded microcapillary column was placed in-line with a Quaternary Agilent 1100 series HPLC pump and a LTQ linear ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (ThermoScientific). Ten-step MudPIT was performed on the ionized peptides as described (Florens and Washburn, 2006). Tandem mass (MS/MS) spectra were interpreted using ProLuCID and searched against a non-redundant protein *D. melanogaster* database (NCBI, 02-20-2013) containing 160 usual contaminants (human keratins, IgGs, and proteolytic enzymes). To estimate false discover rates (FDRs), the amino acid sequence of each non-redundant protein was randomized. Peptide/spectrum matches were sorted and selected using DTASelect (Zhang et al., 2010) with the following criteria set: spectra/peptide matches were retained only if they had a DeltCn of at least 0.8, and minimum XCorr of 1.8 for singly, 2.0 for doubly, and 3.0 for triply charged spectra. Additionally, the peptides had to be minimum 7 amino acids in length and fully tryptic. Peptide hits from multiple runs were compared using CONTRAST (Tabb et al., 2002). The distributed normalized spectral abundance factors (dNSAF) were used to estimate relative protein levels.

Southern Blot

DNA samples are run on 1% agarose gel in 1x TBE. The gel is incubated in 0.25M HCl for 30 minutes and then in 0.4M NaOH for 30 minutes. The DNA is transferred to Hybond XL membrane (Amersham Pharmacia) overnight through capillary action. The membrane is washed with 2x SSC with Tris-HCl (300mM NaCl, 30mM NaCitrate pH 7,

100mM Tris-HCl pH 7.5) and dried at 50C for 15 minutes. Dried membrane is incubated with Church Buffer (1% BSA, 1mM EDTA, 500mM phosphate buffer, 7% SDS) for 30 minutes at 60C before adding radioactive PCR probes prepared by fill-in reactions. After overnight incubation, the membrane is washed with 2x SSC with 0.2% SDS and 0.2x SSC with 0.2% SDS until background signal is gone. The membrane is dried, wrapped in saran wrap, and exposed to a phosphorimager screen (Kodak) and visualized with the PharosFX Plus (Bio-Rad).

DAVID bioinformatics analysis

GenInfo Identifier is taken from MudPIT mass spectrometry results and converted to UNIPROT Identifiers using UNIPROT ID mapping (www.uniprot.org/uploadlists/). UNIPROT Identifiers are inputted into DAVID Bioinformatics Resources 6.8 (david.ncifcrf.gov/tools.jsp) and UP_KEYWORDS, GOTERM_BP_DIRECT, GOTERM_CC_DIRECT, GOTERM_MF_DIRECT, and INTERPRO annotations are used for functional clustering of the gene list.

Chapter Three:

Identification of Novel Regulators of the
D. melanogaster Histone Cluster

Abstract

The precision of S phase restriction for canonical histone gene expression is a well-known biological phenomenon dating back to the early 1990's. However, despite efforts to study the control of histone gene expression both *in vivo* and *in vitro*, how the cell achieves this regulation is still not understood. Imaging experiments have provided some answers regarding how the hierarchy of certain factors involved in the histone cluster is set up, but it has also prompted questions as to the specific mechanism that separates the regulation of *H2A* and *H1* gene activation. To understand the molecular mechanism behind histone gene regulation, we adapted the dCAS9 mediated chromatin associated protein purification method as an unbiased method to enrich for proteins associated with the *Drosophila H2A/H2B* promoter. By using a mass spectrometry approach coupled with RNA interference screens, we identified VIG, VIG2, and BRAHMA as possible novel transcription and post-transcriptional regulators of *Drosophila* histone gene expression.

Introduction

Drosophila histone genes are tightly regulated from the transcriptional to the translational level (Marzluff and Duronio, 2002). This regulation is shared between all canonical histone genes partly due to their structural similarities. For example, all canonical histone gene transcripts lack a poly-A tail and instead have a unique 3' UTR that contains a Histone Downstream Element (HDE) and a stem loop that are both conserved across evolutionary species (Dominski et al., 2002). Both elements are important for the regulation of the histone mRNA. The HDE is involved in the maturation of the histone mRNA, and the stem loop is necessary for the recruitment of the proper post transcriptional processing factors and the stability of the mature mRNA by the binding of Stem Loop Binding Protein (SLBP) (Martin et al., 1997). SLBP is regulated at the translation level during the cell cycle, and because its protein level oscillates in the same manner as histone mRNA levels, it has been proposed as a regulator of mRNA degradation (Marzluff, 2005). However, the specific mechanism of whether SLBP acts as a check on degradation or acts as a recruiter for active degradation is still unknown.

Unlike the post-transcription regulation of the histone mRNAs, there does not seem to be a single motif or transcription factor that neatly explains the tight initiation at the start of S phase and the different active transcription durations between the *H1* and *H2A* genes (Guglielmi et al., 2013; Marzluff, 2005). Histone gene transcription has been long known to be restricted to the S phase of the cell cycle (Harris et al., 1991). Although it was assumed that all histone genes are actively transcribed throughout the entire S phase, an imaging study from our lab recently showed that while the *H1* gene is actively transcribed the entire time, *H2A* is only transcribed for the two hours of the six hour long S phase (Guglielmi et al., 2013). This different transcription regulation naturally leads to the question of which cellular components mediate this phenomenon. Multiple factors have been suggested to be a transcription regulator of the *Drosophila* histone cluster including GAPDH, and OCT1 (Lee et al., 2010; Zheng et al., 2003). And most recently, Multi Sex Combs (MXC) also was identified as being necessary for transcription activation of the *Drosophila* histone cluster (White et al., 2011). Still, all of these proteins are described as general, not specific, activators of the histone cluster and do not differentiate between the various histone gene promoters. While studies have shown that there are differences in the pre-initiation complex (PIC) components distinguishing between the *H1* and *H2A* promoter, the PIC has not been shown to initiate transcription on its own *in vitro* and it is thought that sequence-specific transactivators are needed to influence PIC formation (Albright and Tjian, 2000).

We want to understand the unique transcription regulation of the *H1* and *H2A* gene as well as other regulators of the histone cluster. To approach this, we developed a dCAS9 based reverse-ChIP method as described in the previous chapter and adapted it to specifically target the *H2A/H2B* promoter.

Results

Isolation of the *Drosophila* histone cluster

Having shown that the dCAS9-FLAG pull down method is capable of purifying targeted DNA sequences and its associated proteins within the genome, we moved to adopt this technique for the *Drosophila* histone cluster. Because we are interested in the activators and other potential regulators of the histone cluster and because canonical histone gene expression is restricted to S phase, we synchronized the *Drosophila* S2 cells with a two block method of Ponasterone A and hydroxyurea followed by a 2.5 hour release (FIG 1A). This synchronization method is efficient in generating S phase cells. After release from hydroxyurea block, the proportion of cells in S phase is enriched from approximately 20% to 80% of the total population while not significantly affecting viability (FIG 1B, 1C). After synchronization, the S2 cells are fixed with 1% formaldehyde for 10 minutes, washed with cold 1x PBS, and flash frozen into cell pellets for subsequent pull down assays. We started by optimizing the chromatin shearing conditions for the histone cluster. To our surprise, when we visualize the histone cluster with a *H2A*/*H2B* southern blot probe, the histone locus separates into two distinct fragments upon sonication (FIG 3A). The longer and more distinct fragment is approximately four to six kilobases long while the faint smaller fragment is between 300 to 600 bases long. This suggests that the majority of the histone cluster is extremely compact after formaldehyde crosslinking and is not very susceptible to mechanical sonication. Nevertheless, we attempted an initial pull down following a general protocol similar to the telomere pull down experiments (FIG 2A).

To start, we designed and *in vitro* transcribed sgRNAs targeting the *H1* and *H2A*/*H2B* promoter (FIG 2B). As we are interested in proteins binding and regulating the promoter region of the *H2A*/*H2B* gene we avoided targeting any sgRNA there to prevent possible steric clashes between the dCAS9-FLAG itself and potential covalently crosslinked proteins. We performed an initial trial by separating the mixture of sgRNA into three pools and performing the dCAS9-FLAG pull down (FIG 2B). By using qPCR to assess enrichment at the regions around the *H2A*/*H2B* promoter, we can see that sgRNAs within pool 2 and pool 3 seem to enrich, albeit weakly, for regions around the *H2A*/*H2B* gene whereas pool 1 does not as expected from their respective sgRNA targeting locations (FIG 3B). Surprisingly, the use of one sgRNA gave similar enrichment of the chromatin regions surrounding the *H2A*/*H2B* promoter relative to the pooled sgRNA, so we used a single sgRNA for the troubleshooting experiments (FIG 3D). One noticeable issue is the high amount of background associated with the non-specific pull downs. We attempted to solve this by trying common steps in regular chromatin immunoprecipitation protocols such as pre-clearing the resin with ssDNA and high salt washes. However, neither attempts had significant effect on the reduction of non-specific background, in addition, high salt washes seem to significantly reduce the amount of dCAS9-FLAG that bind to the target chromatin despite previous reports of dCAS9 binding that can withstand 6M urea washes (FIG 3C, 3D) (Sternberg et al., 2014).

Numerous proteins are necessary for the proper transcription and processing of histone mRNAs within a tight physical space (Duronio and Marzluff, 2017). So we thought one reason for the high background could be due to the histone cluster being a compact and sticky mix of protein, DNA, and RNA making it hard to fragment as suggested from the southern blot result (FIG 3A). To test this hypothesis we more than doubled the sonication time and performed the dCAS9-FLAG pull down with a single sgRNA targeting the *H2A* gene body. Indeed, increased sonication time dramatically reduced the amount of background in the non-specific dCAS9-FLAG sample (FIG 4A). Additional increases in sonication was also tested, and while it seems to further reduce the non-specific background, it also reduced the enrichment for fragments of interest (FIG 4A). Once we managed to reduce the background, we retested whether using a mix of sgRNAs tiling the *H2A*/*H2B* promoter region will help pull down efficiency. The use of four different sgRNA targeting around the promoter region was indeed able to significantly help pull down enrichment when the chromatin is extensively sonicated (FIG 4B). And by titrating the amount of dCAS9-FLAG protein used for the pull down, we can further increase the enrichment without substantially increasing non-specific background levels (FIG 4C). With the current optimization we are able to enrich targeted histone cluster sequence by three to five fold over background. To see if we can continue to improve the enrichment we tested additional sgRNA targets near the *H2A*/*H2B* promoter (FIG 5A). Using a combination of eight different sgRNAs, we are able to improve our targeted dCAS9-FLAG pull down signal to noise ratio to greater than ten fold for the desired promoter region (FIG 5B).

Identification of regulators of the *Drosophila* histone cluster

Having achieved a significant enrichment of the *H2A*/*H2B* promoter, we prepared pull down samples of histone cluster specific and non-specific targeting from 500 million synchronized cells each, verified enrichment via qPCR, and submitted the samples for MudPIT mass spectrometry analysis (FIG 6A). In summary, 75 proteins were identified only in the histone cluster pull down, 673 proteins were identified in both samples, and 184 proteins were identified only in the non-specific control (FIG 6B). Like the telomere pull down in HeLa cells, we identified proteins that are enriched for our target by grouping those that are found only within the histone cluster specific pull down with those that are found in both samples but are enriched with an dNSAF ratio of 1.5 or more (Zhang et al., 2010). 342 proteins are identified by this criterion, but this is still a list that is too large to individually verify *in vivo*. To further narrow down the list we utilized the DAVID bioinformatics database to identify any proteins with a GO term directly associated to nucleic acids (Huang et al., 2009a, 2009b). With these criteria, we generated a list of 17 proteins that we can verify by an *in vivo* screen through double stranded RNA (dsRNA) knockdowns.

Canonical histone mRNA levels are dynamically regulated throughout the cell cycle, so by checking mRNA steady state levels in a heterogeneous population we can get a quick readout on whether transcription regulation or mRNA degradation is affected by a specific knockdown. We performed dsRNA mediated RNA interference for all 17

gene on S2 cells, isolated RNA, and checked *H2A* mRNA levels relative to Tub84b as a reference gene via RT-qPCR. Out of 17 genes, *CG11844*, *Brahma*, and *Vig* had a negative effect on the levels of *H2A* mRNA suggesting that they are either positive transcription regulators of the gene or regulators preventing mRNA degradation (FIG 7A). Surprisingly, *CG11844* is also known as *Vig2* and is a paralog of *Vig* in *Drosophila melanogaster* (Carvalho et al., 2015). Because *Vig* and *Vig2* are paralogs of each other and both individual knockdowns had a significant effect on the expression of *H2A*, it is possible the two proteins have redundant molecular functions and can compensate for one another when they are individually knocked down. To test this, we performed double knockdown experiments of both *Vig* and *Vig2* in S2 cells and saw that the decrease in *H2A* mRNA levels was indeed significantly greater (FIG 7B). Interestingly, the triple knockdown of *Vig, Vig2*, and *Brahma* also had a slightly greater reduction in *H2A* mRNA level than *Brahma* knockdown alone (FIG 7B).

Previous literature have shown that Brahma is the *Drosophila* homolog of the yeast SNF2/SWI2 protein that is part of the SNF/SWI nucleosome remodeling complex (Peterson and Tamkun, 1995; Tamkun et al., 1992). Its previously described role in regulating the activation of homeotic genes suggest that *Brahma* might be regulating the histone cluster as a whole, much like HERS repression mechanism, instead of regulating *H2A* alone (Ito et al., 2012). To test this, we looked at the mRNA expression level of *H1*, *H2A*, and *H3* after *Brahma* knockdown in S2 cells. Indeed, *Brahma* knockdown has a significant effect on all three steady state mRNA levels suggesting that *Brahma* is acting as an activator for general histone transcription (FIG 7C).

Vig is identified as a member of the RNA interference complex in *Drosophila*, and *Vig* and *Vig2* publications are mostly focused on *Vig*'s role as a member of the RNA interference machinery (Caudy et al., 2002, 2003; Tomari, 2005). While there is one published report on *Vig* and *Vig2*'s role in affecting heterochromatin formation on an organismal level, none of these give any insight as to what role they might be playing at the *Drosophila* histone cluster (Gracheva et al., 2009). *Vig* and *Vig2* overexpression in S2 cells also did not give any additional hints as to how these proteins can interact with the histone cluster as they are both primarily localized in the cytoplasm (FIG 8A). Lacking information, we turned to *Vig* and *Vig2* homologs in other organisms. The mammalian homolog of both genes was identified in 2001 as *Serbp1* (Heaton et al., 2001). The 2001 publication and subsequent papers also describe SERBP1 as an RNA binding protein that plays a role in regulating the stability of the RNAs it interacts with (Ahn et al., 2015). Given *Serbp1*'s function in mammalian cells, we believed it is possible that *Vig* and *Vig2* might specifically interact with *H2A* mRNA to regulate its stability. We tested this potential interaction by generating and using stable VIG-HALO-V5 and VIG2-HALO-V5 overexpression cell lines and performed a ribonucleoprotein (RNP) immunoprecipitation via the V5 tag. As hypothesized from homology and their identification from the histone cluster pull down, VIG and VIG2 specifically bind and enrich for *H2A* mRNA when compared to reference gene mRNAs such as *Tub84b*, *Actin*, and *Rpl32* (FIG 8B). Interestingly, while both VIG and VIG2 show some affinity for *H3* and *H4* mRNA, neither protein bound to *H1* mRNA over background levels (FIG 8B).

While processing and verifying the first mass spectrometry data *in vivo*, we submitted another set of histone cluster specific and non-specific dCAS9-FLAG pull down samples for MudPIT analysis. In order to identify proteins that are weakly enriched we doubled the amount of material used and processed one billion synchronized S2 cells per pull down. Significantly more proteins were identified in the second mass spectrometry experiment, as expected with the increased material. To group the proteins that were enriched in the second pull down experiment, we used the same criteria as outlined for the first set. Then we took the enriched protein list and submitted it to the same DAVID bioinformatics analysis to look at the enriched protein groups and compared with the initial experiment. Many of the functional clusters were the same, such as ribosomal protein, proteasome complex, and spliceosome, however, well documented protein contaminations from studies such as the CRAPome suggest that these highly abundant and enriched proteins are likely not valid hits (FIG 9A) (Mellacheruvu et al., 2013; Ohta et al., 2010). Indeed, if we look at the results for the same bioinformatic analysis performed on the telomere specific pull down many of the top clustering groups are the same despite spatially and functionally different genomic loci being targeted in each experiment (FIG 9A).

Despite the large amount of non-specific hitchhiker proteins that seem to come down with any sort of DNA enrichment, when we compare the top 15 functional clusters for the first and second histone specific pull down there were differences that appear (FIG 17A) (Ohta et al., 2010). Compared to the first, the second histone cluster pull down sample has a high enrichment of proteins with the LisH dimerization motif (Cluster 7) and proteins involved in germ line maintenance/ COP9 signalosome (Cluster 13). Interestingly, MULTI SEX COMBS (MXC) was one of the eight LisH motif containing proteins enriched and has been previously shown to both physically localize to the histone cluster and also be required for proper histone gene transcription (Terzo et al., 2015; White et al., 2011). In addition, a large subset of the proteins within the germ line maintenance/ COP9 signalosome function group are also found in a table of genes that the White et al. study lists as potential regulators for proper *Drosophila* histone gene expression. While it is encouraging that two separate screening methods found overlapping proteins, additional experiments are necessary to understand how these proteins are involved in the regulation of the histone cluster.

