THE USE OF REPORTERS TO IDENTIFY FACTORS INVOLVED IN 3' END PROCESSING OF NONPOLYADENYLATED RNA POLYMERASE II TRANSCRIPTS

Brandon Douglas Burch

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2010

Approved by:

Advisor: Dr. William F. Marzluff

Reader: Dr. Jeanette G. Cook

Reader: Dr. Zbigniew Dominski

Reader: Dr. Robert J. Duronio

Reader: Dr. Scott M. Hammond

Reader: Dr. Stephen L. Rogers

ABSTRACT

BRANDON DOUGLAS BURCH: The Use of Reporters to Identify Factors Involved in 3' End Processing of Nonpolyadenylated RNA Polymerase II Transcripts
(Under the direction of Dr. William F. Marzluff)

Metazoan replication-dependent histone mRNAs are unique in that, unlike other mRNAs in the cell, they are not polyadenylated, ending instead in a conserved 3' stem loop. Because histone pre-mRNAs do not contain introns, they require only a single endonucleolytic processing event to form the mature histone mRNA. The reaction that forms the 3' end of histone mRNAs requires an assembly of several factors, including the stem loop binding protein (SLBP) and the U7 snRNP, which contains U7 snRNA, two U7 snRNP-specific ring proteins, Lsm10 and Lsm11, and, in mammals, ZFP100. This complex on the pre-mRNA recruits a cleavage factor that contains CPSF73, CPSF100, Symplekin, and possibly other factors, which cleaves the histone pre-mRNA. However, the full composition of the processing complex remains incompletely defined. Here I describe the creation and use of reporter constructs to identify factors that are required for histone pre-mRNA 3' end formation in *Drosophila*, as well as other reporters to investigate histone transcription and the 3' end formation of another RNA polymerase II transcript, U7 snRNA. Further, I present the results of a reporter-based genome-wide RNAi screen to identify factors required for histone pre-mRNA processing. In this study, we identified 24 proteins that have some role in the production of mature histone mRNA, most of which had not been implicated before. Finally,

I describe the characterization of the *Drosophila* ortholog of FLASH, a newly discovered processing factor that binds to Lsm11. Through biochemical studies, I have defined the regions of Lsm11 required to bind FLASH and the regions of FLASH required to bind Lsm11. I have also determined that the binding interaction between dFLASH and dLsm11 is not required for recruitment of dFLASH to the histone locus body (HLB). I have used reporters to help characterize the function of dFLASH in *Drosophila* cultured cells. These studies have demonstrated that removing a portion of the N-terminus of FLASH results in a protein that acts as a dominant negative for histone pre-mRNA processing in vivo, presumably because it binds Lsm11 but cannot interact with other essential factors for processing.

ACKNOWLEDGEMENTS

This volume is the culmination of several years of work and would not have been possible without the help and support of many people along the way. It is not possible to list here everyone who has had an impact on my research and my life, but the significance of their contributions is not diminished. These include the family and friends that have given me love and encouragement over the years, as well as the many colleagues that have helped shape me as a scientist and a person. To these, I offer my heartfelt thanks.

To my wife, Erin Burch, who has given me so much during our years together and has supported me throughout my graduate career, I cannot thank you enough. You have endured so much and have given me the strength to press on even when times were tough. Your love and dedication is truly appreciated and I can never express how much our relationship means to me.

To my parents, Doug and Bonnie Burch, who have always been a beacon of support and love throughout my life, I offer my sincere thanks. Your encouragement has been instrumental to my success and I appreciate all you have taught me through the years. Though graduate school has kept me from many family gatherings, you have always been supportive of my work and understanding of my reasons for not being there. Thank you for all the love you have shown me.

To my "Gaddy parents," Gary and Jill Gaddy, I thank you for your love and encouraging words and for welcoming me into your family. I am blessed to have two such wonderful people in my life.

To my advisor, Bill Marzluff, thank you for being such a great mentor and advocate for me during my time in the lab. You have been a tremendous influence on my development as a scientist, not only for your keen insight and knowledge, but also for your ability to view your researchers as people first.

To the director of the Curriculum in Genetics and Molecular Biology, Bob Duronio, thank you for your dedication to the program, for serving on my committee, and for all your help with my research. The collaboration between our two labs has been wonderful.

To Zbig Dominski, thank you for all your help and kind support, especially during the last two years. I thank you for serving on my committee and for making me a better scientist. It has been a great experience to share the FLASH project with you.

To my other committee members, Jean Cook, Scott Hammond, and Steve Rogers, thank you for all of your help and input over the years, for challenging me to become a better scientist, and for your encouragement.

To the members of the Marzluff Lab, past and present, thank you for your support and your friendship and for helping me through my time in the lab. I have greatly enjoyed knowing all of you and sharing the past several years with you.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

Act_{proc}	Actin promoter-driven histone pre-mRNA processing reporter
Act_{trans}	Actin promoter-driven transcriptional reporter
Cdk	Cyclin dependent kinase
CPSF	Cleavage and polyadenylation specificity factor
FLASH	FLICE-associated huge protein
HDE	Histone downstream element
His_{proc}	Histone promoter-driven histone pre-mRNA processing reporter
His_{trans}	Histone promoter-driven transcriptional reporter
HLB	Histone locus body
Lsm	Sm-like protein
MPM-2	Mitotic protein monoclonal #2
Mxc	Multiple sex combs
NPAT	Nuclear protein, ataxia-telangiectasia locus
SLBP	Stem loop binding protein
snRNA	Small nuclear ribonucleic acid
snRNP	Small nuclear ribonucleoprotein
V5	Simian virus 5 sequence used to tag recombinant proteins

CHAPTER I

INTRODUCTION

Among the most important events occurring within a mitotic cell is duplication of its genome in anticipation of cellular division. This process is strictly governed by a host of regulatory mechanisms that ensure not only the fidelity of the DNA sequence but further, that the genome is replicated once and only once per cell cycle. These controls include factors and pathways that monitor the timing and level of replication, as well as any introduction of damage to the DNA. If perturbations are detected, these factors allow the cell to mount a proper response to repair the damage. Using these control mechanisms, a normal cell is able to proceed efficiently through an entire cell cycle, the culmination of which is its division into two identical daughter cells.