Discussion

After the proof of principle experiments targeting telomeres in HeLa cells, we demonstrated the versatility of our method by successfully purifying the *Drosophila* histone cluster with minimal changes to the protocol (FIG 14). From our initial mass spectrometry experiment, we were able to identify three potentially novel *H2A* regulators, *Brahma*, *Vig*, and *Vig2*, by assessing knockdown effects on *H2A* transcript levels (FIG 15A). *Brahma* is a well-known member of the SNF2/SWI2 nucleosome remodeling complex and was initially identified in *D. melanogaster* as a required activator of homeotic genes (Tamkun et al., 1992). Our preliminary data suggest that BRAHMA has a specific function in regulating overall histone cluster activation because knockdown of BRAHMA causes a decrease in the steady state mRNA levels of *H1*, *H2A*, and *H3* genes (FIG 15C). The potential regulation of the entire histone cluster by BRAHMA is reminiscent of the HERS mediated repression of the histone cluster (Ito et al., 2012). It is tempting to speculate that these two proteins play opposite roles to facilitate active and repressive chromatin context at the histone cluster as some reports suggest that the mammalian BRAHMA homolog, BRG1, mediates euchromatin formation (Ho et al., 2011; Singhal et al., 2010). However, additional experiments such as ChIP of various histone modifications will be necessary to confirm and further dissect BRAHMA's role in the regulation of *Drosophila* histone cluster.

Finding both *Vig* and *Vig2* at the *H2A*/*H2B* gene region of the histone cluster was interesting as little is known about their function in D. melanogaster. *Vig* was identified as a part of the RNA interference complex and studies in *Drosophila* have focused on its described role in silencing (Caudy et al., 2002; Tomari, 2005). Even less is known about the molecular function of *Vig2* as publications are limited to the identification of VIG2 in the structure of the *D. melanogaster* 80S ribosome and a possible role in affecting global heterochromatin formation along with VIG (Anger et al., 2013; Gracheva et al., 2009). While not much is known about *Vig* and *Vig2*, *Serbp1*, the mammalian homolog, gave us a hint that these proteins might be binding and regulating histone mRNA (Ahn et al., 2015). And that is indeed the case as both VIG and VIG2 specifically bind to *H2A* mRNA when compared to reference genes such as *Actin*, *Tub84b*, and *Rpl32* (FIG 16B). SERBP1 is shown to bind to RNA through a RGG box 3' of the evolutionary conserved region (Ahn et al., 2015; Heaton et al., 2001). Although an exact RGG domain homolog cannot be found in either VIG or VIG2, similar domains can be identified in comparable locations 3' of the conserved region. It remains to be seen whether VIG and VIG2 RNA binding is mediated through this region and which part of the histone mRNA it recognizes, but VIG and VIG2 binding of the histone mRNA is likely to be near the 3' UTR based on reported SERBP1 mRNA binding and general binding regions of RNPs that regulate mRNAs (Siomi and Dreyfuss, 1997). The preferential binding of VIG and VIG2 for *H2A* but not *H1* mRNA is consistent with the timing difference in active transcription of the two genes during S phase. We speculate that VIG and VIG2 might act as an additional level of protection against mRNA degradation as H2A mRNA levels could be more sensitive to degradation due to lack of continuous transcription during S phase. Additional experiments such as RNA binding mutant immunoprecipitation and finding the *H2A*

mRNA sequence that VIG and VIG2 binds to are necessary to further build on this hypothesis.

We also identified MXC with mass spectrometry when additional starting material was used (FIG 17B). The identification of MXC serves as a good positive control as it is a previously described transcription regulator of the *Drosophila* histone cluster even though it is a non-DNA binding protein (White et al., 2011). The discovery of MXC stemmed partially from its Lis homology (LisH) domain and this domain has been shown to be important for the recruitment of protein to the histone cluster (Terzo et al., 2015). LisH domains are also known to be involved in protein dimerization and the assembly of multi-protein complexes (Cerna and Wilson, 2005; Kim et al., 2004; Mateja et al., 2006; Mikolajka et al., 2006). Because of the importance of this domain for MXC recruitment to the histone cluster and its prominent role in dimerization, it is currently proposed that the LisH domain helps MXC assemble into an oligomeric network that provides a scaffold for other components of the histone cluster to assemble (Duronio and Marzluff, 2017). Interestingly, in addition to MXC we identified seven other proteins with a recognizable LisH domain out of the 18 possible proteins in the *D. melanogaster* proteome. Given the domain's role in dimerization, it will be interesting to test whether these other LisH containing proteins also localize and play a role at the histone cluster.

Figure 1.

A

Time — Transfer cells into suspension culture — 24 hours — 0.2nM Ponasterone A — 18 hours — 1.5mM Hydroxyurea — 2.5hr — Fresh media release — Collection

B

Asynchronous cells

Synchronized Cells

C

= Asynchronous
= Synchronous
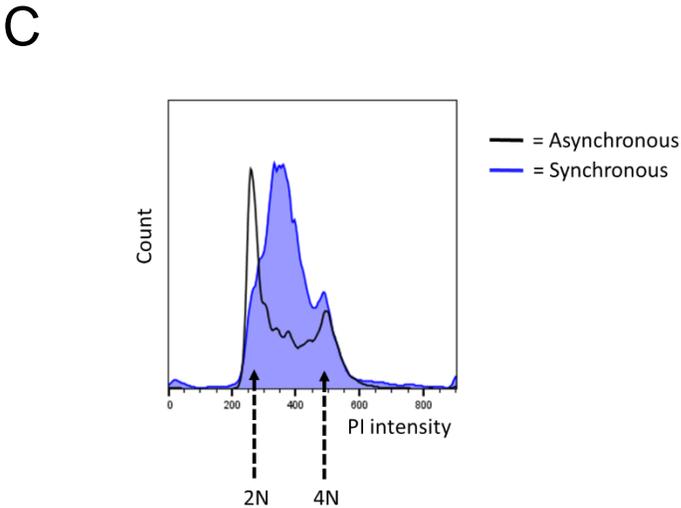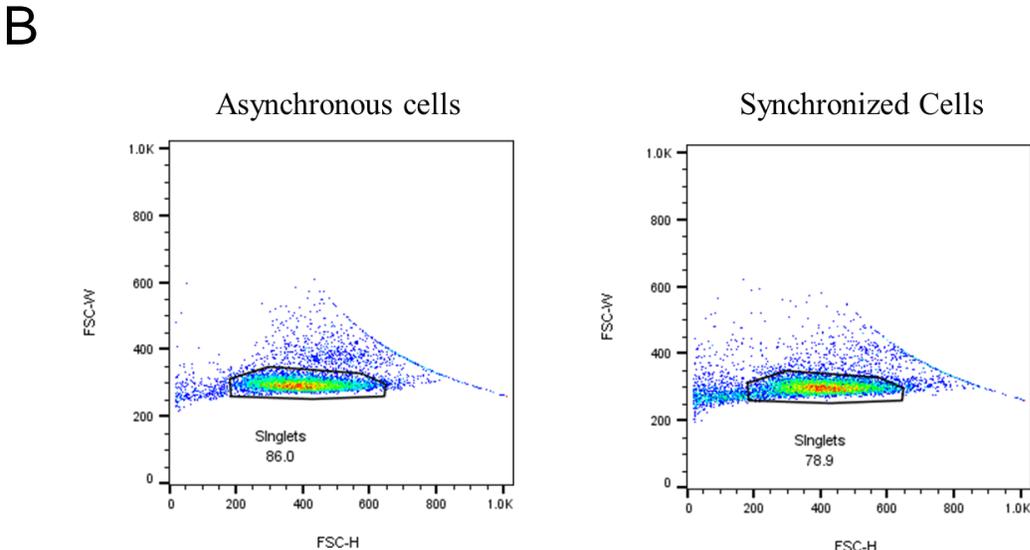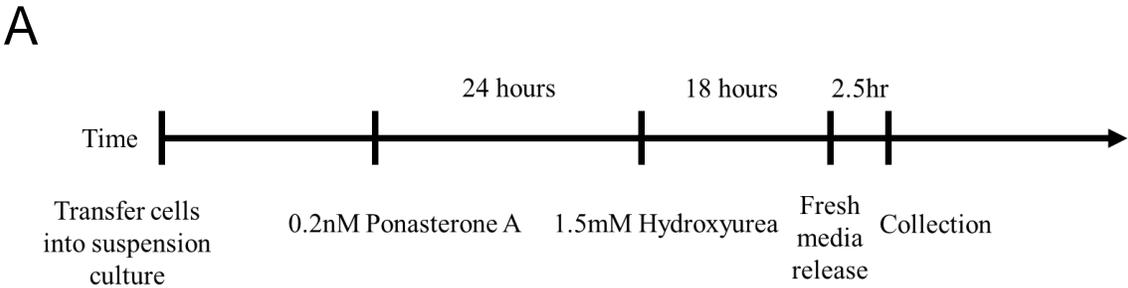
**Figure 1. S2 cells can be properly synchronized in suspension cultures without greatly affecting cell viability. A.** Synchronization scheme for S2 cells in suspension culture. **B.** Comparison of the FSC-W and FSC-H of sorted S2 cells to distinguish proportion of viable singlet cells. **C.** Synchronized and asynchronous S2 cells are stained with propidium iodide (PI) and the intensity is measured by the flow cytometer.

Figure 2.

A
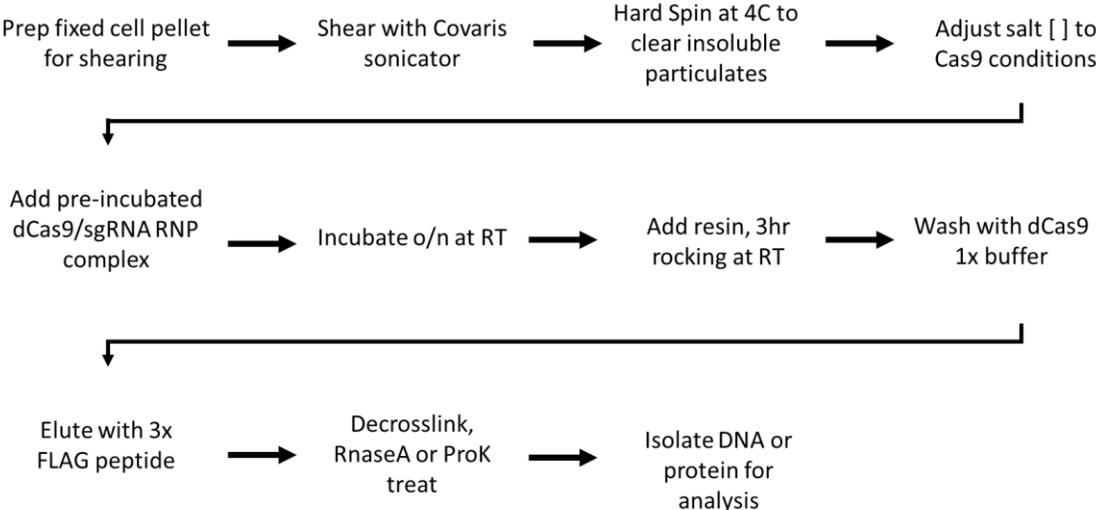
B
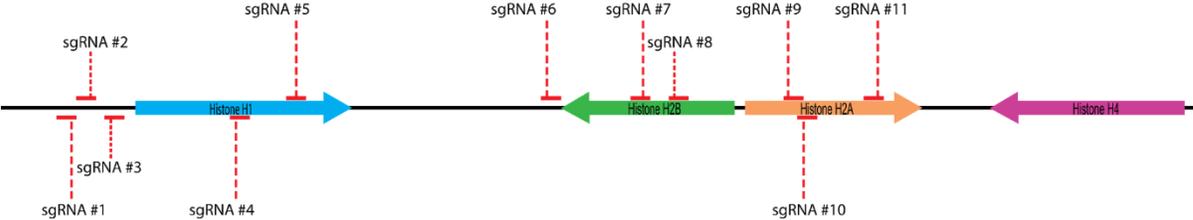
Figure 2. General dCAS9-3x FLAG chromatin purification protocol and layout of sgRNA placements across the *H1*, *H2A*, and *H2B* gene in the histone cluster. A. General protocol for dCAS9-3xFLAG chromatin purification method. **B.** A schematic of the histone cluster and the sgRNA targeting locations.

Figure 3.

A

C

D
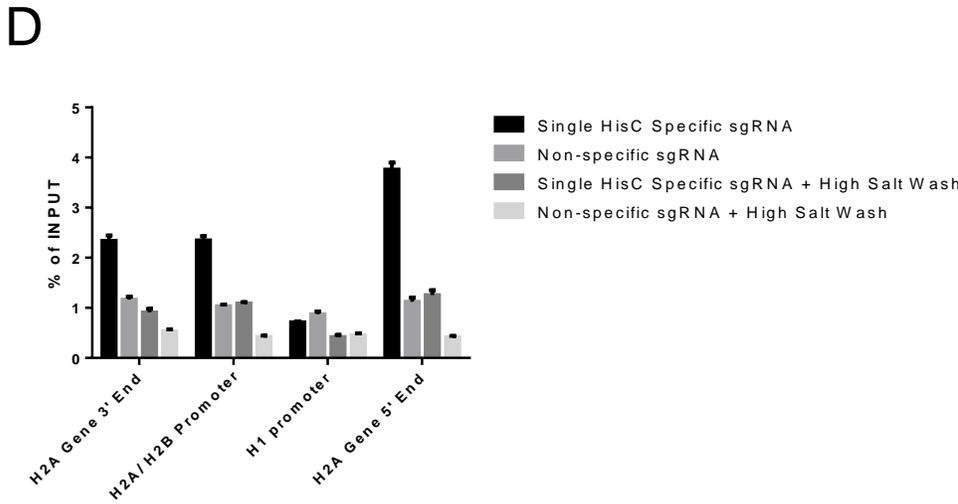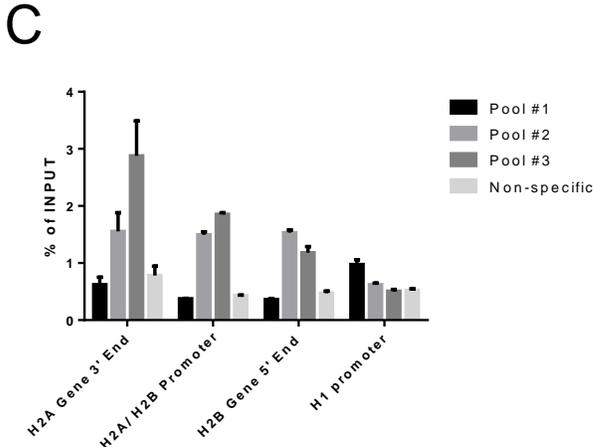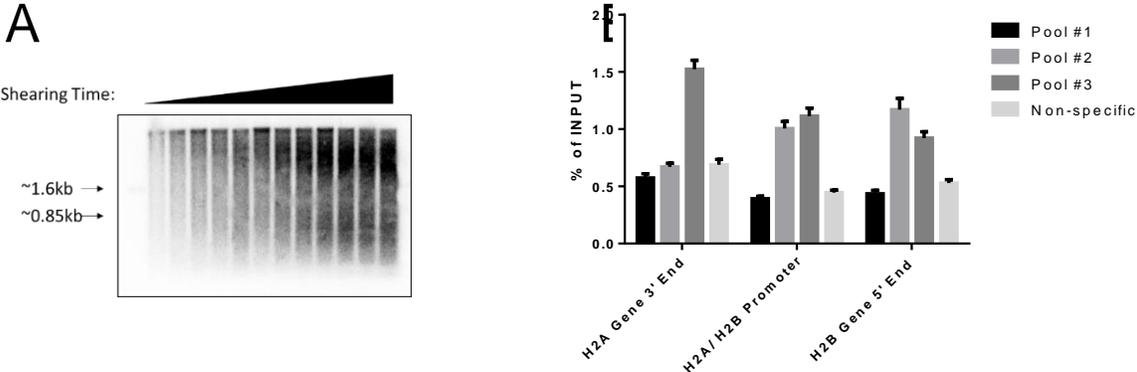
Figure 3. Pull down the *Drosophila* histone cluster results in low enrichment and high background. A. Synchronized S2 cells fixed with 1% formaldehyde is sheared with increasing amount of time. Radioactive probes against the *H2A/H2B* genomic region is used to visualize the sheared DNA lengths of the histone locus. **B.** 11 different sgRNA targeting the histone cluster is pooled into three different groups and used to specifically enrich for the histone cluster. Enrichment is assessed by qPCR against different regions of the *H2A/H2B* genomic locus. **C.** Same pull down assay as in (A) except that the anti FLAG resin is blocked overnight with salmon sperm DNA. **D.** A single sgRNA (sgRNA #10) was used to enrich for the *H2A/H2B* histone cluster region and seems to be as effective as using a pool of sgRNAs. High salt washes incorporated into the purification method slightly decreases the background but also dramatically decrease the enrichment signal.

Figure 4.

A

B

C
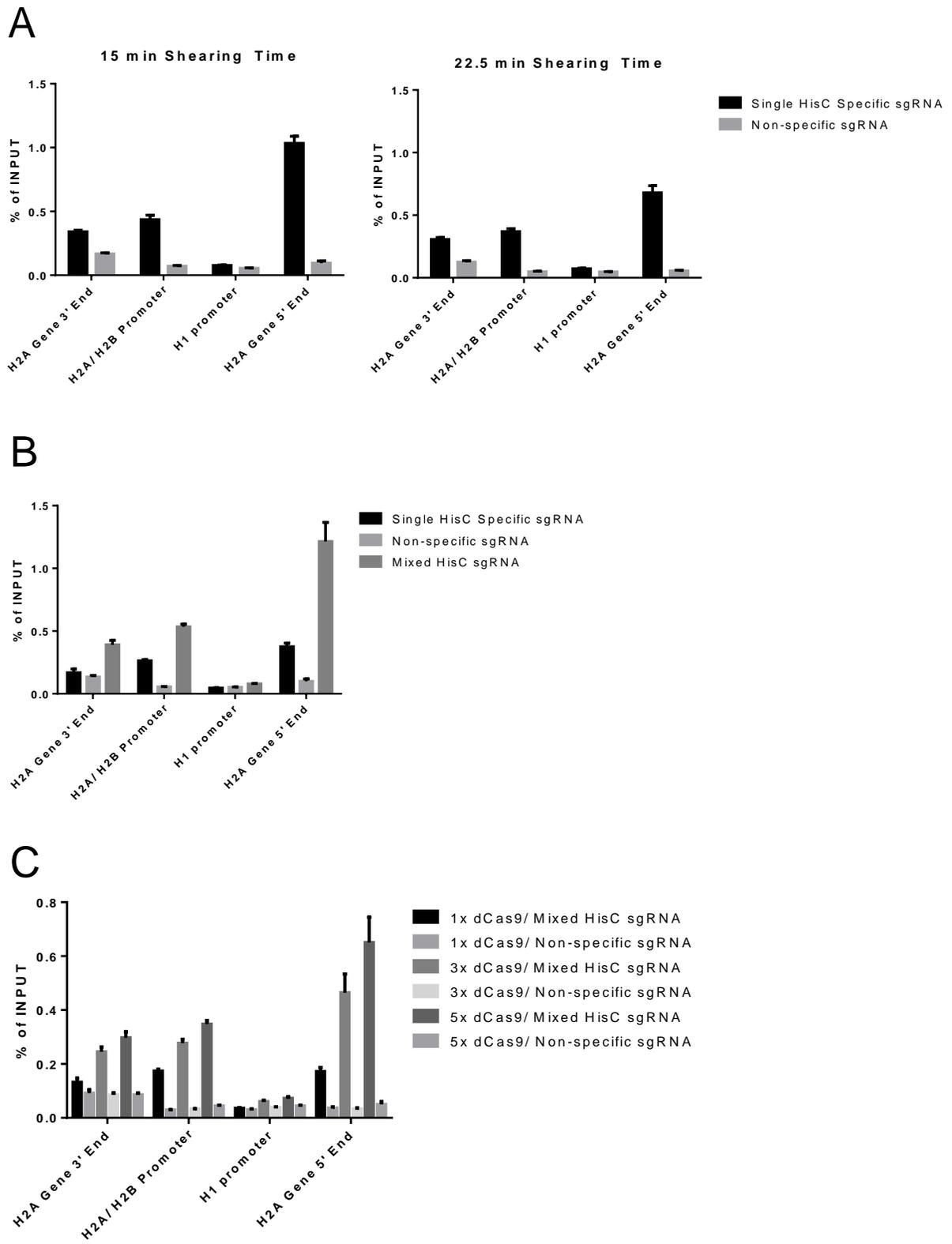
Figure 4. Increase in shearing time can dramatically decrease background associated with histone cluster pull down. Using a pool of sgRNA can also greatly increase pull down efficiency when highly sheared S2 chromatin is used. A. Increasing the shearing time from six minutes to 15 minutes can dramatically decrease the amount of background associated with *H2A/H2B* specific pull down with a single sgRNA. Further increase in shearing time is not associated with significant improvement in background reduction. **B.** The efficiency of dCAS9-3x FLAG bound to a single sgRNA targeting the histone cluster is compared to a mix of four different sgRNAs targeting the same region. **C.** Increasing amounts of dCAS9-3x FLAG is used with the same amount of crosslinked and sheared S2 chromatin to optimize pull down efficiency.

Figure 5.

A

B
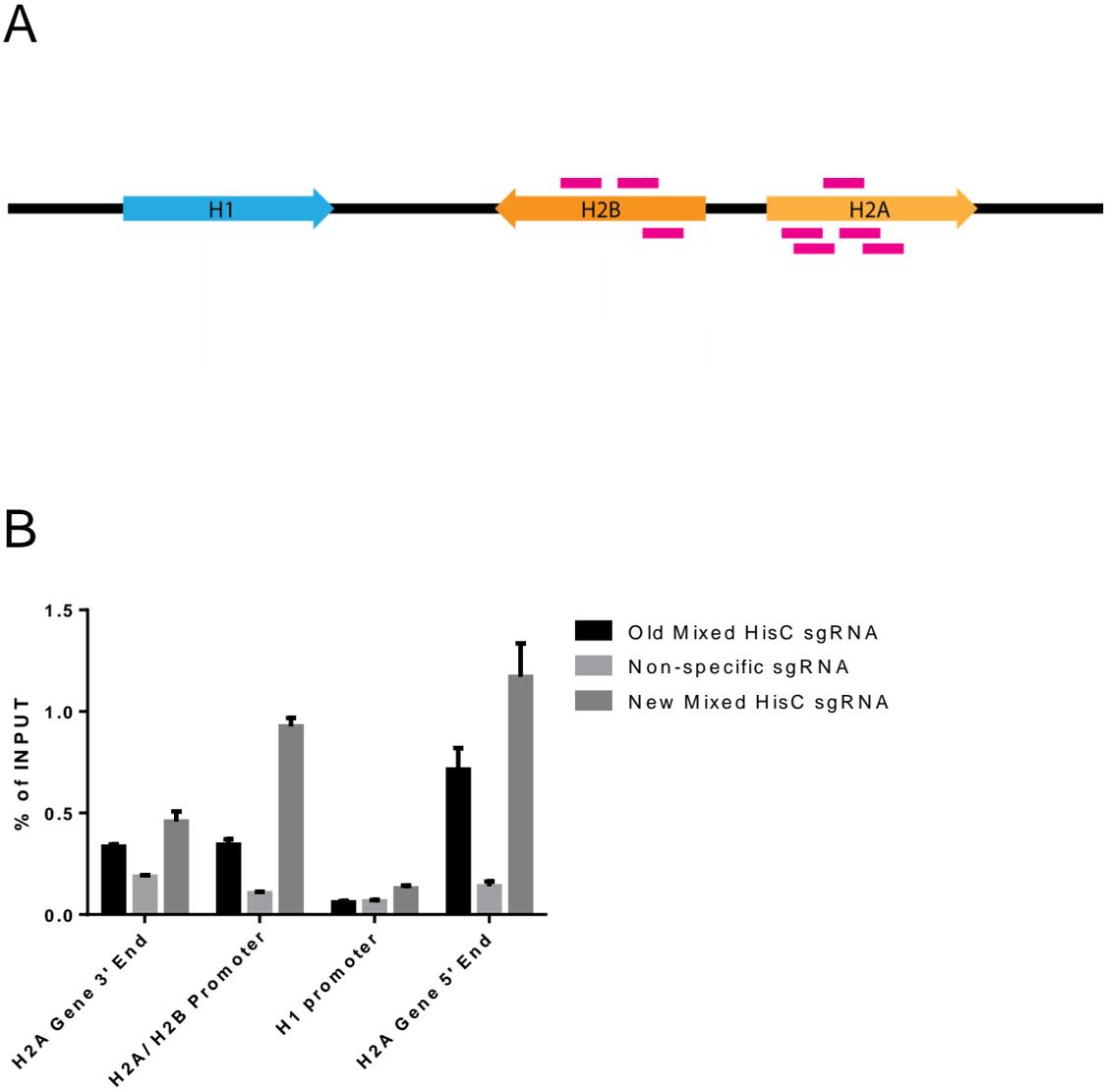
Figure 5. Increasing number of sgRNA targeting the *H2A/H2B* genomic region increases efficiency of enrichment. A. Graphical representation of the eight different genomic targets of the sgRNA used in the new mixed histone cluster pool. **B.** Comparison of *H2A/H2B* locus enrichment when four sgRNAs versus eight sgRNA are bound to the same amount of dCAS9-3x FLAG.

Figure 6.

A

B
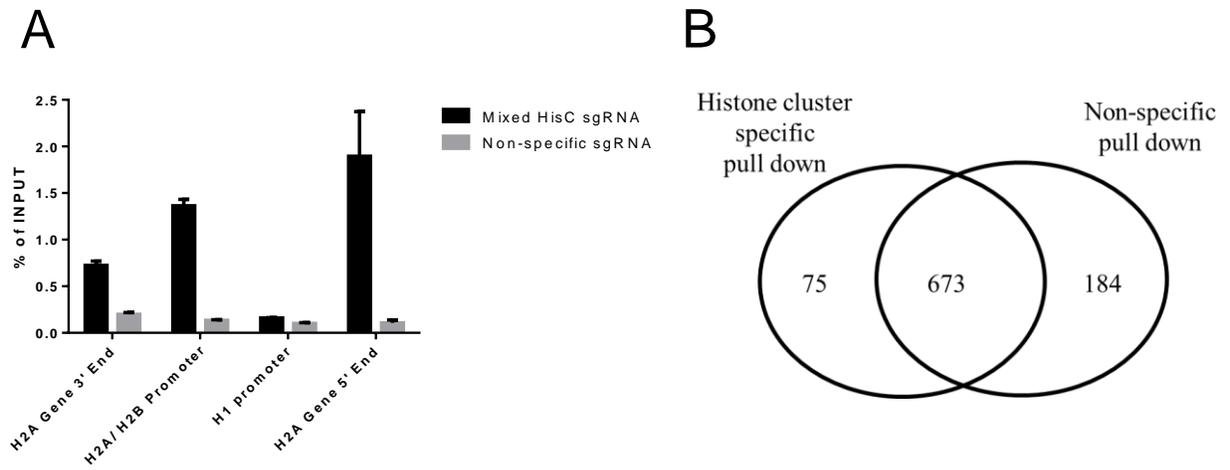
C

Protein Name	Histone cluster/ Non-specific dNSAF ratio
CG3262	> 999
CG11844	> 999
CG12608	> 999
Bin1	> 999
BubR1	> 999
Cp190	> 999
CG8771	> 999
Syncrip	> 999
Dbp21E2	> 999
Vig	5.69
Row	3.15
Brahma	2.93
Cdc5	2.38
Nap1	2.28
Bap60	2.14
CKIIalpha	2.14
Rox8	2

Figure 6. MudPIT mass spectrometry on *H2A/H2B* enriched samples identified a large number of unique/ enriched proteins and 17 proteins that are associated with nucleic acids. A. qPCR analysis of the histone cluster specific sample vs non-specific sample sent for mass spectrometry. **B.** Venn diagram of the total proteins found only in histone cluster specific pull down, non-specific pull down, and those that are found within both samples. **C.** Table of proteins and their associated dNSAF ratio values that are at least enriched by 1.5 fold and associated to nucleic acids by DAVID analysis.

Figure 7.

A
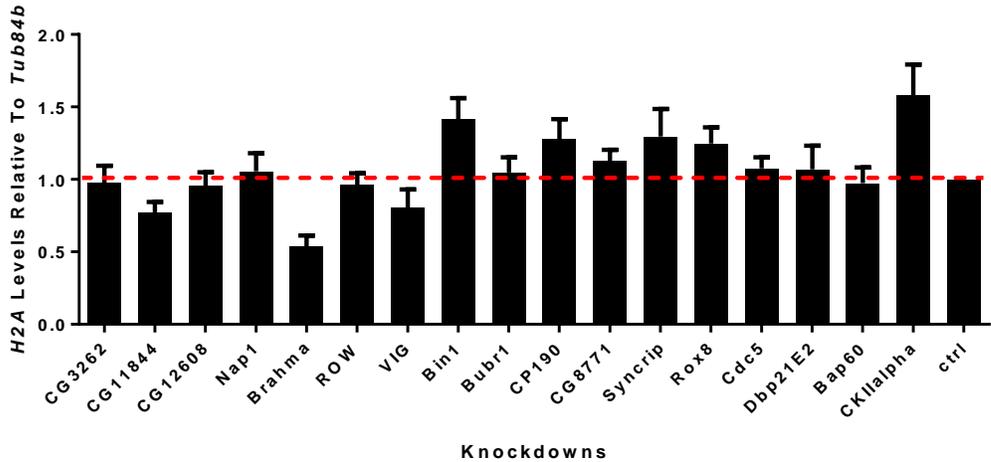
B
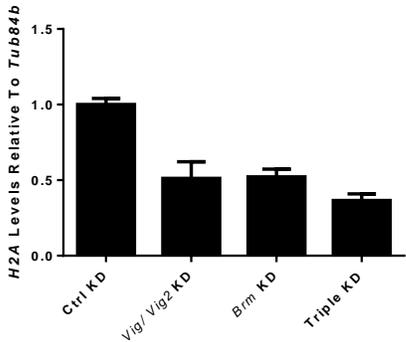
C
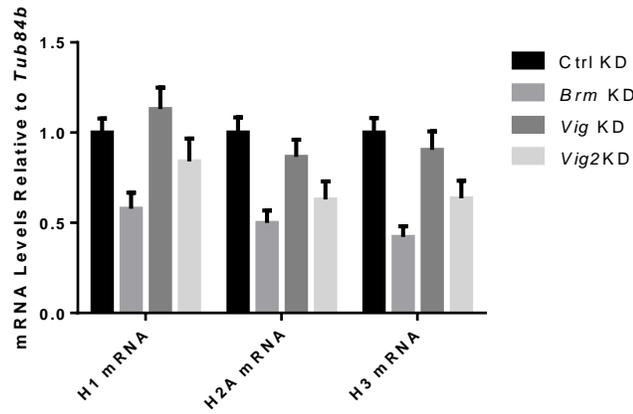
Figure 7. Knockdown assays reveal potential histone gene regulators that have specific effects on different histone mRNAs. A. S2 cells are treated with dsRNA against the specified gene for three days and the levels of *H2A* mRNA is assessed by qPCR. Total RNA is isolated by TRIZOL extraction and cDNA is synthesized with iSCRIPT reverse transcriptase kit. **B.** The effect on *H2A* mRNA is compared when double and triple knockdowns are performed on potential regulators of *H2A* expression. **C.** The effect of the individual knockdowns on three different histone genes is checked by qPCR.

Figure 8.

A
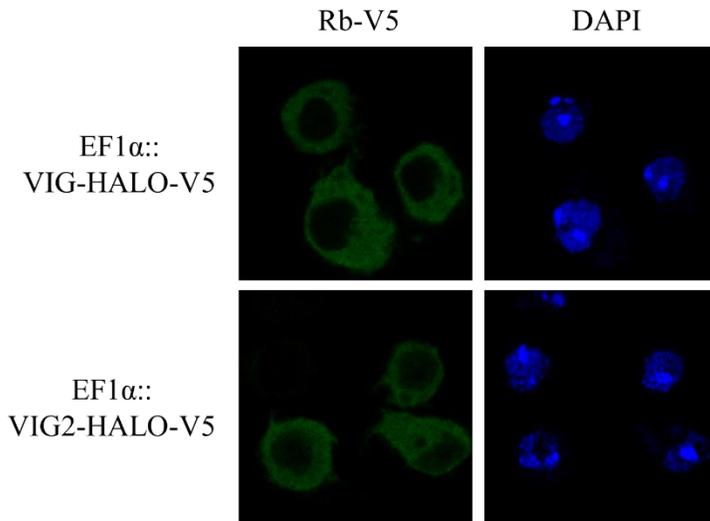
B
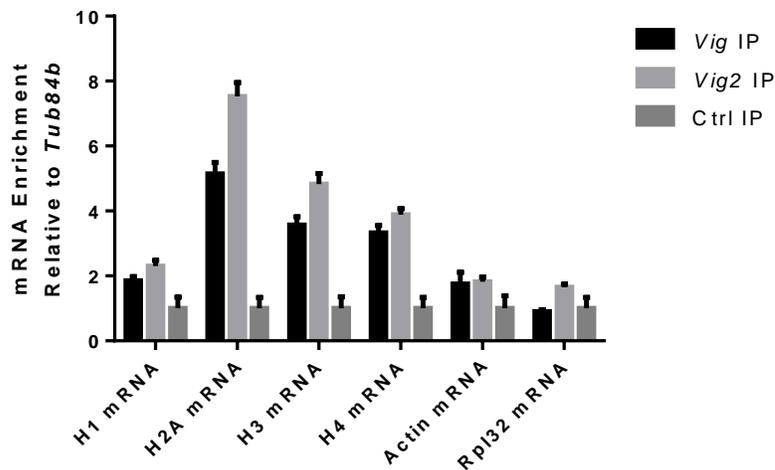
Figure 8. Localization of *Vig* and *Vig2* and its ability to specifically bind to *H2A* mRNA. A. VIG and VIG2 cDNA is overexpressed as a HALO-V5 fused protein driven by an EF1α promoter in S2 cells. Cells are then fixed and the overexpressed protein is visualized by Rabbit anti V5 antibody. **B.** Cells expressing VIG-HALO-V5 and VIG2-HALO-V5 are lysed and a anti V5 resin is used to immunoprecipitate the fusion protein. RNA is isolated by TRIZOL and iSCRIPT reverse transcriptase is used to generate cDNA. RNA enrichment is assessed by qPCR after reverse transcription.

Figure 9.

A

DAVID Analysis Rank	Telomere Specific Pull Down	1st Histone Cluster Pull Down	2nd Histone Cluster Pull Down
1.	Mitochondrial Proteins	Ribosomal Proteins	Nucleotide Binding
2.	Nucleoplasm	Proteasome Complex	Ribosomal Proteins
3.	RNA processing, Splicing	Protein Biosynthesis	Protein Biosynthesis
4.	Ribosomal Proteins	Protein Import	Integrator Complex
5.	Protein Transport	Protein Transport	Spliceosome
6.	Nucleotide Binding	Spliceosome	mRNA binding
7.	Cytoskeleton Proteins	Nucleotide Binding	LisH Dimerization Motif
8.	mRNA Transport	Aminoacyl tRNA Synthethase	Mitochondrion
9.	GTP Binding	Pyridoxal Phosphate Binding	Proteasome
10.	Cellular Respiration	Fatty Acid Degradation	Protein Import
11.	Nuclear body	GTP Binding	WD Repeat Containing Domains
12.	Golgi	Cytoskeleton	Helicases
13.	Protein Import	Isomerase	COP9 Signalosome
14.	Cell Cycle	WD Repeat Containing Domains	Cell Cycle
15.	Pigment Granule	COPI Vesicle Coat	Hydrogen Ion Transport

B

LisH Domain Containing Proteins
CG31357
CG3295
EBI
CG7611
RanBPM
SMU1
MAHJ
MXC

Figure 9. Comparison of top DAVID analysis clusters between Telomere specific pull down in HeLa cells and two histone cluster pull downs show similarities, but LisH motif containing proteins are highly enriched in one histone cluster pull down. A. Table of the top 15 clusters when the proteins enriched by at least 1.5 dNSAF ratio is analyzed by DAVID. **B.** List of the LisH domain containing proteins enriched in the second histone cluster pull down.