Closely linked to DNA replication is the production of replication-dependent histone proteins during S phase of the cell cycle (Robbins and Borun, 1967; Spalding *et al.*, 1966). These proteins are required for packaging the newly synthesized DNA into a highly condensed form within the nucleus and they play a crucial role in gene expression through covalent modifications that culminate in the histone code (Strahl and Allis, 2000). There are five classes of replication-dependent histone proteins, including core histones H2A, H2B, H3, and H4 that comprise the histone octamer and histone H1, which acts as a linker protein outside the core octamer. The histone octamer and 146 nt of the newly replicated DNA make up the basic unit of chromatin, known as the nucleosome. Nucleosomes are linked together

by another histone protein, histone H1. At this level of compaction, the chromatin fiber, often referred to as "beads on a string," is approximately 10 nm in diameter. These fibers further condense into higher order chromatin structures, including 30nm fibers and 300nm fibers, finally culminating in the most compacted form, the chromosome.

THE ROLE OF HISTONE PROTEINS IN GENE REGULATION

Protruding from the octamer are the N-terminal tails of the histone proteins, which participate not only in facilitating higher order chromatin structures (Peterson and Laniel, 2004), but also play a key role in gene regulation. These tails are approximately 20-35 amino acids in length, contain a large number of basic amino acids, and are the target of various posttranslational modifications that have been collectively dubbed the histone code (Lopez-Rodas *et al.*, 1993; Tordera *et al.*, 1993; Turner, 1993; Loidl, 1994; Mizzen *et al.*, 1998; Strahl and Allis, 2000). These modifications are crucial for defining regions of active chromatin and gene transcription and, conversely, for demarcating heterochromatic regions of the genome (Peters *et al.*, 2001; Shiio and Eisenman, 2003). This complex system of gene expression allows for a greater diversity of outcomes and control than would be possible from the DNA alone.

There are several types of posttranslational modifications that have been described for histone tails, including acetylation (Allfrey *et al.*, 1964; Grunstein, 1997; Kuo and Allis, 1998), methylation (Jenuwein and Allis, 2001; Peters *et al.*, 2001), phosphorylation (Roth and Allis, 1992; Dou and Gorovsky, 2000; Ren and Gorovsky, 2001; Peterson and Laniel, 2004), ubiquitination (Robzyk *et al.*, 2000; Henry *et al.*, 2003; Hwang *et al.*, 2003; Wood *et al.*, 2003), sumolyation (Shiio and Eisenman, 2003), biotinylation (Stanley *et al.*, 2001;

Camporeale *et al.*, 2004), and ADP-ribosylation (Ueda *et al.*, 1975; Burzio *et al.*, 1979; Riquelme *et al.*, 1979). These modifications are not stand-alone signaling marks, however, as the same modification can lead to different gene expression outcomes, indicating that this code is context-dependent and likely multi-layered. Indeed, it appears that not only a particular modification, but other surrounding modifications on the same molecule and within the same nucleosome and the topology of the chromatin in the region are crucial for achieving a given readout. As a result, there has been a suggestion to change the term "histone code" into the more inclusive "chromatin code" (Benecke, 2006).

In addition to posttranslational modification of canonical histone proteins, incorporation of variant histone proteins into the chromatin adds yet another level of complexity to chromatin structure and gene regulation. Unlike replication-dependent histone proteins, expression of the variant histones is not linked to DNA replication. Indeed, these proteins are constitutively expressed throughout the cell cycle (Wu *et al.*, 1981). Further, the genes that encode the variant histones, unlike replication-dependent histone genes, give rise to mRNAs that are polyadenylated, many of which also contain introns (Engel *et al.*, 1982; Woudt *et al.*, 1983; Brush *et al.*, 1985). Because of this, expression of replication-independent histone genes is not regulated like the replication-dependent histone genes (Sittman *et al.*, 1983). These histone variants participate in a wide array of important regulatory mechanisms in the cell, including activation or repression of gene expression, response to DNA damage, and formation of chromosomal structures like centromeres.

HISTONE EXPRESSION IS LINKED TO THE CELL DIVISION CYCLE AND DNA REPLICATION

The eukaryotic cell division cycle is a highly regulated temporal program that allows a cell to replicate its genome and other cellular components and separate these into two new daughter cells. The canonical cell cycle consists of two phases, interphase and mitosis, each divided into smaller subphases. Directly after the cell divides, each new daughter cell enters interphase. This phase consists three subphases, the first of which is a gap phase, termed G1, in which the cell assembles all the factors it will require to initiate and sustain genome replication. Next, the cell enters a synthesis (S) phase, which is a time of genome duplication and cell growth. Interphase then concludes with one more gap phase (G2), in which the cell prepares to enter mitosis. Mitosis is itself comprised of several subphases, each corresponding to a particular state of genomic separation. These phases are known as prophase (chromosome condensation and nuclear envelope breakdown), metaphase (alignment of chromosome along the metaphase plate by spindle fibers radiating from duplicated centrosomes, one in each pole of the cell), anaphase (separation of sister chromatids and migration toward the spindle poles), and telophase (reformation of the nuclear envelope around the segregated chromosomes). In normally cycling cells, these steps are directly followed by cytokinesis, in which the cell divides into two daughter cells, and the cell cycle resets to G1.