Materials and Methods

Purification of recombinant dCAS9 fusion proteins

dCAS9 fusion proteins were cloned into pET302 NT-His vectors (Thermo Fisher) and transformed into BL21-Codon Plus RIPL competent cells (Agilent). Bacterial cultures were induced at 0.6OD for incubation at 18C overnight with 0.3mM IPTG. Cell pellets were lysed in lysis buffer (500mM NaCl, 50mM HEPES, pH7.5, 5% Glycerol, 10mM 2-mercaptoethanol, 1% Triton X100, 10mM imidazole, and protease inhibitors). Lysates are frozen at -80C overnight and sonicated. Sonicated lysates were cleared by ultracentrifugation and incubated with Ni-NTA resin overnight at 4C. Resin is then washed with 20x resin volume of 250mM NaCl wash buffer (250mM NaCl, 50mM HEPES pH7.5, 5% glycerol, 10mM 2-mercaptoethanol, and 25mM imidazole) and eluted with 250mM NaCl wash buffer + 250mM imidazole. Peak elution fractions are pooled and applied to a POROS HS20 column (Applied Biosystems) and subjected to a linear gradient from 0.25M NaCl to 1M NaCl. Eluted fractions are analyzed by SDS-PAGE followed by PageBlue staining (Thermo Fisher). Peak fractions are pooled and dialyzed to 200mM NaCl, 50mM HEPES pH 7.5, 5% glycerol, and 1mM DTT. Samples are aliquoted and flash frozen for storage in -80C.

RNA isolation, reverse transcription, and real time PCR analysis

Total RNA was extracted and purified using TRIzol reagent (Life Technologies), according to manufacturers' protocol. cDNA synthesis was performed with 1 µg of total RNA using iScript cDNA Synthesis Kit (Bio-Rad) and diluted 10-fold. Real time PCR analysis was carried out with SYBR Select Master Mix for CFX (Life Technologies) using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Gene specific primer sequences are provided in the appendix.

Drosophila S2 cell synchronization

Two confluent T150 flask of *Drosophila* S2 cells are dissociated from the flask and cultured in a Wheaton double side arm spinner flask (Fisher) with 75mL of M3BPYE media with 5% heat inactivated fetal bovine serum. Cells are kept growing in suspension at a density of 1 million to 3 million cells per mL. To synchronize, 0.2nM of Ponasterone A (Sigma) is added to the suspension culture. After 24 hours, the S2 cells are spun down at 800g for 5 minutes, washed once with 1x PBS, and then resuspended in fresh media containing 1.5mM Hydroxyurea (Sigma). After 18 hours the cells are spun down, washed with 1x PBS, and resuspended in fresh media only. Cells are collected after 2.5 hours in fresh media.

S2 cell culture

S2 cells were cultured in M3BPYE media supplemented with 5% heat inactivated fetal bovine serum.

In vitro sgRNA transcription and purification

The 19 base pair targeted DNA sequence is inserted into the middle of a 58 base pair primer behind a T7 promoter sequence (5'-TTAATACGACTCACTATAGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGC-3'). The custom primer is then used with a reverse template (5'-AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC- 3') in a DNA polymerase extension reaction to generate a dsDNA template. The dsDNA template is used with the HiScribe T7 High Yield RNA synthesis kit (NEB) to generate single stranded RNA of approximately 100 bases in length. The reaction is DNaseI treated and full length RNA is purified by isolating the correct length after running on a denaturing polyacrylamide gel with 8M urea.

S2 cell immunofluorescence

18mm coverslips are cleaned with MeOH and EtOH washes and then incubated with 0.01% poly-lysine solution in water for 15mins. Cells are grown on poly-lysine treated coverslips until ~70% confluency then fixed with 4% paraformaldehyde in 1x PBS for 10 minutes. The fixed samples are washed 1x PBS, permeabilized with 0.5% Triton X100 in 1x PBS, and blocked with 3% BSA in 1x PBS. Primary antibody is added to the samples in 1x PBS with 0.1% Triton X100 and incubated at 4C overnight. Samples are washed and incubated with the appropriate Alexa Fluorophore secondary for one hour at room temperature. The samples are then washed, briefly incubated with 300nM DAPI, and then prepped with ProLong Gold mounting media (Thermo Fisher) for confocal imaging.

dsRNA preparation and *Drosophila* S2 RNAi knockdown assays

dsRNA templates were generated by placing a T7 promoter in front of PCR primers against an exon region of the targeted gene and performing PCR. The resulting template is visualized and isolated by QIAquick Gel Extraction Kit (Qiagen). 100ng of template DNA is used with the HiScribe T7 High Yield RNA synthesis kit (NEB). The reaction mixture is DNaseI treated and purified using TRIzol reagent (Life Technologies), according to manufacturers' protocol. Resulting RNA is resuspended in water, heated to 65C for 30 minutes, and slowly cooled to room temperature to anneal and make dsRNA. S2 cells are resuspended in serum free M3BPYE and cultured with dsRNA for 30 minutes

at room temperature and then 10% FBS M3BPYE is added to get a final concentration of 3.75% FBS. Cells are incubated at 27C for 72 hours before TRIzol extraction for total RNA.

VIG and VIG2 immunoprecipitation and RT-qPCR

S2 cells are dissociated from flasks, spun down, and lysed in Lysis Buffer (20mM Tris-HCl pH 7.5, 100mM KCl, 5mM MgCl2, 0.5% NP40, 800 units RnaseIN per mL). Lysates are incubated on ice and then spun down at 4C to clear insoluble particles. 100uL of supernatant is taken and added to anti-V5 agarose beads (Sigma) that have been blocked with 5% BSA and are resuspended in 900uL of 1x NT2 buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 1mM MgCl2, 0.05% NP40, 0.04 units RnaseIN per mL). Mixture is rocked overnight at 4C and then resin is washed with 1x NT2 buffer adjusted to 200mM NaCl. 20 units of DnaseI (NEB) is added to washed resin in 1x NT2 buffer with 150mM NaCl and incubated at 37C for 30 minutes. SDS is added to mixture to get 0.1% final concentration and treated with 2.5uL of Proteinase K (Thermo Fisher) at 56C for 1 hour. RNA is isolated by using TRIzol reagent (Life Technologies), according to manufacturers' protocols. cDNA synthesis was performed with 50ug of total RNA using iScript cDNA Synthesis Kit (Bio-Rad) and diluted 10-fold. Real time PCR analysis was carried out with SYBR Select Master Mix for CFX (Life Technologies) using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Gene specific primer sequences are provided in the appendix.

MudPIT mass spectrometry and analysis

The TCA precipitated proteins were urea denatured, reduced, alkylated, and digested with recombinant endoproteinase Lys-C (Promega) and modified trypsin (Promega)(Florens and Washburn, 2006; Washburn et al., 2001). Peptides were loaded onto 100-μm fused silica (Polymicro Technologies) capillary column packed with 3 cm of 5-μm reverse phase (RP) C18 resin (Aqua, Phenomenx), 4 cm of 5-μm strong cation exchange resin (Partisphere SCX, Whatman), and 8 cm of RP C18 resin. The loaded microcapillary column was placed in-line with a Quaternary Agilent 1100 series HPLC pump and a LTQ linear ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (ThermoScientific). Ten-step MudPIT was performed on the ionized peptides as described (Florens and Washburn, 2006). Tandem mass (MS/MS) spectra were interpreted using ProLuCID and searched against a non-redundant protein *D. melanogaster* database (NCBI, 02-20-2013) containing 160 usual contaminants (human keratins, IgGs, and proteolytic enzymes). To estimate false discover rates (FDRs), the amino acid sequence of each non-redundant protein was randomized. Peptide/spectrum matches were sorted and selected using DTASelect (Zhang et al., 2010) with the following criteria set: spectra/peptide matches were retained only if they had a DeltCn of at least 0.8, and minimum XCorr of 1.8 for singly, 2.0 for doubly, and 3.0 for triply charged spectra. Additionally, the peptides had to be minimum 7 amino acids in length and fully tryptic.

Peptide hits from multiple runs were compared using CONTRAST (Tabb et al., 2002). The distributed normalized spectral abundance factors (dNSAF) were used to estimate relative protein levels.

Southern Blot

DNA samples are run on 1% agarose gel in 1x TBE. The gel is incubated in 0.25M HCl for 30 minutes and then in 0.4M NaOH for 30 minutes. The DNA is transferred to Hybond XL membrane (Amersham Pharmacia) overnight through capillary action. The membrane is washed with 2x SSC with Tris-HCl (300mM NaCl, 30mM NaCitrate pH 7, 100mM Tris-HCl pH 7.5) and dried at 50C for 15 minutes. Dried membrane is incubated with Church Buffer (1% BSA, 1mM EDTA, 500mM phosphate buffer, 7% SDS) for 30 minutes at 60C before adding radioactive PCR probes prepared by fill-in reactions. After overnight incubation, the membrane is washed with 2x SSC with 0.2% SDS and 0.2x SSC with 0.2% SDS until background signal is gone. The membrane is dried, wrapped in saran wrap, and exposed to a phosphorimager screen (Kodak) and visualized with the PharosFX Plus (Bio-Rad).

DAVID bioinformatics analysis

GenInfo Identifier is taken from MudPIT mass spectrometry results and converted to UNIPROT Identifiers using UNIPROT ID mapping (www.uniprot.org/uploadlists/). UNIPROT Identifiers are inputted into DAVID Bioinformatics Resources 6.8 (david.ncifcrf.gov/tools.jsp) and UP_KEYWORDS, GOTERM_BP_DIRECT, GOTERM_CC_DIRECT, GOTERM_MF_DIRECT, and INTERPRO annotations are used for functional clustering of the gene list.

PI Stain and cell cycle analysis

Cells are collected and resuspended into 1x PBS and fixed with ice cold 70% EtOH for at least 2 hours. The samples are then washed and with 1x PBS and resuspended into PI/Triton X100 solution (0.1% Triton X100, 0.2mg/mL Rnase A, 0.02mg/mL Propidium Iodide in 1x PBS), incubate at 37C for 15 minutes. Fluorescence is detected using BD LSRFortessa (BD Biosciences).

References

Ahn, J.W., Kim, S., Na, W., Baek, S.J., Kim, J.H., Min, K., Yeom, J., Kwak, H., Jeong, S., Lee, C., et al. (2015). SERBP1 affects homologous recombination-mediated DNA repair by regulation of CtIP translation during S phase. Nucleic Acids Res. *43*, 6321–6333.

Albright, S.R., and Tjian, R. (2000). TAFs revisited: More data reveal new twists and confirm old ideas. Gene *242*, 1–13.

Amati, B., Dalton, S., Brooks, M.W., Littlewood, T.D., Evan, G.I., and Land, H. (1992). Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. Nature *359*, 423–426.

Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. Nature *497*, 80–85.

Antao, J.M., Mason, J.M., Dejardin, J., and Kingston, R.E. (2012). Protein landscape at Drosophila melanogaster telomere-associated sequence repeats. Mol Cell Biol *32*, 2170–2182.

Birchmeier, C., Folk, W., and Birnstiel, M.L. (1983). The terminal RNA stem-loop structure and 80 bp of spacer DNA are required for the formation of 3??? termini of sea urchin H2A mRNA. Cell *35*, 433–440.

Bond, U.M., Yario, T.A., and Steitz, J.A. (1991). Multiple processing-defective mutations in a mammalian histone pre-mRNA are suppressed by compensatory changes in U7 RNA both in vivo and in vitro. Genes Dev. *5*, 1709–1722.

Boyer, L. a, Mathur, D., and Jaenisch, R. (2006). Molecular control of pluripotency. Curr. Opin. Genet. Dev. *16*, 455–462.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. a, et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853.

Byrum, S.D., Raman, A., Taverna, S.D., and Tackett, A.J. (2012). ChAP-MS: A method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep. *2*, 198–205.

Carvalho, A.B., Vicoso, B., Russo, C.A.M., Swenor, B., and Clark, A.G. (2015). Birth of a new gene on the Y chromosome of Drosophila melanogaster. Proc. Natl. Acad.

Sci. U. S. A. *112*, 12450–12455.

Caudy, A.A., Myers, M., Hannon, G.J., and Hammond, S.M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev. *16*, 2491–2496.

Caudy, A. a, Ketting, R.F., Hammond, S.M., Denli, A.M., Bathoorn, A.M.P., Tops, B.B.J., Silva, J.M., Myers, M.M., Hannon, G.J., and Plasterk, R.H. a (2003). A micrococcal nuclease homologue in RNAi effector complexes. Nature *425*, 411–414.

Cerna, D., and Wilson, D.K. (2005). The structure of Sif2p, a WD repeat protein functioning in the SET3 corepressor complex. J. Mol. Biol. *351*, 923–935.

Chai, W., Ford, L.P., Lenertz, L., Wright, W.E., and Shay, J.W. (2002). Human Ku70/80 associates physically with telomerase through interaction with hTERT. J. Biol. Chem. *277*, 47242–47247.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Chew, J., Loh, Y., Zhang, W., Chen, X., Tam, W., Yeap, L., Li, P., Ang, Y., Robson, P., Ng, H., et al. (2006). Reciprocal Transcriptional Regulation of Complex in Embryonic Stem Cells Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4 / Sox2 Complex in Embryonic Stem Cells. Mol. Cell. Biol. *25*, 6031–6046.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320.

Courey, A.J., Holtzman, D.A., Jackson, S.P., and Tjian, R. (1989). Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell *59*, 827–836.

Davis, A.J., Chi, L., So, S., Lee, K.J., Mori, E., Fattah, K., Yang, J., and Chen, D.J. (2014). BRCA1 modulates the autophosphorylation status of DNA-PKcs in S phase of the cell cycle. Nucleic Acids Res. *42*, 11487–11501.

Déjardin, J., and Kingston, R.E. (2009). Purification of Proteins Associated with Specific Genomic Loci. Cell *136*, 175–186.

Deng, W., Shi, X., Tjian, R., Lionnet, T., and Singer, R.H. (2015). CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. Proc. Natl. Acad. Sci. U. S. A. *112*, 11870–11875.

Dominski, Z., and Marzluff, W.F. (1999). Formation of the 3′ end of histone mRNA. Gene *239*, 1–14.

Dominski, Z., Yang, X.C., Raska, C.S., Santiago, C., Borchers, C.H., Duronio, R.J., and Marzluff, W.F. (2002). 3' end processing of Drosophila melanogaster histone pre-mRNAs: requirement for phosphorylated Drosophila stem-loop binding protein and coevolution of the histone pre-mRNA processing system. Mol Cell Biol *22*, 6648–6660.

Duronio, R.J., and Marzluff, W.F. (2017). Coordinating cell cycle-regulated histone gene expression through assembly and function of the Histone Locus Body. RNA Biol. *0*, 1–13.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Fletcher, C., Heintz, N., and Roeder, R.G. (1987). Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. Cell *51*, 773–781.

Florens, L., and Washburn, M.P. (2006). Proteomic analysis by multidimensional protein identification technology. New Emerg. Proteomic Tech. Methods Mol. Biol. (Clifton, N.J.) *328*, 159–175.

Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA Repair Complex Functions as an Oct4/Sox2 Coactivator in Embryonic Stem Cells. Cell *147*, 120–131.

Frey, M.R., and Matera, A.G. (1995). Coiled bodies contain U7 small nuclear RNA and associate with specific DNA sequences in interphase human cells. Proc. Natl. Acad. Sci. U. S. A. *92*, 5915–5919.

Fujita, T., and Fujii, H. (2011). Direct identification of insulator components by insertional chromatin immunoprecipitation. PLoS One *6*, 4–9.

Fujita, T., and Fujii, H. (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochem. Biophys. Res. Commun. *439*, 132–136.

Gallia, G.L., Johnson, E.M., and Khalili, K. (2000). Puralpha: a multifunctional single-stranded DNA- and RNA-binding protein. Nucleic Acids Res. *28*, 3197–3205.

Gauthier, L.R., Granotier, C., Hoffschir, F., Etienne, O., Ayouaz, A., Desmaze, C., Mailliet, P., Biard, D.S., and Boussin, F.D. (2012). Rad51 and DNA-PKcs are involved in the generation of specific telomere aberrations induced by the quadruplex ligand 360A that impair mitotic cell progression and lead to cell death. Cell. Mol. Life Sci. *69*, 629–640.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-

Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Golldack, D., Lüking, I., and Yang, O. (2011). Plant tolerance to drought and salinity: Stress regulating transcription factors and their functional significance in the cellular transcriptional network. Plant Cell Rep. *30*, 1383–1391.

Gracheva, E., Dus, M., and Elgin, S.C.R. (2009). Drosophila RISC component VIG and its homolog Vig2 impact heterochromatin formation. PLoS One *4*, 1–11.

Griesenbeck, J., Boeger, H., Strattan, J.S., and Kornberg, R.D. (2003). Affinity purification of specific chromatin segments from chromosomal loci in yeast. Mol. Cell. Biol. *23*, 9275–9282.

Grolimund, L., Aeby, E., Hamelin, R., Armand, F., Chiappe, D., Moniatte, M., and Lingner, J. (2013). A quantitative telomeric chromatin isolation protocol identifies different telomeric states. Nat. Commun. *4*, 2848.

Gu, P., Goodwin, B., Chung, A.C., Xu, X., Wheeler, D.A., Price, R.R., Galardi, C., Peng, L., Latour, A.M., Koller, B.H., et al. (2005). Orphan Nuclear Receptor LRH-1 Is Required To Maintain Oct4 Expression at the Epiblast Stage of Embryonic Development Orphan Nuclear Receptor LRH-1 Is Required To Maintain Oct4 Expression at the Epiblast Stage of Embryonic Development. Mol. Cell. Biol. *25*, 3492–3505.

Guglielmi, B., LaRochelle, N., and Tjian, R. (2013). Gene-specific transcriptional mechanisms at the histone gene cluster revealed by single-cell imaging. Mol. Cell *51*, 480–492.

Gunjan, A., and Verreault, A. (2003). A Rad53 Kinase-Dependent Surveillance Mechanism that Regulates Histone Protein Levels in S. cerevisiae. Cell *115*, 537–549.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a Link Between Genetic and Biochemical Analyses of RNAi. Science (80-.). *293*, 1146–1150.

Hamperl, S., Brown, C.R., Perez-Fernandez, J., Huber, K., Wittner, M., Babl, V., Stöckl, U., Boeger, H., Tschochner, H., Milkereit, P., et al. (2014). Purification of Specific Chromatin Domains from Single-Copy Gene Loci in Saccharomyces cerevisiae. J.C. Stockert, J. Espada, and A. Blázquez-Castro, eds. (Totowa, NJ: Humana Press), pp. 329–341.

Harris, M.E., Bohni, R., Schneiderman, M.H., Ramamurthy, L., Schumperli, D., and Marzluffl, W.F. (1991). Regulation of Histone mRNA in the Unperturbed Cell Cycle: Evidence Suggesting Control at Two Posttranscriptional Steps. Mol. Cell. Biol. *11*, 2416–2424.

Heaton, J.H., Dlakic, W.M., Dlakic, M., and Gelehrter, T.D. (2001). Identification

and cDNA Cloning of a Novel RNA-binding Protein that Interacts with the Cyclic Nucleotide-responsive Sequence in the Type-1 Plasminogen Activator Inhibitor mRNA. J. Biol. Chem. *276*, 3341–3347.

Hentschel, C.C., and Birnstiel, M.L. (1981). The organization and expression of histone gene families. Cell *25*, 301–313.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hindley, C., and Philpott, A. (2012). Co-ordination of cell cycle and differentiation in the developing nervous system. Biochem. J. *444*, 375–382.

Hirsch, J.D., Eslamizar, L., Filanoski, B.J., Malekzadeh, N., Haugland, R.P., Beechem, J.M., and Haugland, R.P. (2002). Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: Uses for protein labeling, detection, and isolation. Anal. Biochem. *308*, 343–357.

Ho, L., Miller, E.L., Ronan, J.L., Ho, W.Q., Jothi, R., and Crabtree, G.R. (2011). esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. Nat. Cell Biol. *13*, 903–913.

Hořejší, Z., Takai, H., Adelman, C.A., Collis, S.J., Flynn, H., Maslen, S., Skehel, J.M., de Lange, T., and Boulton, S.J. (2010). CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability. Mol. Cell *39*, 839–850.

Huang, D.W., Lempicki, R. a, and Sherman, B.T. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Hutvágner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056–2060.

Ide, S., and Dejardin, J. (2015). End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. Nat. Commun. *6*, 6674.

Im, S.H., and Lee, J. (2005). PLP-1 binds nematode double-stranded telomeric DNA. Mol. Cells *20*, 297–302.

Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007). Transcription of histone gene cluster by differential core-promoter factors. Genes Dev. *21*, 2936–2949.

Ito, S., Fujiyama-Nakamura, S., Kimura, S., Lim, J., Kamoshida, Y., Shiozaki-Sato, Y., Sawatsubashi, S., Suzuki, E., Tanabe, M., Ueda, T., et al. (2012). Epigenetic Silencing of Core Histone Genes by HERS in Drosophila. Mol. Cell *45*, 494–504.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Jasinskas, A., and Hamkalo, B.A. (1999). Purification and Initial Characterization of Primate Satellite Chromatin. Chromosom. Res. *7*, 341–354.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity. Science *337*, 816–822.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2013*, 1–9.

Kadonaga, J.T., and Tjian, R. (1986). Affinity purification of sequence-specific DNA binding proteins. Proc. Natl. Acad. Sci. U. S. A. *83*, 5889–5893.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell *51*, 1079–1090.

Kim, B., and Little, J.W. (1992). Dimerization of a specific DNA-binding protein on the DNA. Science *255*, 203–206.

Kim, M.H., Cooper, D.R., Oleksy, A., Devedjiev, Y., Derewenda, U., Reiner, O., Otlewski, J., and Derewenda, Z.S. (2004). The structure of the N-terminal domain of the product of the lissencephaly gene Lis1 and its functional implications. Structure *12*, 987–998.

Kremer, H., and Hennig, W. (1990). Isolation and characterization of a Drosophila hydei histone DNA repeat unit. Nucleic Acids Res. *18*, 1573–1586.

Lee, M.C., Toh, L.L., Yaw, L.P., and Luo, Y. (2010). Drosophila octamer elements and Pdm-1 dictate the coordinated transcription of core histone genes. J. Biol. Chem. *285*, 9041–9053.

Lei, E.P., and Corces, V.G. (2006). RNA interference machinery influences the nuclear organization of a chromatin insulator. Nat. Genet. *38*, 936–941.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: Transcription enters a new era. Cell *157*, 13–25.

Lifton, R.P., Goldberg, M.L., Karp, R.W., and Hogness, D.S. (1976). The organization of the histone genes in drosophila melanogaster: functional and evolutionary implications.

Liu, J.L., Murphy, C., Buszczak, M., Clatterbuck, S., Goodman, R., and Gall, J.G. (2006). The Drosophila melanogaster Cajal body. J. Cell Biol. *172*, 875–884.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

Lonze, B.E., and Ginty, D.D. (2002). Function and regulation of CREB family transcription factors in the nervous system. Neuron *35*, 605–623.

Los, G. V, Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., et al. (2008). HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. ACS Chem. Biol. *3*, 373–382.

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

Lustig, a J., and Petes, T.D. (1986). Identification of yeast mutants with altered telomere structure. Proc. Natl. Acad. Sci. U. S. A. *83*, 1398–1402.

Ma, T., Van Tine, B.A., Wei, Y., Garrett, M.D., Nelson, D., Adams, P.D., Wang, J., Qin, J., Chow, L.T., and Harper, J.W. (2000). Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. Genes Dev. *14*, 2298–2313.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-Guided Human Genome Engineering via Cas9. Science (80-.). *339*, 823–826.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. *78*, 7634–7638.

Martin, F., Schaller, A., Eglite, S., Schü, D., and Mü Ller, B. (1997). The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein and (ii) a highly conserved 26 bp sequence encompassing a not essential for maximal processing efficiency (Mowry. EMBO J. *16*, 769–778.

Marzluff, W.F. (2005). Metazoan replication-dependent histone mRNAs: A distinct set of RNA polymerase II transcripts. Curr. Opin. Cell Biol. *17*, 274–280.

Marzluff, W.F., and Duronio, R.J. (2002). Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences William F Marzluff*. 1–8.

Mateja, A., Cierpicki, T., Paduch, M., Derewenda, Z.S., and Otlewski, J. (2006).

The dimerization mechanism of LIS1 and its implication for proteins containing the LisH motif. J. Mol. Biol. *357*, 621–631.

Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.-P., St-Denis, N.A., Li, T., Miteva, Y. V, Hauri, S., Sardiu, M.E., Low, T.Y., et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat. Methods *10*, 730–736.

Mikolajka, A., Yan, X., Popowicz, G.M., Smialowski, P., Nigg, E.A., and Holak, T.A. (2006). Structure of the N-terminal Domain of the FOP (FGFR1OP) Protein and Implications for its Dimerization and Centrosomal Localization. J. Mol. Biol. *359*, 863–875.

Minucci, S., Botquin, V., Yeom, Y.I., Dey, a, Sylvester, I., Zand, D.J., Ohbo, K., Ozato, K., and Scholer, H.R. (1996). Retinoic acid-mediated down-regulation of Oct3/4 coincides with the loss of promoter occupancy in vivo. EMBO J. *15*, 888–899.

Mowry, K.L., and Steitz, J.A. (1987). Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNA's. Science *238*, 1682–1687.

Nordhoff, V., Hübner, K., Bauer, A., Orlova, I., Malapetsa, A., and Schöler, H.R. (2001). Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences. Mamm. Genome *12*, 309–317.

Ohta, S., Bukowski-Wills, J.C., Sanchez-Pulido, L., Alves, F. de L., Wood, L., Chen, Z.A., Platani, M., Fischer, L., Hudson, D.F., Ponting, C.P., et al. (2010). The Protein Composition of Mitotic Chromosomes Determined Using Multiclassifier Combinatorial Proteomics. Cell *142*, 810–821.

Okamoto, K., Okazawa, H., Okuda, a, Sakai, M., Muramatsu, M., and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell *60*, 461–472.

Ovitt, C.E., and Schöler, H.R. (1998). The molecular biology of Oct-4 in the early mouse embryo. Mol. Hum. Reprod. *4*, 1021–1031.

Palm, W., and de Lange, T. (2008). How Shelterin Protects Mammalian Telomeres. Annu. Rev. Genet. *42*, 301–334.

Pestova, T. V., and Kolupaeva, V.G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. Genes Dev. *16*, 2906–2922.

PETERSON, C., and TAMKUN, J. (1995). The SWI-SNF complex: a chromatin remodeling machine? Trends Biochem. Sci. *20*, 143–146.

Pillai, R.S., Will, C.L., Lührmann, R., Schümperli, D., and Müller, B. (2001). Purified

U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. EMBO J. *20*, 5470–5479.

Pourfarzad, F., Aghajanirefah, A., deBoer, E., TenHave, S., BrynvanDijk, T., Kheradmandkia, S., Stadhouders, R., Thongjuea, S., Soler, E., Gillemans, N., et al. (2013). Locus-specific proteomics by TChP: Targeted chromatin purification. Cell Rep. *4*, 589–600.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. Nature *311*, 532–537.

Rozario, D., and Siede, W. (2012). Saccharomyces cerevisiae Tel2 plays roles in TORC signaling and telomere maintenance that can be mutationally separated. Biochem. Biophys. Res. Commun. *417*, 1182–1187.

Sànchez, R., Marzluff, W.F., and Sa, R. (2002). The Stem-Loop Binding Protein Is Required for Efficient Translation of Histone mRNA In Vivo and In Vitro The Stem-Loop Binding Protein Is Required for Efficient Translation of Histone mRNA In Vivo and In Vitro. Society *22*, 7093–7104.

Sawyer, I.A., and Dundr, M. (2016). Nuclear bodies: Built to boost. J. Cell Biol. *213*, 509–511.

Singh, R.K., Liang, D., Gajjalaiahvari, U.R., Kabbaj, M.H.M., Paik, J., and Gunjan, A. (2010). Excess histone levels mediate cytotoxicity via multiple mechanisms. Cell Cycle *9*, 4236–4244.

Singhal, N., Graumann, J., Wu, G., Araúzo-Bravo, M.J., Han, D.W., Greber, B., Gentile, L., Mann, M., and Schöler, H.R. (2010). Chromatin-remodeling components of the baf complex facilitate reprogramming. Cell *141*, 943–955.

Siomi, H., and Dreyfuss, G. (1997). RNA-binding proteins as regulators of gene expression. Curr. Opin. Genet. Dev. *7*, 345–353.

Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J. *11*, 4047–4057.

So, M. kyung, Yao, H., and Rao, J. (2008). HaloTag protein-mediated specific labeling of living cells with quantum dots. Biochem. Biophys. Res. Commun. *374*, 419–423.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. Cell *136*, 731–745.

Spycher, C., Streit, A., Stefanovic, B., Albrecht, D., Koning, T.H.W., and Schümperli, D. (1994). 3′ End processing of mouse histone pre-mRNA: evidence for additional base-pairing between U7 snRNA and pre-mRNA. Nucleic Acids Res. *22*, 4023–4030.

van Steensel, B., Delrow, J., and Henikoff, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase. Nat. Genet. *27*, 304–308.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62–67.

Sui, J., Lin, Y.F., Xu, K., Lee, K.J., Wang, D., and Chen, B.P.C. (2015). DNA-PKcs phosphorylates hnRNP-A1 to facilitate the RPA-to-POT1 switch and telomere capping after replication. Nucleic Acids Res. *43*, 5971–5983.

Taatjes, D.J., Marr, M.T., and Tjian, R. (2004). Regulatory diversity among metazoan co-activator complexes. Nat. Rev. Mol. Cell Biol. *5*, 403–410.

Tabb, D.L., McDonald, W.H., and Yates, J.R. (2002). DTASelect and Contrast: Tools for Assembling and Comparing Protein Identifications from Shotgun Proteomics. J. Proteome Res. *1*, 21–26.

Takai, H., Wang, R.C., Takai, K.K., Yang, H., and de Lange, T. (2007). Tel2 Regulates the Stability of PI3K-Related Protein Kinases. Cell *131*, 1248–1259.

Takai, K.K., Hooper, S., Blackwood, S., Gandhi, R., and de Lange, T. (2010). In vivo stoichiometry of shelterin components. J. Biol. Chem. *285*, 1457–1467.

Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C., and Kennison, J.A. (1992). brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2SWI2. Cell *68*, 561–572.

Tatomer, D.C., Terzo, E., Curry, K.P., Salzler, H., Sabath, I., Zapotoczny, G., McKay, D.J., Dominski, Z., Marzluff, W.F., and Duronio, R.J. (2016). Concentrating pre-mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis. J. Cell Biol. *213*, 557–570.

Terzo, E. a, Lyons, S.M., Poulton, J.S., Temple, B.R.S., Marzluff, W.F., and Duronio, R.J. (2015). Distinct self-interaction domains promote Multi Sex Combs accumulation in and formation of the Drosophila histone locus body. Mol. Biol. Cell *26*, 1559–1574.

Tesar, P.J., Chenoweth, J.G., Brook, F. a, Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Thomas, M.C., and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

Ting, N.S.Y., Pohorelic, B., Yu, Y., Lees-Miller, S.P., and Beattie, T.L. (2009). The human telomerase RNA component, hTR, activates the DNA-dependent protein kinase to phosphorylate heterogeneous nuclear ribonucleoprotein A1. Nucleic Acids Res. *37*, 6105–6115.

Tomari, Y. (2005). Perspective: machines for RNAi. Genes Dev. *19*, 517–529.

Townley-Tilson, W.H.D., Pendergrass, S.A., Marzluff, W.F., and Whitfield, M.L. (2006). Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. RNA *12*, 1853–1867.

Unnikrishnan, A., Gafken, P.R., and Tsukiyama, T. (2010). Dynamic changes in histone acetylation regulate origins of DNA replication. Nat. Struct. Mol. Biol. *17*, 430–437.

Vincenz, C., Fronk, J., Tank, G.A., and Langmore, J. (1991). Nucleoprotein hybridization: A method for isolating active and inactive genes as chromatin. Nucleic Acids Res. *19*, 1325–1336.

Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nat. Protoc. *2*, 1467–1478.

Waldrip, Z.J., Byrum, S.D., Storey, A.J., Gao, J., Byrd, A.K., Mackintosh, S.G., Wahls, W.P., Taverna, S.D., Raney, K.D., and Tackett, A.J. (2014). A CRISPR-based approach for proteomic analysis of a single genomic locus. Epigenetics *9*, 1207–1211.

Walther, T.C., and Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. J. Cell Biol. *190*, 491–500.

Washburn, M.P., Wolters, D., and Yates, J.R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. *19*, 242–247.

White, A.E., Burch, B.D., Yang, X.C., Gasdaska, P.Y., Dominski, Z., Marzluff, W.F., and Duronio, R.J. (2011). Drosophila histone locus bodies form by hierarchical recruitment of components. J. Cell Biol. *193*, 677–694.

Whitfield, M.L., Zheng, L.X., Baldwin, a, Ohta, T., Hurt, M.M., and Marzluff, W.F. (2000). Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms. Mol. Cell. Biol. *20*, 4188–4198.

Wierer, M., and Mann, M. (2016). Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. Hum. Mol. Genet. *25*, R106–R114.

Workman, J.L., and Langmore, J.P. (1985). Nucleoprotein hybridization: a method for isolating specific genes as high molecular weight chromatin. Biochemistry *24*, 7486–7497.

Wortman, M.J., Johnson, E.M., and Bergemann, A.D. (2005). Mechanism of DNA binding and localized strand separation by Purα and comparison with Pur family member, Purβ. Biochim. Biophys. Acta - Mol. Cell Res. *1743*, 64–78.

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., and Felsenfeld, G. (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes Dev. *24*, 2543–2555.

Yeo, J.-C., and Ng, H.-H. (2013). The transcriptional regulation of pluripotency. Cell Res. *23*, 20–32.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, a, Ohbo, K., Gross, M., Hübner, K., and Schöler, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*, 881–894.

Zhang, Y., Wen, Z., Washburn, M.P., and Florens, L. (2010). Refinements to label free proteome quantitation: How to deal with peptides shared by multiple proteins. Anal. Chem. *82*, 2272–2281.

Zhao, J., Kennedy, B.K., Lawrence, B.D., Barbie, D.A., Gregory Matera, A., Fletcher, J.A., and Harlow, E. (2000). NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. Genes Dev. *14*, 2283–2297.

Zheng, L., Roeder, R.G., and Luo, Y. (2003). S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. Cell *114*, 255–266.

Zhou, X., Zhang, X., Xie, Y., Tanaka, K., Wang, B., and Zhang, H. (2013). DNA-PKcs Inhibition Sensitizes Cancer Cells to Carbon-Ion Irradiation via Telomere Capping Disruption. PLoS One *8*, 1–7.

Appendix A

The Identification of Novel Activator Binding Site
in the Proximal Enhancer of Oct4

Abstract

Oct4 is one of the three master regulators of the mammalian stem cell state and its role in establishing the stem cell identity through transcription regulation has been studied in detail. When epiblast cells were successfully cultured *in vitro*, researchers were able to confirm that OCT4 is also necessary to maintain the undifferentiated epiblast state. Though OCT4 plays a similar role in naïve embryonic stem cell and epiblast stem cells, the regulation of *Oct4* gene expression dramatically changes from activation through the well known distal enhancer in embryonic stem cells to the less understood proximal enhancer in epiblast cells. How this enhancer switching occurs and what activators mediate this switching is currently unknown. To better describe the DNA sequences responsible for the proximal enhancer activation, I took an enhancer bashing approach to dissect the *Oct4* proximal enhancer in a cell type that has been previously described to preferentially use the proximal enhancer for *Oct4* activation. With detailed mutagenesis analysis and transcription activation assays, I was able to isolate a novel potential activator binding site in the *Oct4* proximal enhancer.