Among the most critical and highly regulated steps within the cell replication cycle is duplication of the genome during S phase. It is imperative that the cell accurately and efficiently replicate its DNA before entering mitosis; otherwise, catastrophic consequences may ensue, including DNA damage, genomic instability, and cell death. However, accurate

replication of the cell's DNA is not enough; indeed, the cell must replicate its entire chromatin, which requires proper coordination of DNA synthesis levels with production of canonical histones. For this reason, there are a number of cellular controls that monitor not only proper initiation, fidelity, and termination of DNA replication, but also the state of the chromatin environment within the cell.

The cell prepares for DNA replication during G1 phase with the identification of replication origins. In budding yeast, these origins consist of special sequences, known as autonomously replicating sequences (ARS) (Brewer and Fangman, 1987; Huberman *et al.*, 1988; Marahrens and Stillman, 1992), that are around 100-200 basepairs long, contain well conserved sequence elements, and are spaced approximately every 30 kilobases throughout the genome (Bell, 1995). In higher eukaryotes, the replication origins are poorly defined, and the location and identity of the origins appears to be more stochastic, possibly relying more on the local chromatin environment than a particular conserved sequence (DePamphilis, 1993; Bogan *et al.*, 2000; Bell, 2002; Antequera, 2004; Schwob, 2004; Cvetic and Walter, 2005; MacAlpine and Bell, 2005; Marheineke *et al.*, 2005). In all eukaryotes, the replication origins are bound by the multisubunit origin recognition complex (ORC), which acts as a marker of replication origins (Bell and Stillman, 1992; Diffley and Cocker, 1992; Romanowski *et al.*, 1996; Rowles *et al.*, 1996; Carpenter *et al.*, 1996; Dutta and Bell, 1997; Lee and Bell, 1997). Once the ORC has identified the origin, Cdc6 (Coleman *et al.*, 1996) and Cdt1 (Hofmann and Beach, 1994; Nishitani *et al.*, 2000) are recruited, which in turn are responsible for recruiting the Mcm complex, possibly through ATPase-induced conformational changes in ORC (Chong *et al.*, 1995; Kubota *et al.*, 1997; Nishitani *et al.*, 2000). The MCM complex contains the helicase activity that will unwind the DNA for

replication (Koonin, 1993) and its recruitment to the replication origin completes the formation of the pre-replicative complex (pre-RC) (Blow and Dutta, 2005).

Formation of the pre-RC is regulated by several proteins that ensure the proper timing of DNA replication during the cell cycle. One of these, geminin, binds Cdt1 and inhibits its ability to associate with ORC (McGarry and Kirschner, 1998; Wohlschlegel *et al.*, 2000). During G1, geminin levels are reduced through ubiquitination by the anaphase promoting complex (APC) (McGarry and Kirschner, 1998), which targets geminin for degradation (Li and Blow, 2004), allowing Cdt1 to bind to ORC. In addition, other proteins, including cyclin dependent kinases, are involved in regulating pre-RC formation and reformation, likely through phosphorylation of Cdt1, which targets this protein for ubiquitiation and destruction. This function is especially important during S phase to prevent rereplication of DNA sequences.

Once the pre-RC has been formed and the cell has prepared for entry into S phase, an activated cyclinE/Cdk2 complex (Jackson *et al.*, 1995) results in activation of the S phase cyclinA/Cdk2 complex, which is partially responsible for activating replication origins (Girard *et al.*, 1991). In addition to cyclinA/Cdk2, activation of another protein kinase, Cdc7, through its interaction with Dbf4, is required (Bousset and Diffley, 1998; Donaldson *et al.*, 1998). In this complex, Dbf4 not only stimulates Cdc7 kinase activity, but may also target Cdc7 to replication origins, allowing for phosphorylation of Mcm components. Importantly, at about the same time as the replication origins are being activated, a critical component for histone expression, NPAT, is also activated by cyclin E/Cdk2 phosphorylation (Ma *et al.*, 2000), linking the processes of DNA replication and histone synthesis.

As S phase begins, other proteins localize to the pre-RC, including Mcm10 (Merchant *et al.*, 1997; Aves *et al.* ,1998; Homesley *et al.*, 2000), the Cdc45 complex (Aparicio *et al.*, 1999), Dbp11/TopBP1 (Kamimura *et al.*, 1998; Kumagai *et al.*, 2006; Pospiech *et al.*, 2010), and the GINS complex (Takayama *et al.*, 2003), thus forming the preinitiation complex. The preinitiation complex recruits DNA polymerases, including the primase (DNA pol α), which interacts with Cdc45 and Mcm10, pol ε, which is bound by Dpb11, and pol δ (Aparicio *et al.*, 1999; Walter and Newport, 2000; Ricke and Bielinsky, 2004). DNA pol α is responsible for creating replication primers at the start of replication (Spadari and Weissbach, 1975; Conaway and Lehman, 1982), while pol δ (Byrnes *et al.*, 1976) and pol ε (Syväoja *et al.*, 1990) extend these primers and continue replicating the leading and lagging strands at the replication fork.

Organization of the Histone Genes in Mammals and Flies

Along with regulation of DNA synthesis, histone expression is also regulated, with much of this regulation taking place at the site of the histone genes. The general organization of histone genes has been conserved in metazoans, with these genes existing in multi-gene clusters that make up a histone locus. Though the clustered nature of the histone genes is well conserved, the precise location of these genes and the number of histone genes encoded in the genome is not. In mammals, there are approximately 75 replication-dependent histone genes encoded in the genome, existing primarily in two major clusters covering over two megabases of DNA (Green *et al.*, 1984; Tripputi *et al.*, 1986). In contrast, the *Drosophila* genome contains only five replication-dependent histone genes. These genes are located together on chromosome 2 and are arranged in an approximately 5 kb cluster (Fig. 1) that is tandemly repeated about100 times (Lifton *et al.*, 1978).

The organization of the histone genes into a histone gene cluster likely offers the cell the ability to rapidly modulate histone expression levels by targeting a single location in the genome, rather than several. This may be especially important during periods of replication stress or DNA damage, when histone levels must be reduced quickly. One important feature of the *Drosophila* histone gene cluster are cryptic polyadenylation signals within the intergenic regions downstream of the normal cleavage site (Fig. 1). These signals are used if normal histone pre-mRNA processing becomes disrupted and probably evolved in response to the close proximity of the histone genes to each other as a way to insulate them from transcriptional readthrough and prevent the creation of chimeric histone proteins.

The Histone Locus Body

During G1 phase, assembly of a subnuclear structure, known as the histone locus body (HLB), occurs at the site of the histone genes. This body has been best described in *Drosophila* (Liu *et al.*, 2006). There is clearly also an HLB in mammalian cells, although this was first described as a Cajal body localized near histone genes (Frey and Matera, 1995). A primary component of this structure is the $p220^{NPAT}$ protein, found only at histone genes in mammals, which forms discrete foci at the histone locus during mid-G1 phase and around 6 hours before the onset of S phase in normally cycling mammalian cells (Zhao *et al.*, 2000; Ma *et al.*; 2000; Zheng *et al.*, 2003; Ghule *et al.*, 2008). Just before the G1/S transition, NPAT becomes phosphorylated by cyclin E/Cdk2 and remains phosphorylated into S phase (Ma *et al.*, 2000). In this form, NPAT promotes transcription of histone genes during S phase (Ma *et al.*, 2000; Wei and Harper, 2003). Since NPAT is phosphorylated by the same cyclin/Cdk that promotes DNA replication uncovers an important link between histone expression and DNA synthesis. In addition, it provides a key regulatory mechanism within S

phase, as inhibition of DNA synthesis results in dephosphorylation of NPAT and its subsequent delocalization from the HLB (Su *et al.*, 2004). As a result, histone expression decreases, preventing the accumulation of excess histone proteins under these conditions. Similarly, inhibiting histone expression causes a concomitant decrease in DNA replication, further suggesting a tight link between these two processes.

In addition to NPAT, the HLB contains many of the factors that are required for transcribing the histone genes and processing the 3' end of the pre-mRNAs to form mature histone mRNAs. Each of these factors, then, represents a potential target for regulation. These include the U7 snRNP and the protein FLASH, identified as a putative transcription factor that colocalizes with NPAT in mammalian cells (Barcaroli *et al.*, 2006). In contrast, SLBP, which is required for the cleavage reaction that leads to the formation of mature histone mRNA, is not concentrated in the HLB. In mammals, SLBP protein is cell cycle regulated, with its highest levels of expression occurring from the G1/S phase transition through S phase (Whitfield *et al.*, 2000). At the end of S phase, this protein becomes degraded, effectively downregulating histone expression. In this way, SLBP provides another important link between histone expression and the cell cycle.

REGULATION OF REPLICATION-DEPENDENT HISTONE EXPRESSION

Given the importance of the canonical replication-dependent histone proteins for DNA compaction and gene expression, expression of the histone genes is highly regulated (Fig. 2). This regulation ensures proper levels of histone proteins during S phase and prevents expression of these proteins outside of S phase. Aberrant expression of histone proteins results in catastrophic consequences for the cell, including chromosome loss in yeast

(Meeks-Wagner and Hartwell, 1986), DNA damage in mammals (Ye *et al.*, 2003), and developmental arrest in *Drosophila* (Sullivan *et al.*, 2001). In support of a tight linkage between histone expression and DNA replication, it has been observed that blocking histone synthesis inhibits DNA replication (Nelson *et al.*, 2002) and further, that blocking DNA replication correspondingly inhibits histone expression (Sittman *et al.*, 1983; Heintz *et al.*, 1983; Baumbach *et al.*, 1987). Because of the tight association, several targets for histone expression regulation have evolved that help ensure proper coordination of histone expression with DNA replication and the cell cycle.

Histone Gene Transcription

Regulation of histone expression occurs largely at the mRNA level, being both transcriptional and post-transcriptional. Overall, an increase in the transcription of histone genes accounts for 3.5-fold of a total 35-fold increase in total histone message upon entry into S phase in mammals (Harris *et al.*, 1991), making it a highly significant control target for histone expression. Such control likely stems from cell cycle regulation of HLB formation and the existence of of histone-specific transcription factors. Despite these unique regulatory properties, transcription of histone genes also shares common factors with transcription of non-histone mRNA. Histone mRNAs are, like other mRNAs in the cell, transcribed by DNA-dependent RNA polymerase II. Further, these messages become modified at their 5' ends by a 7-methyl-guanisine cap that is added cotranscriptionally and is required for translation of the message in the cytoplasm.

As the first step in histone expression (Fig. 3), histone gene transcription presents a key target for regulation. Regulation of histone transcription occurs in a cell cycle-dependent manner. At the G1/S phase transition, NPAT, which localizes to the site of the histone genes,

becomes phosphorylated. As S phase begins, histone mRNA transcription by DNA-dependent RNA polymerase II increases (DeLisle *et al.*, 1983; Heintz *et al.*, 1983), likely due to the interaction of transcription factors with specific sequence elements within histone gene promoters (LaBella *et al.* 1988; Heintz, 1991) or within the coding region itself (Hurt *et al.*, 1991; Bowman and Hurt 1995; Eliassen *et al.*, 1998). A few transcriptional regulators have been identified, including factors such as YY1 (Palko *et al.*, 2004) and NPAT (Zhao *et al.*, 2000), as well as some subtype-specific factors, including the NPAT-interacting proteins HiNF-P (Mitra *et al.*, 2003) for histone H4 and OCA-S for histone H2B (Zheng *et al.*, 2003). The most recent addition to our understanding of histone transcription is from a study in *Drosophila* revealing the recruitment of POU-domain protein in D. melanogaster 1 (Pdm-1) to elements within core histone promoters (Lee *et al.*, 2010). Together with dmOct-1 coactivator in S-phase (dmOCA-S), this protein localizes to the HLB only during S phase, permitting transcription from the core histone promoters and helping to regulate their S phase-specific expression (Lee *et al.*, 2010).