Introduction

Embryonic stem cells (ESCs) possess the unique characteristics of pluripotency and self-renewal, making them an ideal starting point for regenerative medicine and biomedical studies (Boyer et al., 2006; Yeo and Ng, 2013). Understanding the mechanisms of pluripotency is a necessary step towards using ESCs to their full potential. Towards that goal, mouse embryonic stem cells (mESCs) have been the work horse in pluripotency studies due to both the ease of culture conditions since being isolated more than three decades ago (Evans and Kaufman, 1981; Martin, 1981). *In vivo*, cultured mESCs reflect the inner cell mass (ICM) during the blastocyst stage of embryonic development. Within an additional 24 hours of development, the blastocyst implants into the uterine wall and the ICM differentiates into a distinct cell type called the epiblast, which gives rise to all cell types of the eventual organism (Ovitt and Schöler, 1998). Epiblast cells are also pluripotent and share many major characteristics with mESCs, such as the expression of key pluripotency genes OCT4, SOX2, and NANOG. The recent *in vitro* establishment of epiblast cell lines (Brons et al., 2007; Tesar et al., 2007), commonly referred to as epiblast stem cells (EpiSCs), paved the way for researchers to unravel the differences between these two distinct and, yet equally pluripotent, cell types.

Significant progress has been made toward identifying factors governing pluripotency. OCT4 (also known as POU5F1) in combination with SOX2 and NANOG have been shown to be the crucial transcription factors for controlling numerous pathways connected to pluripotency, self-renewal, and cell fate determination (Loh et al., 2006). Despite our continued progress on understanding how OCT4 regulates a vast array of genes in stem cells, the specific mechanism for the transcription regulation of the *Oct4* gene between the two *in vivo* pluripotent cell types is not well understood.

The transcription control of *Oct4* was first described in 1996 (Yeom et al., 1996). The 2.8 kilobase upstream region of the *Oct4* gene can be broken up into two distinct regions; the distal enhancer (DE) and the proximal enhancer (PE). *In vivo* enhancer-lac-Z fusion transgenic studies have shown distinct expression patterns for the two enhancers in the two pluripotent tissues of the early mouse embryo; the DE drives expression in the ICM at the blastocyst stage and the PE drives expression in the epiblast (Yeom et al., 1996). Interestingly, mouse embryonal carcinoma cell lines named F9 and P19 also show *Oct4* activation using the distal and proximal enhancer, respectively (Yeom et al., 1996).

How the proximal enhancer is activated has been a compelling question ever since the discovery that human embryonic stem cells activate the *Oct4* gene through the proximal enhancer instead of the distal enhancer (Yeom et al., 1996). Various proteins have been proposed as the factor responsible, such as PRDM14 and LRH1/ SF1, however PRDM14 is not present in mouse EpiSC, and null mutant mice have no early embryonic phenotype (Chia et al., 2010; Gu et al., 2005). While LRH1 is expressed in EpiSCs, it is also expressed in ESCs where the PE is not active. Our exploratory experiments to determine if it is responsible for proximal enhancer activation in P19 or F9 cells suggest that it is not the sole factor responsible. SF1 is a related family member to LRH1 proposed to be responsible for PE activation in P19 cells, but our knockdown experiments do not produce an effect on the activation of a proximal enhancer reporter

construct in P19 cells. Furthermore, SF1 overexpression constructs do not have any activation effect on proximal enhancer reporters in F9 cells despite having a positive transcription effect on the *Oct4* promoter reporter which contains a described SF1 binding site (FIG 3 and FIG 4) (Gu et al., 2005). In this chapter, I describe my efforts in understanding the regulatory elements dictating proximal enhancer activation of *Oct4*.

Results

A reliable way to identify DNA binding factors that are responsible for transcription activation of an enhancer or promoter is to narrow the region of interest down to the actual binding sites so one can multimerize it for DNA affinity purification or *in vitro* transcription assays for screening (Fong et al., 2011; Kadonaga and Tjian, 1986). So we took a similar approach with the distal and proximal enhancer of *Oct4* in mouse cells by utilizing enhancer bashing coupled with luciferase assays as an output. F9 and P19 EC cell lines display preferential enhancer usage when the distal and proximal enhancer of *Oct4* is linked to a reporter gene. Therefore, these two cell lines are used as proxies of the molecular environment necessary for the activation of the respective enhancers (Minucci et al., 1996).

Break Down of the Distal Enhancer

As a positive control, we decided to breakdown the approximately 1600 base pair distal enhancer of *Oct4* despite the key binding sites of OCT4/ SOX2 already described in Chew et al 2005. To determine which parts of the distal enhancer is crucial, we separated the enhancer into three parts, the first encompassing beginning 733 base pairs, the second being the middle 272 base pairs, and the third being the last 626 base pairs (FIG 1A). Luciferase assays in F9 EC cells show that the middle 272 base pairs show the highest enhancer activation. This result is perhaps as expected because it contains the conserved region previously described and also the annotated OCT4/SOX2 binding site (FIG 1B) (Chew et al., 2006; Nordhoff et al., 2001). This conserved region also showed the highest activity in P19 cells when compared to the other two regions (FIG 1C). However, this is not surprising as the main determinant activators of the distal enhancer is presumed to be OCT4 and SOX2 which is also present in P19 EC cells.

We further broke down the middle 266 base pair region of the distal enhancer by separating the putative SP1 binding site from the OCT4/ SOX2 binding site that have been previously described (Chew et al., 2006; Minucci et al., 1996). As expected, the last 137 base pair of the middle region did not provide much activation as it does not contain any putative binding sites (FIG 2A and 2B). But surprisingly, the 96 base pair fragment that contains the identified OCT4/SOX2 binding site does not show any significant activation over the promoter-only control even though the auto regulation of OCT4/ SOX2 protein for the *Oct4* gene has been well described and is thought to be the main driver for the activation of the *Oct4* gene (FIG 2B) (Boyer et al., 2006). The remaining piece of the original 266 base pair distal enhancer contains the G/C rich putative SP1 binding site, however this region alone is also not sufficient to recapitulate the activation levels of the entire 266 base pair region (FIG 2B). One possible reason that the separate parts of the enhancer do not display the activation levels reminiscent of the original could be that we separated binding sites for transcriptional activators that require each other for its full function (Amati et al., 1992). So we combined the described G/C rich binding site with the OCT4/SOX2 site in a 147bp fragment and found it was able to fully recapitulate an activation level similar to the full 266 base pair enhancer (FIG 2B). To ascertain whether

the distal enhancer bashing experiments is specific for *Oct4* distal enhancer activation only in embryonal carcinoma cell lines or is applicable to mouse embryonic stem cells as well, we tested the various constructs in mESCs and found that they reproduce similar results (FIG 5A and 5B).

Break Down of the Proximal Enhancer

To better understand what is responsible for the activation of the proximal enhancer, we took an enhancer bashing approach to distill it down to its essential binding sites. Previous research have published that there is a well conserved region within the first 400 base pairs of the proximal enhancer (Nordhoff et al., 2001). Thus, we split the 930 base pair long enhancer into approximately 400 and 530 base pair fragments. Unsurprisingly, the majority of the enhancer activity resides within the previously described conserved region (FIG 1C). We further dissected the enhancer by dividing it into approximately 200 base pairs each and then an additional construct that further constrict the 3' fragment to the two putative SF1 binding sites previously described (FIG 2A) (Gu et al., 2005). To our surprise, the enhancer fragment that is 30-40 base pairs shorter on each side of the SF1 putative binding sites showed no increased activation compared to the first 200 base pairs of the truncated proximal enhancer that had no putative transcription factor binding sites (FIG 2C). Instead, the putative SF1 binding sites required an additional 30-40 base pairs on each side to have significant activation in P19 cells (FIG 2C). This result suggests that additional DNA sequences are important in the activation of the proximal enhancer in P19 EC cells in addition to the reported SF1 binding sites.

In order to understand which portions of the truncated 200 base pair proximal enhancer is important in determining enhancer activation, we performed an 18 base pair mutagenesis scan across the length of the 200 bases and tested its effect on enhancer activation by luciferase assays in P19 cells (FIG 6A). Although the majority of the mutagenesis had some effect on the enhancer's ability to activate transcription, the most striking effect occurred with the D2.2 and D2.3 mutagenesis right next to the described SF1 binding site (FIG 6B). This surprising result suggest that perhaps the critical binding sites for transcription activation lies outside the described SF1 region. To further narrow down the possible binding sites we performed a 7 base pair mutagenesis scan across the 36 base pair region determined to have the most drastic effect on enhancer activation (FIG 6A). Again, although detailed mutagenesis suggests all mutations had an overall negative effect on enhancer activation, mutations for the middle 16 base pairs stood out as having the most significant decrease in activation (FIG 6C). It has been previously shown that transcription factor binding sites that lead to activation often have synergistic effects upon multimerization (Courey et al., 1989). To test if these potential binding sites bind transcription activators we multimerized the first 25 base pairs, the last 25 base pairs, and the middle 16 base pairs four times in a row for luciferase assays with a minimal promoter. As expected from the mutagenesis results, multimerizing the binding sites all gave some degree of synergistic activation. While this result suggests that each of the multimerized sequence can interact with a transcription activator, the middle 16 base pair

sequence might best represent the entire binding site as it has the greatest activation when multimerized (FIG 7A)

With a new region of activator dependent sequence found in the proximal enhancer, we tested whether this potential activator remains cell type specific. To our dismay, the multimerized 16 base pair binding site luciferase construct showed high activation in not only P19 cells but also F9, HeLa, and D3 mouse embryonic stem cells (FIG 7B).

Discussion

Oct4 is crucial to the establishment of embryonic stem cells in both human and mouse systems (Okamoto et al., 1990; Yeo and Ng, 2013). And understanding how *Oct4* gene activation is regulated is an essential part of understanding pluripotency and stem cell reprogramming. To this end, we attempted to address the long standing question of what is responsible for activating the *Oct4* gene through the proximal enhancer.

We decided to take a reductionist approach to this question and performed detailed enhancer bashing and mutagenesis to uncover the key DNA sequences responsible for proximal enhancer activation (FIG 6A). We were able to narrow the key sequence down to a 16 base pair region, and to our surprise it is a G/C rich region next to but distinct from the putative SF1/LRH1 binding domain previously described (FIG 6A and 6B) (Gu et al., 2005). Although it seems probable that this region is acting as an binding site for a transcription activator due to its ability to synergistically increase activation upon multimerization, it is likely that the identity of this factor is an ubiquitously expressed transcription factor rather than a cell type specific one (FIG 7B). If the proximal enhancer of *Oct4* does rely, in part, on a ubiquitous activator it brings up the question of why the proximal enhancer is not active in cell types where the distal enhancer is utilized. I postulate that this is because the chromatin context of the proximal enhancer in cell types such as F9's and ESCs impose a repressive context on the proximal enhancer, preventing transcription factors from accessing their respective binding sites. And indeed, a study published in 2014 sees large chromatin reorganization upon the transition from a mESC state to an Epiblast state. This reorganization allows for transcription factors such as OCT4 to bind and access new sites that were previously inaccessible suggesting that a similar mechanism could regulate the appearance of OCT4 binding at the PE after the mESC to epiblast transition (Buecker et al., 2014).

Figure 1.
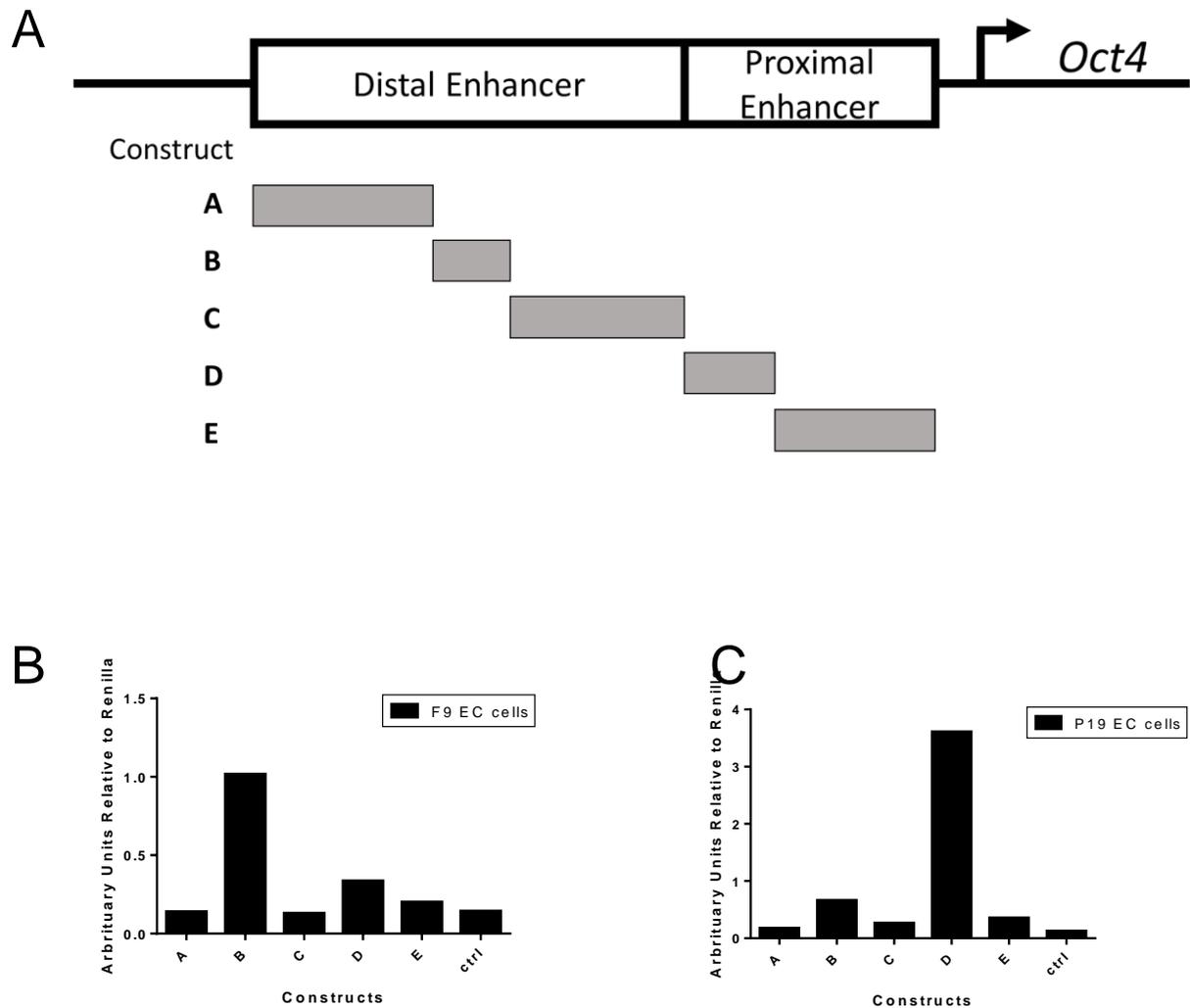
Figure 1. Initial dissection of the canonical distal and proximal enhancer of *Oct4*.
(A) Graphical description of the different enhancer breakdowns. The pieces are
separated based on known binding sites and previously described conserved regions. (B,
C) Truncated enhancers are cloned in front of an SV40 minimal promoter and transfected
into the respective cell lines via Lipofectamine 2000 and lysed to assess luciferase activity
relative to renilla control.

Figure 2.

A
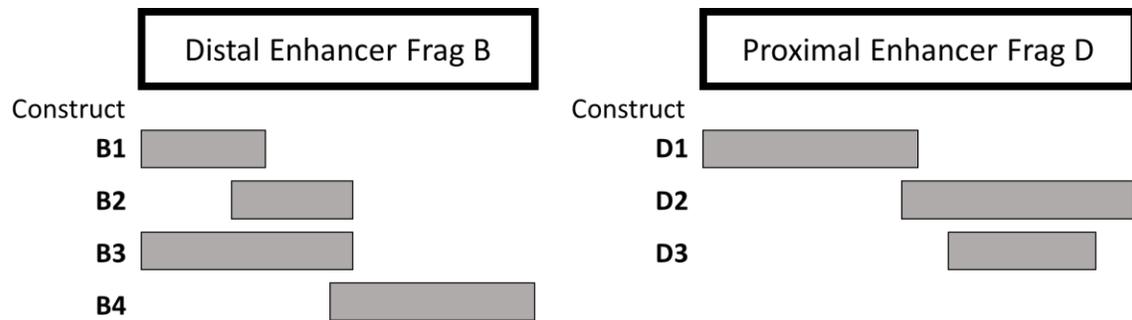
B
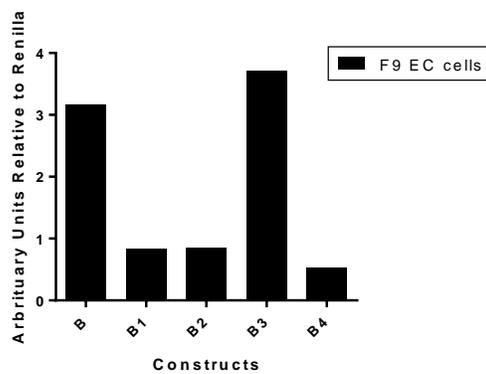
C
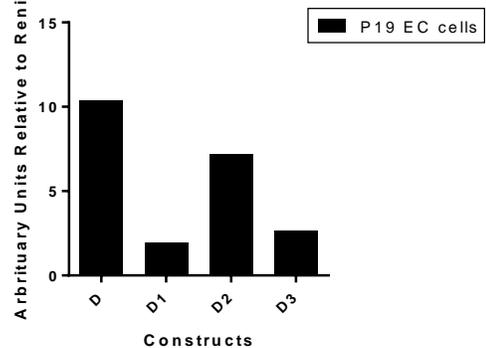
Figure 2. Further breakdown of the active enhancer region. (A) Graphical description of the continued enhancer breakdown for each of the active enhancer region from Figure 1A. **(B, C)** Truncated enhancer regions are cloned in front of a minimal SV40 promoter and transfected into the respective cell lines for luciferase activation analysis. Both the distal and proximal enhancer truncation required regions of the enhancer not included in the previously described binding sites.