Central to the regulation of histone expression is the coordination of this process with rates of DNA synthesis during S phase. Indeed, when cultured cells are treated with the DNA replication inhibitors hydroxyurea or cytosine arabinonucleoside, histone gene transcription is reduced by 80 % (Sittman *et al.*, 1983). As mentioned above, one of the most solid links between histone expression and DNA replication occurs at the level of transcription and involves NPAT, a transcriptional regulator of histone transcription in mammalian cells that localizes to the HLB from mid-G1 through S phase (Zhao *et al.*, 2000; Ma *et al.*, 2000; Zheng *et al.*, 2003, Ghule *et al.*, 2008). NPAT expression is regulated by the E2F transcription factor (Gao *et al.*, 2003), with the highest levels of NPAT mRNA achieved

at the G1/S boundary (Gao *et al.*, 2003; Zhao *et al.*, 1998). NPAT is activated at the end of G1 through phosphorylation by the cyclin E-Cdk2 complex (Zhao *et al.*, 1998; Zhao *et al.*, 2000; Ma *et al.*, 2000), which is itself activated in late G1 and is crucial for S-phase entry (Morgan, 1997; Ekholm and Reed, 2000; Roberts and Sherr, 2003). In addition to linking histone transcription to DNA replication under normal conditions, NPAT also does so after DNA damage (Su *et al.*, 2004). Upon exposure of cells to ionizing radiation, DNA replication is inhibited and histone mRNA levels are significantly decreased (Su *et al.*, 2004). These responses coincide with dephosphorylation of NPAT and its dissociation from the histone locus (Su *et al.*, 2004), suggesting coupling through NPAT. The dephosphorylation of NPAT occurs in a p53 and $p21^{CIP1/WAF1}$-dependent manner and likely results from inhibition of cyclin E/Cdk2 upon DNA damage (Su *et al.*, 2004). Indeed, using chemical Cdk2 inhibitors without the presence of DNA damage results in the same loss of NPAT from the HLB and downregulation of histone expression (Su *et al.*, 2004).

Histone pre-mRNA Processing

Upon entry into S phase in mammals, there is a net 35-fold increase in histone mRNA expression, of which, only 3.5-fold is related to transcription, while the majority, around 10-fold, occurs at the level of histone pre-mRNA processing (Harris *et al.*, 1991). There is a single processing reaction that involves cleavage of a longer histone transcript, the histone pre-mRNA, to form the mature histone mRNA (Birchmeier *et al.*, 1984; Krieg and Melton, 1984; Fig. 3). Histone pre-mRNAs do not contain any introns and as such, this single endonucleolytic event is all that is required to form the mature histone mRNA (Fig. 4). The cleavage reaction occurs cotranscriptionally (Adamson and Price, 2003) and, while reminiscent of the reaction required for processing polyadenylated mRNAs, requires a unique

complement of cis sequences and trans factors. The result is a histone mRNA with a 4-5 nt ssRNA overhang that remains nonpolyadenylated. Following the cleavage reaction, the 3' end of the histone mRNA is further trimmed to a 2-3 nt overhang, most likely by the 3'hExo (Dominski *et al.*, 2003a; Yang *et al.*, 2006). The mature 3' end of the histone message formed by the cleavage reaction is important for downstream events in histone mRNA metabolism, including export of the histone mRNA from the nucleus (Eckner *et al.*, 1991; Sullivan *et al.*, 2009a) and its subsequent translation (Sun *et al.*, 1992) and degradation (Pandey and Marzluff, 1987) in the cytoplasm, as well as coordination of histone mRNA levels with DNA synthesis (Levine *et al.*, 1987).

Histone mRNA Export

Once the histone pre-mRNA has been processed, forming the mature histone mRNA, it must be exported from the nucleus into the cytoplasm, where it will be translated (Fig. 3). Export of histone mRNAs is not well characterized, but certain details are known. Export is a very rapid process (Schochetman and Perry, 1972) that requires the presence of the mature histone mRNA 3' end (Eckner *et al.*, 1991; Sun *et al.*, 1992; Williams *et al.*, 1994) and SLBP (Sullivan *et al.*, 2009a). In addition, other factors, such as the SR proteins 9G8 and SRp20 may have a role in histone mRNA export, as binding sequences for these proteins have been found within the histone ORF and these factors are known to have a role in exporting intronless RNAs in both mammalian cells and Xenopus oocytes (Huang and Steitz, 2001). These proteins each contain an RNA recognition motif that is adjacent to a positively-charged, arginine-rich region that facilitates interaction with the export factor Tip-associated protein (TAP) (Hargous *et al.*, 2006). In agreement with this, export of histone mRNA has been shown to require the activity of TAP (Erkmann *et al.*, 2005), which functions in export

by traversing the nuclear pore complex (NPC) through interactions with NPC proteins (Katahira *et al.*, 1999; Bachi *et al.*, 2000).

Histone mRNA Translation

Once the mature histone mRNA has been successfully exported to the cytoplasm, it is translated (Fig. 3). For polyadenylated messages, this process requires the presence of both the 5' cap and the poly(A) tail (Gallie, 1991), which is bound by poly(A) binding protein (PABP). These components participate in the circularization of the mRNA (Wells *et al.*, 1998), which stabilizes the message and facilitates efficient turnover of terminating ribosomes. Circularization is made possible by interactions between the cap structure and eIF4E, which interacts with eIF4G, which itself interacts with the PABP bound to the poly(A) tail (Imataka *et al.*, 1998). Though histone mRNAs are translated via a similar mechanism (Cakmakci *et al.*, 2008) and do contain the 5' cap, they do not have a poly(A) tail, ending instead in a stem loop structure that is bound by SLBP. This means the histone 3' end/SLBP complex must interact with some other factor, other than PABP, that allows interaction with eIF4G and circularization of the histone mRNA.