Figure 3.
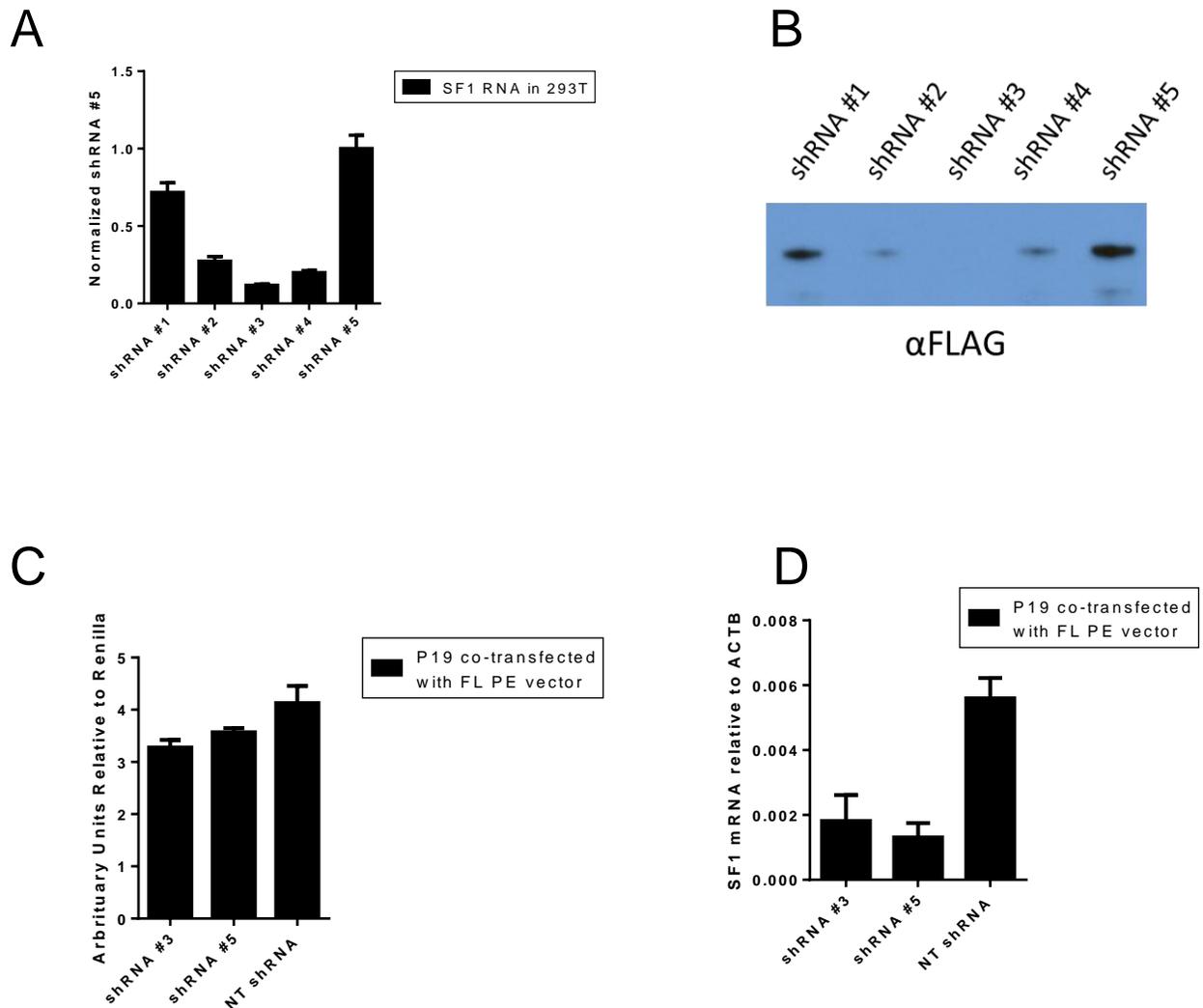
Figure 3. SF1 is not the main protein responsible for driving proximal enhancer activation in P19 EC cells. (A) shRNA knockdown control experiment. SF1 cDNA targged with FLAG is overexpressed in 293T cells by a CMV expression vector alongside the shRNA expression vector. shRNA #1 to 5 target the cDNA sequence while shRNA #5 targets the 5' UTR and serves as a negative control. (B) Cells are transfected with labeled shRNA vector and SF1 cDNA tagged with FLAG peptide. Cells are collected 48 hours after transfection and lysed directly into 1x sample buffer. (C) Full length proximal enhancer luciferase construct is transfected alongside shRNA #3, #5, and non-targeting (NT). Cells are lysed and assessed for luciferase activity relative to renilla control. (D) SF1 mRNA quantification after shRNA knockdown. Cells are transfected with the same constructs as in FIG 3C but RNA is isolated instead and SF1 mRNA levels are measured relative to ACTIN B levels.

Figure 4.

A

B

C
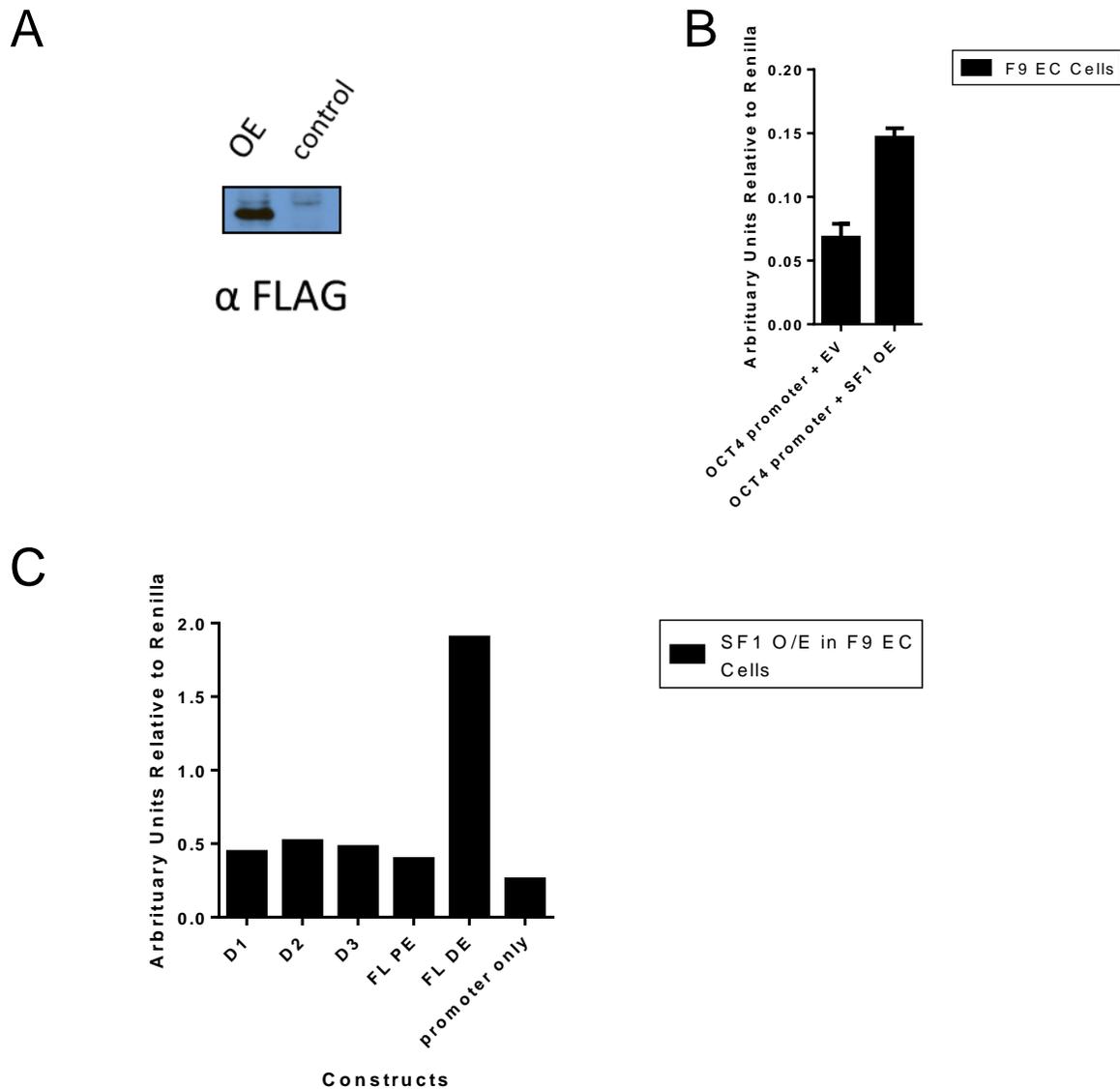
Figure 4. SF1 overexpression in F9 EC cells do not activate proximal enhancer. (A) SF1 cDNA fused to FLAG peptide is overexpressed in F9 EC cells with CMV promoter. SF1 protein overexpression is detected by mouse anti FLAG antibody with a western blot. **(B)** A *Oct4* promoter driving luciferase vector is transfected along with an empty overexpression vector or vector driving SF1 cDNA. SF1 overexpression in F9 EC cells is capable of activating *Oct4* promoter as previously described. **(C)** SF1 cDNA overexpression vector is co-transfected into F9 EC cells with different *Oct4* enhancers driving luciferase. SF1 cDNA overexpression does not have any significant effect on proximal enhancer activation in F9 EC cells.

Figure 5.

A

B
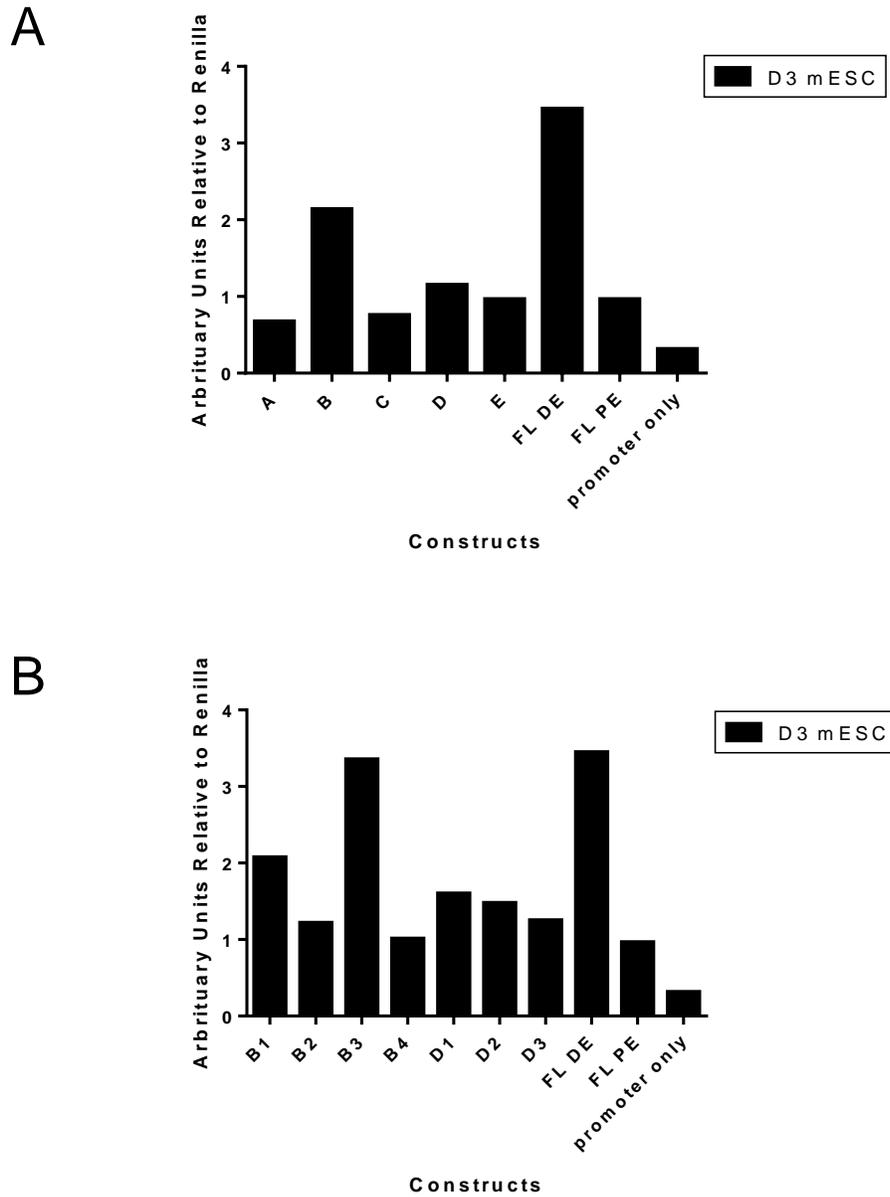
**Figure 5. Distal and proximal enhancer activities are recapitulated in mouse embryonic stem cells. (A, B)** The various enhancer truncations driving luciferase is transfected into D3 mouse embryonic stem cells. The respective activities seen in F9 embryonal carcinoma cells are recapitulated in mouse embryonic stem cells.
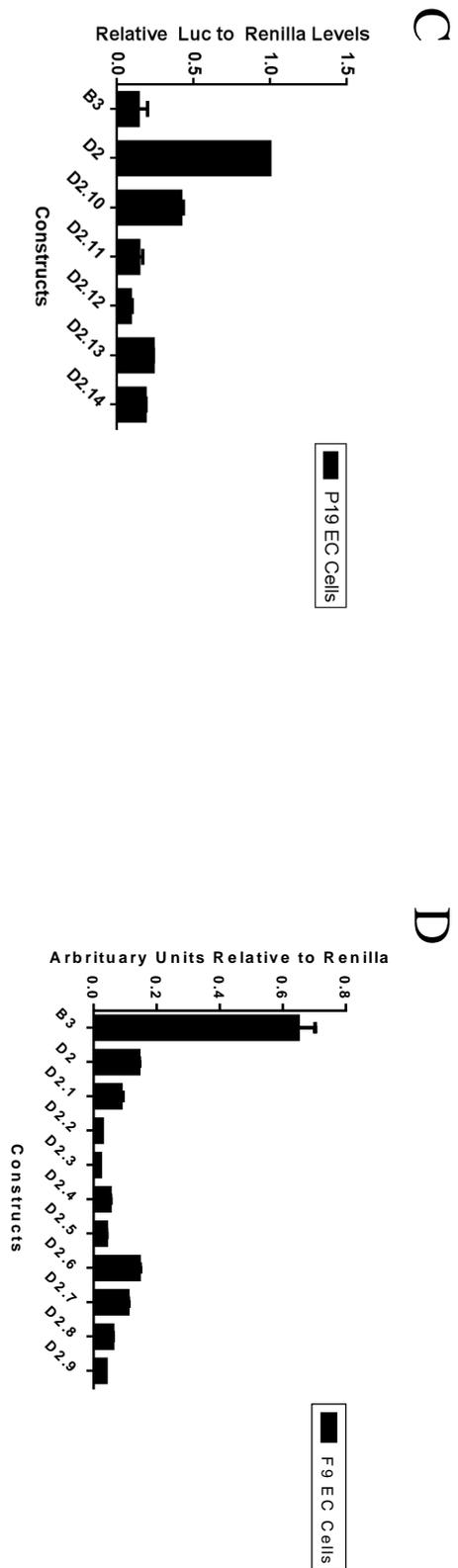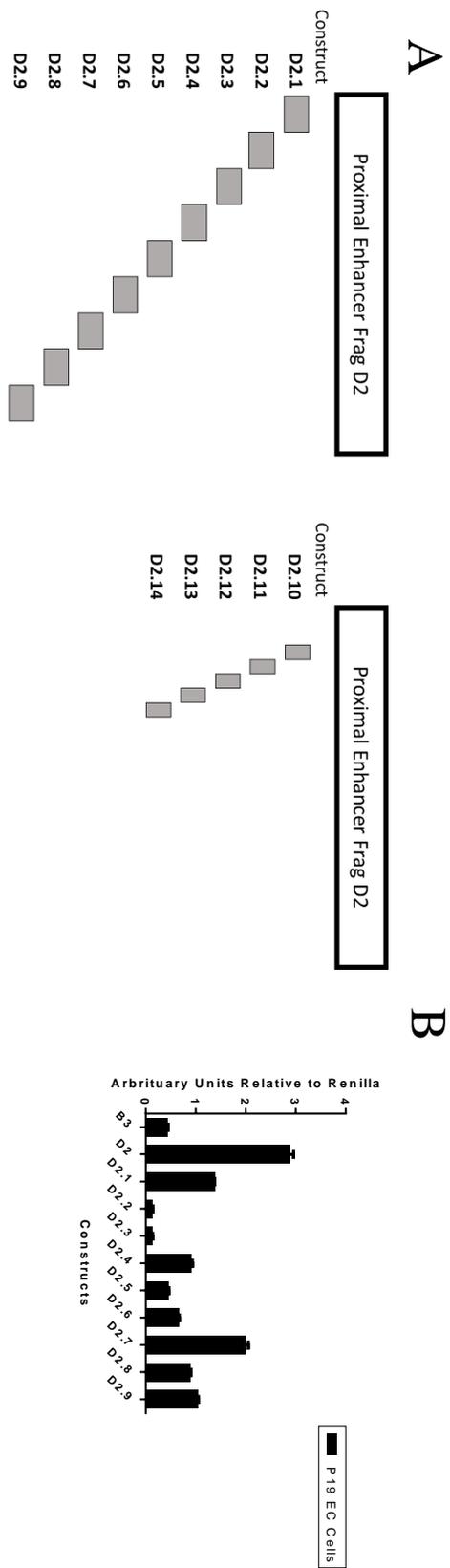
Figure 6.