The mature histone mRNA 3' end is required for localization of histone mRNA to polyribosomes (Sun *et al.*, 1992), where it is associated with SLBP (Whitfield *et al.*, 2004). SLBP stimulates translation of histone message both *in vitro* and *in vivo* (Sanchez and Marzluff, 2002) and is required for binding another translation factor, SLBP-interacting protein 1 (SLIP1), which is itself required for efficient translation of histone mRNA (Cakmakci *et al.*, 2008). It is likely that this factor performs the function for histone mRNA translation that PABP performs for polyadenylated mRNAs, to bridge the 5' and 3' ends of the mRNA to facilitate circularization and efficient translation of the message.

Histone mRNA Degradation/Stability

At the end of S phase or in response to inhibiting DNA replication or the introduction of DNA damage, histone mRNA half-life is reduced from 45-60 minutes to around 10 minutes, and the mRNA is rapidly degraded (Sittman *et al.*, 1983; Fig. 2; Fig. 3). In a normal cell that has reached the end of S phase, this mechanism of degradation prevents aberrant expression of histone proteins as the cell enters G2 of the cell cycle. In a damaged or arrested cell, degradation of histone message prevents the toxic accumulation of excess histone proteins. This gives the cell the chance to mount a proper response to the damage before allowing DNA replication and histone synthesis to proceed. A common way to induce histone mRNA degradation is to treat cells with a DNA synthesis inhibitor such as hydroxyurea or aphidicolin.

Histone mRNA is degraded by a mechanism that requires active translation (Graves *et al.*, 1987), a correctly spaced stem loop close to the termination codon and the presence of SLBP (Kaygun and Marzluff, 2005a). The nonsense mediated decay (NMD) factor Upf1 is also required for histone mRNA degradation and is recruited to the histone mRNA 3' end by SLBP at the end of S phase or when DNA replication has been inhibited (Kaygun and Marzluff, 2005b). Further, both ataxia telangiectasia and Rad3 related (ATR) (Kaygun and Marzluff, 2005b) and DNA-activated protein kinase (DNA-PK) (Müller *et al.*, 2007) are involved in controlling histone mRNA degradation in response to replication stresss. It is likely that these proteins are required for degradation of histone mRNA under stress conditions through a mechanism that involves phosphorylation of Upf1, but not necessarily SLBP (Kaygun and Marzluff, 2005b; Müller *et al.*, 2007). These replication stress response

proteins may represent two pathway of histone mRNA degradation control are not completely redundant (Müller *et al.*, 2007).

Once hyperphosphorylated Upf1 is recruited to the histone mRNA via its interaction with SLBP, Upf1 likely mediates recruitment of a terminal uridyl transferase (TUTase) that catalyzes the addition of a short poly(U) tail onto the 3' end of the histone message (Mullen and Marzluff, 2008). Following this, the oligo(U) tail becomes bound by the Lsm1-7 complex, which recruits the decapping complex, leading to 5'-3' degradation of the histone mRNA. In addition, it is probable that the exosome is recruited to the degradation complex, facilitating 3'-5' degradation of the histone message. Once the Lsm1-7 complex has been recruited, the histone mRNA is likely degraded by the same mechanism as other mRNAs. Indeed, evidence for both mechanisms of decay has been attained for histone mRNAs (Mullen and Marzluff, 2008).

FACTORS REQUIRED FOR HISTONE PRE-MRNA PROCESSING

The focus of this dissertation primarily lies on the histone pre-mRNA processing reaction and identification of the many factors that are required to facilitate cleavage. When I started this project, the two cis elements, the stem loop and histone downstream element (HDE) were known and it was known that they bound two factors, SLBP and the U7 snRNP, respectively, that are essential for processing (Fig. 5A). Further, an additional factor, ZPF100, which in mammals binds to both SLBP and the U7 snRNP protein, Lsm11, had been described. However, the nuclease was not known at the time. Since then, our understanding of histone pre-mRNA processing and the factors that participate in it has grown (Fig. 5B). Now we know that a core cleavage factor, comprised of CPSF73,

CPSF100, Symplekin, and potentially other factors, is recruited to the site of cleavage, though the mechanism by which this recruitment occurs is currently unknown. Another protein, FLASH, which makes contacts with the long N-terminus of Lsm11, has also been shown to be essential for the cleavage reaction. In addition, many more factors have been implicated for a role in the cleavage reaction and await characterization.

The Stem Loop and Histone Downstream Element

There are two sequence elements within the histone pre-mRNA that are required for the processing reaction to occur and which have been conserved in metazoans (Marzluff, 1992) from sea urchin (Birchmeier *et al.*, 1982; Birchmeier *et al.*, 1983; Birchmeier *et al.*, 1984) to mammals (Birchmeier *et al.*, 1984; Mowry and Steitz, 1987; Cotten *et al.*, 1988; Mowry *et al.*, 1989; Vasserot *et al.*, 1989). These sequences lie less than 100 nucleotides downstream of the stop codon and include a stem loop structure 4-5 nt upstream of the cleavage site and a U-rich element 15-20 nt downstream of the stem loop known as the histone downstream element (HDE; Fig. 4).

The stem loop structure is highly conserved both among the various histone mRNAs within a single organism and also between metazoans. This structure consists of a stem that is 6 basepairs in length and a 4-nt loop. The stem loop is cotranscriptionally bound by stem loop binding protein (SLBP), also known as hairpin binding protein (HBP). Together, the stem loop and SLBP are required for processing histone pre-mRNA (Dominski *et al.*, 1995; Hanson *et al.*, 1996; Wang *et al.*, 1996; Wang *et al.*, 1999; Zanier *et al.*, 2002) and other aspects of histone mRNA metabolism including nuclear export (Eckner *et al.*, 1991; Sun *et al.*, 1992; Williams *et al.*, 1994; Sullivan *et al.*, 2009a), translation (Sun *et al.*, 1992; Sanchez

and Marzluff, 2002, Whitfield *et al.*, 2004,Cakmakci *et al.*, 2008) and degradation (Kaygun and Marzluff, 2005a).

The second critical cis element, the HDE, resides approximately 15-20 nucleotides downstream of the stem loop. The sequence of the HDE is somewhat variable between metazoans, and between individual histone genes in the same species, although the core of the sequence, containing a region rich in purine residues, is much better conserved (Birchmeier *et al.*, 1983; Birnstiel *et al.*, 1985; Mowry and Steitz, 1987; Cotten *et al.*, 1988; Mowry *et al.*, 1989). The role of the HDE in the histone pre-mRNA processing reaction is to bind the U7 snRNP particle, which is essential for the cleavage reaction to occur. This interaction occurs through basepairing interactions between the HDE and the 5' end of the U7 snRNA component of the U7 snRNP. Because of this important interaction, the identity of the bases within this HDE, as well as the location of the HDE within the histone pre-mRNA, are crucial for correct processing to occur. Indeed, mutation of the sequence (Mowry *et al.*, 1989; Vasserot *et al.*, 1989) or moving the sequence downstream of its normal location (Georgiev and Birnstiel, 1985; Scharl and Steitz, 1994; Scharl and Steitz, 1996) leads to inefficient or abolished cleavage.

SLBP

The histone stem loop structure is bound by the stem loop binding protein (SLBP; Fig. 5), which participates in multiple aspects of histone mRNA metabolism, including histone pre-mRNA processing, export, translation, and degradation (Fig. 3). Its role in processing is essential *in vivo* (Pandey *et al.*, 1994; Lanzotti *et al.*, 2002) and is mediated in part by its ability to bind the stem loop structure near the 3' end of the histone pre-mRNA. This interaction requires a unique RNA binding domain (RBD) that is approximately 70

amino acids in length and is located in the central portion of the protein. Once bound to the stem loop, SLBP interacts indirectly with the U7 snRNP and stabilizes its interaction with the histone pre-mRNA substrate, thus facilitating the cleavage reaction (Melin *et al.*, 1992; Streit *et al.*, 1993; Spycher *et al.*, 1994; Dominski *et al.*, 1999).

Expression of SLBP is tightly linked to the cell cycle in mammals, with levels highest from the end of G1 through S phase, then rapidly declining at the end of S phase (Whitfield *et al.*, 2000; Fig. 2). Reduction of SLBP levels at the end of S phase is a result of phosphorylation of two threonine residues located at amino acids 60 and 61 (Zheng *et al.*, 2003; Koseoglu *et al.*, 2008), with phosphorylation of T60 by casein kinase 2 (CK2) and T61 by cyclinA/Cdk1 (Koseoglu *et al.*, 2008). Destruction of SLBP at the end of S phase is coincident with histone mRNA degradation during this time. In this way, mammalian SLBP represents a key regulatory target that helps coordinate the cell cycle regulation of histone expression. In contrast to the mammalian protein, however, *Drosophila* SLBP exists throughout the cell cycle, at least during development (Lanzotti *et al.*, 2004), suggesting that these cells likely utilize an alternative mode of regulating the cell cycle expression of histone proteins.

U7 snRNP

The histone downstream element (HDE) acts as a binding site for the U7 snRNP complex via basepairing with U7 snRNA (Fig. 5). The *Drosophila* U7 snRNA is a 71 nt snRNA that contains a trimethylguanosine cap at its 5' end and, like the spliceosomal snRNA genes U1-U5, is transcribed by RNA polymerase II. In *Drosophila*, the U7 snRNA gene is located within a large intron of the ecdysone-induced protein 63E (Eip63E) gene on chromosome 3L (Dominski *et al.*, 2003b). The mature U7 snRNA contains a 3' stem loop of

30 nt, immediately preceded by an Sm binding region that differs from that of the spliceosomal snRNPs, but it very well conserved with the Sm binding sites of U7 snRNAs in other organisms. This unique binding site allows the U7 snRNA to bind 5 canonical Sm proteins and two U7 snRNP-specific Sm-like (Lsm) proteins instead of the 7 Sm proteins bound by the spliceosomal snRNPs. This difference is crucial to the ability of the U7 snRNP to participate in processing histone pre-mRNAs. Upstream of the Sm binding site is a string of nucleotides that are complementary to the histone downstream element (HDE) within histone pre-mRNAs. The U7 snRNA is able to use this complementarity to bind the HDE and participate in the cleavage reaction, for which it is an essential component. Indeed, null mutation of the U7 snRNA gene in *Drosophila* results in production of only polyadenylated histone message (Godfrey *et al.*, 2006).

Until recently, very little was known about the 3' end processing reaction that gives rise to the mature U7 snRNA. Despite many of the spliceosomal snRNA processing complexes being at least partially characterized, nothing has come to light regarding such a complex for processing U7 snRNA. However, very recent data suggests that U7 snRNA 3' end formation proceeds in a similar manner to other snRNAs, with a multisubunit complex, termed Integrator, responsible for a co-transcriptional cleavage reaction resulting in formation of the snRNA and releasing the transcript from the template. In Chapter II, I will discuss experiments that were performed using a variety of reporters to address characteristics about U7 pre-snRNA processing, including whether or not the processing reaction is promoter and/or length-dependent.