Figure 6. Detailed mutagenesis scans across the truncated proximal enhancer region reveals a possible activator binding site. (A) A graphical representation of the mutagenesis constructs made across the D2 fragment of the *Oct4* proximal enhancer. A 16-base pair sequence based on the Lambda phage genome is used for tiling mutagenesis across the length of the D2 fragment, depicted by constructs D2.1 to D2.9. A shorter 7bp version is used for a more detailed mutagenesis across the D2.2 and D2.3 region, as depicted by constructs D2.10 to D2.14. **(B)** Mutagenesis constructs D2.1 to D2.9 is transfected into P19 and F9 EC cells and luciferase activity is measured relative to renilla control. **(C)** More detailed mutagenesis constructs, D2.10 to D2.14, is transfected into P19 EC cells and luciferase activity is measured relative to renilla.

Figure 7.

A
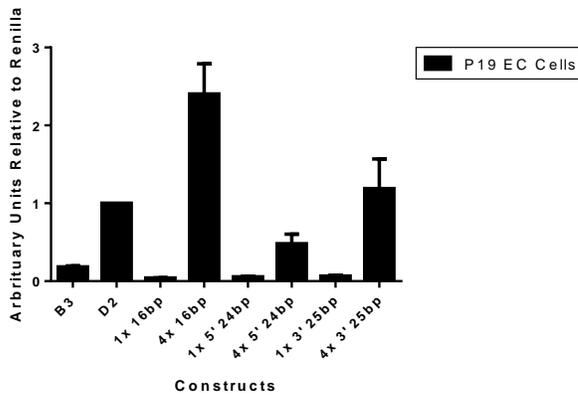
B
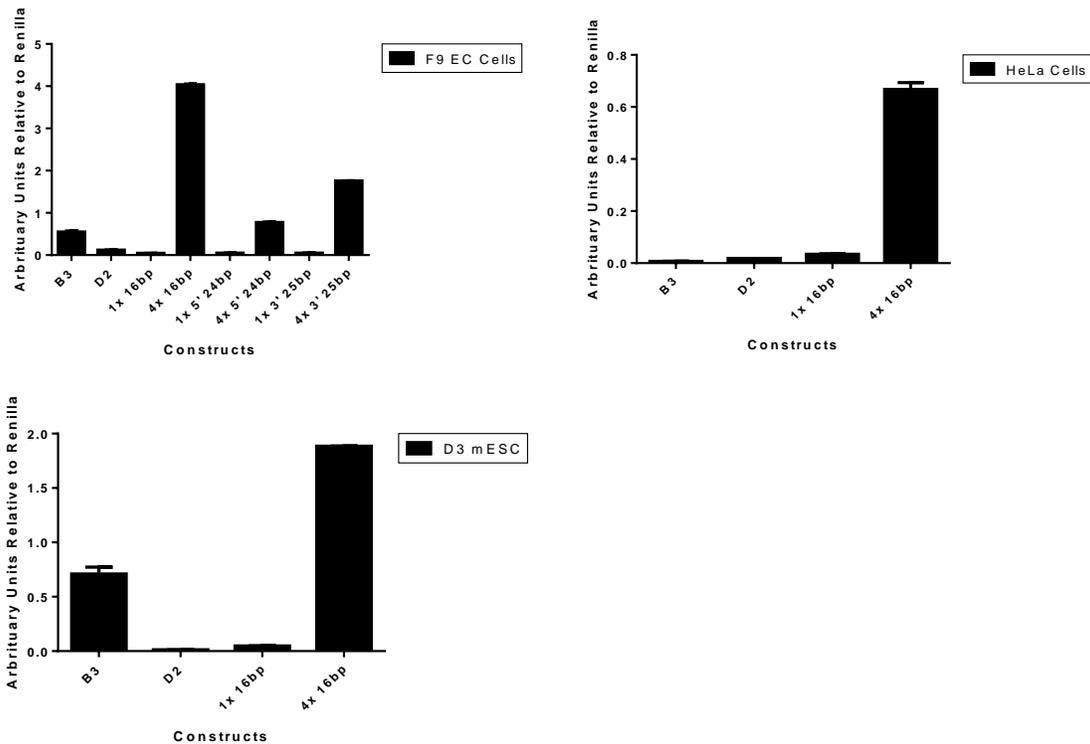
Figure 7. Unique binding site identified synergistically activates transcription when multimerized but activation is not cell type specific. (A-D) The 36 base pairs scanned by D2.10 to D2.14 constructs are split into the first 24 base pairs, the last 25 base pairs, and 16 base pairs in the middle encompassing D2.11 and D2.12. These sites are multimerized four times tandemly and placed in front of a luciferase construct with a minimal promoter. These constructs are transfected into P19, F9, HeLa, and D3 cells and lysate is assessed for luciferase activity compared to renilla control.

Materials and Methods

Cell Culture

Mouse embryonal carcinoma cell lines, F9 and P19, and mouse D3 ESC line was obtained from ATCC. F9, P19, and 293T cells were cultured in DMEM high glucose with GlutaMAX (Life Technologies) supplemented with 10% fetal bovine serum (HyClone). Mouse D3 ESCs were cultured in knockout DMEM (Life Technologies) supplemented with 15% fetal bovine serum (HyClone), 2mM GlutaMAX (Life Technologies), non-essential amino acids (Life Technologies), 0.1mM 2-mercaptoethanol (Sigma Aldrich), and 1000 units of LIF (Millipore) on 0.1% gelatin without feeder cells.

Western Analysis

Cells are lysed in 1x Laemmli sample buffer and ran on a 10% Bis-Tris SDS PAGE gel with 1x MOPS-Tris running buffer. The proteins are transferred onto nitrocellulose membrane (GE Healthcare), blocked with 10% milk in 1x TBS + 0.1% Tween-20 (TBST), and incubated with mouse anti-FLAG antibody (F3165, Sigma) overnight at 4C. The membrane is washed with 1x TBST, incubated with goat anti-mouse IgG coupled with HRP (PI31430, Fisher) for one hour at room temperature, and treated with the Western Lightning ECL + detection system (Perkin Elmer).

Lipofectamine transfection with luciferase assay

Enhancer and subsequence truncations are ligated into pGL3 and pGL4.23 luciferase constructs. Primer sequences are listed in the appendix. The plasmids are transfected into P19 and F9 cells using Lipofectamine 2000 (Thermo Fisher) with Renilla control plasmid. 48 hours later, cells are washed with 1x PBS and lysed in 1x Passive Lysis Buffer supplied by the Dual-Luciferase Reporter Assay System (Promega) and activity is measured according to manufacturer's protocol.

shRNA knockdown, RNA isolation, reverse transcription, and real time PCR analysis

shRNA constructs against NR5A1 is obtained from Sigma and transfected into 293T cells using Lipofectamine 2000. After 72 hours of transfection, total RNA was extracted and purified using TRIzol reagent (Life Technologies), according to manufacturers' protocol. cDNA synthesis was performed with 1 µg of total RNA using iScript cDNA Synthesis Kit (Bio-Rad) and diluted 10-fold. Real time PCR analysis was carried out with SYBR Select Master Mix for CFX (Life Technologies) using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Gene specific primer sequences are provided in the appendix.

Appendix B

Oct4 Enhancer Construct Primers.

Construct Name	Forward (5' → 3')	Reverse (5' → 3')
A	ACGTAAGTACTTCAGACAC CAGAAGAGG	CACACCCAGTTCCTCCCA
B	TGTGGGGAGGTTGTAGCC	GTCATGCTCACCTCCCAAT
C	ATTGGGAGGTGAGCATGAC	TGCAGAGAGCCTACCCTGAA
D	ACTCTAGGGAAGTTCAGGG T	TCTGACTTCAGGTTCAAAGGGG
E	CCCCTTTGAACCTGAAGTC AGA	AAAGCCTGTTGGCACTGCACC CTCTCGG
B 1	TGTGGGGAGGTTGTAGCC	CTGCCCAGAACTCTCA
B 2	TGAGAGTTCTGGGCAG	GCTAGGACGAGAGGGA
B 3	TGTGGGGAGGTTGTAGCC	GCTAGGACGAGAGGGA
B 4	TCCCTCTCGTCCTAGC	GTCATGCTCACCTCCCAAT
D 1	ACTCTAGGGAAGTTCAGGG T	TCACACAAGACTTCCCCAGC
D 2	GCTGGGGAAGTCTTGTGTG A	TCTGACTTCAGGTTCAAAGGGG
D 3	GAGCAGGAAGTTGTCC	GGGCAGGACAATGGCCTT
Single 16bp multimerization	GGGAGCAGGAAGTTGT	
Single 5' 24 bp multimerization	GGGGGTTGGGGAGCAGGA AGTTGT	
Single 3' 25 bp multimerization	AGCAGGAAGTTGTCCCCAG GGGAGC	

Oct4 Enhancer Mutagenesis Primers

Construct Name	Mutagenesis Primer (5' → 3')
D 2.1	CGCTGGGGAAGTCTTGTGTGATTTATGAAAACCCACGTTGGGGGTTGGGGAGCAG
D 2.2	TGTGAGGGGATTGGGGCTCAGGATTTATGAAAACCCACGTTAGTTGTCCCCAGGGGAGC
D 2.3	GGAGGGGGTTGGGGAGCAGGATTTATGAAAACCCACGTTCATCCTGGCCCATTCAAGG
D 2.4	AAGTTGTCCCCAGGGGAGCTTTATGAAAACCCACGTTGGTTGAGTACTTGTTTAGGGT
D 2.5	GCCATCCTGGCCCATTCAAGTTTATGAAAACCCACGTTGGTTAGAGCTGCCCCCTCTG
D 2.6	AGGGTTGAGTACTTGTTTAGTTTATGAAAACCCACGTTTGGGGACCAGGATTGTCCAGC
D 2.7	TAGGGTTAGAGCTGCCCCCTCTTTATGAAAACCCACGTTAGCCAAGGCCATTGTCCTGC
D 2.8	TCTGGGGACCAGGATTGTCCTTTATGAAAACCCACGTTGCCCCCTTCCCCCAGTCCCTC
D 2.9	CCAGCCAAGGCCATTGTCCTTTTATGAAAACCCACGTTCTCCCAGGCCCCTTTGAACC
D 2.10	ATTGGGGCTCAGGAGCCACGTTGGGAGCAGGAAGTTGTC
D 2.11	CTCAGGAGGGGGTTGCCACGTTGGAAGTTGTCCCCAGG
D 2.12	AGGGGGTTGGGGAGCACCACGTTGTCCCCAGGGGAGCCATC
D 2.13	TTGGGGAGCAGGAAGTTCCACGTTGGGGAGCCATCCTGGC
D 2.14	GCAGGAAGTTGTCCCCACCACGTTCATCCTGGCCCATTC

Dot Blot Probe Sequences

	Sequence (5' → 3')
Telomere Probe	TTAGGGTTAGGGTTAGGGTTAGGG
ALU SINE Probe	GTGATCCGCCCGCCTCGGCCTCCCAAAGTG

RT-qPCR Primer Sequences

	Forward (5' → 3')	Reverse (5' → 3')
mSF1	TCATCCTCTTCAGCCTCGAT	GCACAATAGCAACTGCTGGA
mACTB	AGAGGGAAATCGTGCGTGAC	CAATAGTGATGACCTGGCCGT
dmH1	TCCCCAGTGGCTGCCCCACC	GTGACGGCGTCGCAGAGGCT
dmH2A	AGCGTGTTGGTGCAGGCGCT	TGGCCAGTTGCAGATGACGCGG
dmH3	TGGTGGAAAGGCGCCACGCA	AGGCCACGGTTCCAGGGCGA
dmH4	CAGAGCGTACACAACATCCA	GTGTGAAGCGCATATCTGGA
dmTUB84b	CGTGAATGTATCTCTATCCATGT	TTGCCAGCTCCAGTCTCGCT
dmACTIN	GCGTCGGTCAATTCAATCTT	AAGCTGCAACCTCTTCGTCA
dmRPL32	CCAGTCGGATCGATATGCTAA	GTTCGATCCGTAACCGATGT

qPCR Primer Sequences

	Forward (5' → 3')	Reverse (5' → 3')
Telomere Target Plasmid Control	GGGAATTGTGAGCGGATAAC	AGCGAATTCACGTGATGATG
H2A Gene 3' End	AATTATTCCGCGTCATCTGC	GGCCTTCTTCTCGGTCTTCT
H2B Gene 5' End	GTGTCAGGATGGACCTGCTT	CTAGTGGAAAGGCAGCCAAG
H2A/ H2B Promoter	CCTTCACTTTGCCACCTTTT	GTCACCCACCCCTAACTGAA
H1 Promoter	CACTTCAAGCAAACTTCGACA	CTGCCTACCAACCTCCTTTG
H2A Gene 5' End	ACTACGCAGAGCGTGTTGGT	CTTGTTGTCACGAGCAGCAT

sgRNA DNA Targets

Target	Sequence (5' → 3')
Telomere	GTTAGGGTTAGGGTTAGGGTTA
Non-specific	ACATGTTGATTTCCTGAAA
dm sgRNA #1	AAACGTTCATCCCCTAGAA
dm sgRNA #2	TAGATGTTTTTTTATAAAT
dm sgRNA #3	GTCACTTTCCATTTTTAAA
dm sgRNA #4	TTCGGATCCTTTTCCTTCT
dm sgRNA #5	GTGACTGCAGCGAAGCCAA
dm sgRNA #6	GGAGAAAATTATTTAGAGC
dm sgRNA #7	AGAACCTTGTAAATGTAGA
dm sgRNA #8	TTCTTTTTCTTGTCGGTCT
dm sgRNA #9	GAAACTACGCAGAGCGTGT
dm sgRNA #10	TTACGGCAGCTAGGTAAAC
dm sgRNA #11	GAAGGCCTAAACGTTTCAA
dm sgRNA #12	CTGCCTTTCCACTAGTTTT
dm sgRNA #13	GGTGGCAAAGTGAAGGGAA
dm sgRNA #14	GCCGGTCTTCAATTCCCTG
dm sgRNA #15	CTGAGGTTCTCGAGTTGGC

Drosophila RNAi primers

Target	Forward (5' → 3')	Reverse (5' → 3')
BIN1	TTAATACGACTCACTATAGGGAGA ATGGCCAACGTGGAATCTAT	TTAATACGACTCACTATAGGGAGA GTACGGACGCTGGCGCC
BUBR1	TTAATACGACTCACTATAGGGAGA GGCCTGGAATAAGGCAAATG	TTAATACGACTCACTATAGGGAGA CCAACAACTCCTCTGGCT
CP190	TTAATACGACTCACTATAGGGAGA AAGCCTGCTATCGCC	TTAATACGACTCACTATAGGGAGA CGCCTTCTGTTGTGCT
CG3262	TTAATACGACTCACTATAGGGAGA GGAAAGGGTGGTGTTGGAA	TTAATACGACTCACTATAGGGAGA TTGCGCGGTAGAGAACTT
CG8771	TTAATACGACTCACTATAGGGAGA CAACCGGGTGTTACCGAG	TTAATACGACTCACTATAGGGAGA CACGAACTTTGGCTGTTTC
VIG2	TTAATACGACTCACTATAGGGAGA CAATCGTGACAACAGGGGA	TTAATACGACTCACTATAGGGAGA TCCGTCATTGCGGAAGCC
CG12608	TTAATACGACTCACTATAGGGAGA CGCGGAAACAGTCGCAG	TTAATACGACTCACTATAGGGAGA AGGCGATGGCCTTGAC
SYNCRIP ISOFORM C	TTAATACGACTCACTATAGGGAGA GGTCAGCGTAAATACGGC	TTAATACGACTCACTATAGGGAGA TTGTTTGCTCATCCGGCTC
DBP21E2	TTAATACGACTCACTATAGGGAGA GGTGAGGAACTCCAGCAGG	TTAATACGACTCACTATAGGGAGA CAGGATCATCTGGGTGCC
BAP60	TTAATACGACTCACTATAGGGAGA CATCGCTACTGCAGCGC	TTAATACGACTCACTATAGGGAGA GCTTGAACTGCAGCGGC
ROX8	TTAATACGACTCACTATAGGGAGA CCAGTCCCGGCAATCAG	TTAATACGACTCACTATAGGGAGA ACCTCGCTGTTGTGCG
CDC5	TTAATACGACTCACTATAGGGAGA CTCGCAAGTTGAAGCCCG	TTAATACGACTCACTATAGGGAGA GCTAGCAAAGCATCCGTCG
IRBP	TTAATACGACTCACTATAGGGAGA TGTCACGGACGTCAGGGA	TTAATACGACTCACTATAGGGAGA ACAAGCGCTTCGATCCG
NAP1	TTAATACGACTCACTATAGGGAGA AGGACGTCTACAAGCTGGA	TTAATACGACTCACTATAGGGAGA TCTGCTGGGAGTCGTCG
BRM	TTAATACGACTCACTATAGGGAGA GGGACAGCCATTGCCA	TTAATACGACTCACTATAGGGAGA TCGTTCAGCCTCTAGCTTC
CKIIalpha	TTAATACGACTCACTATAGGGAGA ACGACCACGGAAAAGTGC	TTAATACGACTCACTATAGGGAGA ATCGCTTTCGTGAGTGACG
ROW	TTAATACGACTCACTATAGGGAGA AACGGCGACTCCTTCG	TTAATACGACTCACTATAGGGAGA GGCCGCTTGTAGGTGG
VIG	TTAATACGACTCACTATAGGGAGA AGGAAGCGCGAGTTCG	TTAATACGACTCACTATAGGGAGA TGGGCCACGGTTTCCAC