U7 snRNP COMPONENTS – Lsm10 and Lsm11

In addition to U7 snRNA, the U7 snRNP is comprised of a heptameric ring of Sm proteins (Fig. 5), much like the Sm ring of spliceosomal snRNPs. In the spliceosomal snRNPs, this ring consists of seven conserved Sm proteins known as SmD1, SmD2, SmE, SmF, SmG, SmB/B', and SmD3. In the U7 snRNP, however, the SmD1 and SmD2 proteins are replaced by two Sm-like proteins, Lsm10 and Lsm11, respectively. Both of these Lsm proteins are required for proper processing of histone pre-mRNA, and depletion of one via RNAi results in depletion of the other (Wagner *et al.*, 2007), suggesting these proteins exist as a stable heterodimer in the cell. Lsm10, along with SmB and SmD3, binds to a region within the histone pre-mRNA that lies between the normal cleavage site and the U7 snRNA binding site. Together, these three proteins likely function as a type of molecular ruler that helps the processing machinery determine the proper site for cleavage (Yang *et al.*, 2009a).

Directly interacting with other members of the cleavage apparatus, Lsm11 is unique among known Sm and Sm-like proteins because of its long N-terminus, which is required for processing histone pre-mRNAs. In mammalian cells, the Lsm11 N-terminus interacts with a zinc finger protein known as ZFP100 (Fig. 5A), which forms a bridge between the U7 snRNP and SLBP and acts to stabilize the interaction between these two factors (Dominski *et al.*, 2002; Azzouz *et al.*, 2005; Wagner *et al.*, 2006). To date, no ZFP100 homolog has been found outside of mammals, and the sequence within Lsm11 that binds this protein is not well conserved. However, there are many other conserved residues within the N-terminus of Lsm11 that may be required for binding other factors. One such factor, FLASH (Fig. 5B), was recently identified as an Lsm11 binding partner and component of the HLB (Yang *et al.*, 2009b) and is described in Chapter IV.

The Cleavage Factor

All non-histone mRNAs in the cell become polyadenylated through a process that requires recognition of conserved sequence motifs in the pre-mRNA that direct cleavage, followed by the subsequent addition of several adenosine residues to the 3' end of the message, to form the mature messenger RNA. This process of cleavage and polyadenylation requires at least 17 proteins and as many as 100 proteins have been identified as possibly associated with the complex using a proteomics approach, although the precise function of many of these factors remains undefined. However, a subset of these proteins has been described, including the factors that comprise the cleavage and polyadenylation specificity factor (CPSF) complex, the cleavage stimulation factor (CstF) complex, and the cleavage factor (CF) components (Fig. 6).

There are two cis elements within the sequence of mRNAs that are required for efficient cleavage and polyadenylation. These include a polyadenylation signal, AAUAAA, which is highly conserved upstream of the cleavage site, and a G/U-rich region downstream. These sequences are bound by subunits of the cleavage machinery, with cleavage usually occurring 10-30 nt downstream of the polyadenylation signal. Once the message is cleaved, a poly(A) tail is added by poly(A) polymerase (PAP) and the mature mRNA is formed.

The CPSF complex is made up of several proteins, including CPSF-160 (CPSF1), CPSF-100 (CPSF2), CPSF-73 (CPSF3), CPSF-30 (CPSF4), Symplekin, and Fip1. The CPSF-160 subunit is responsible for binding to the polyadenylation signal (Murthy and Manley, 1995), possibly with help from two of the other subunits, CPSF-30 and Fip1 (Kaufmann *et al.*, 2004). CPSF-30 also forms important protein-protein interactions within the CPSF complex and Fip1 (factor interacting with Pap1) may have a further role in

recruitment of PAP to the polyadenylation site. CPSF-73 and CPSF-100 are both members of the metallo-β-lactamase superfamily of hydrolases (Aravind, 1999; Callebaut *et al.*, 2002; Dominski, 2007) and comprise the nuclease core of the cleavage complex. These two proteins form a heterodimer in the cell, interacting with each other through the non-catalytic C-terminus of CPSF-73 (Dominski *et al.*, 2005). Though both CPSF-73 and CPSF-100 contain metallo-β-lacatamase domains, only the domain associated with CPSF-73 appears able to bind zinc (Mandel *et al.*, 2006), with the CPSF-100 containing mutations in this domain that prevent zinc ion binding (Mandel *et al.*, 2006). This observation has led to the common assumption that CPSF-73 is active, while CPSF-100 is not, and the structure of the yeast CPSF-100 protein is consistent with this proposal (Mandel *et al.*, 2006). Indeed, most of the important metal binding residues in yeast CPSF-100 are not conserved (Aravind, 1999). However, in metazoans, there exist many residues within the metallo-β-lactamase domain of CPSF-100 that might be capable of binding metal ions (Kolev *et al.*, 2008). In support of this possibility, recent evidence reveals that mutation within the metallo-β-lactamase domain of either CPSF-73 or CPSF-100 is capable of reducing processing efficiency (Kolev *et al.*, 2008).

In addition to the contribution of the other CPSF subunits, Symplekin plays a key role in processing non-histone pre-mRNAs. Originally characterized as a component of cellular tight junctions (Keon *et al.*, 1996), symplekin likely performs the role of a molecular scaffold in the CPSF complex, not only linking the constituent subunits of the CPSF complex, but perhaps also joining these to the CstF complex via an interaction with CstF-64 (Takagaki and Manley, 2000).

The CstF complex is comprised of three proteins, CstF-50, CstF-64, and CstF-77. The primary function of the CstF complex is carried out by the CstF-64 subunit, which is responsible for binding the G/U-rich sequence on the mRNA (MacDonald *et al.*, 1994). This subunit also contains binding sites for Symplekin and the CstF-77 subunit (Takagaki and Manley, 2000). CstF-77 acts as a central organizer of the CstF complex, containing binding sites for both CstF-64 and CstF-50 (Takagaki and Manley, 2000), and CstF-50 is likely involved in mediating other protein-protein interactions within the cleavage complex. In addition, CstF-50 has been found to interact with the BRCA1-associated protein, BARD1, to form a complex that represses the cleavage reaction following DNA damage (Kleiman and Manley, 2001).

Two other proteins, CFI and CFII, are also required for efficient cleavage of non-histone mRNAs in mammals. CFI, which is comprised of two subunits (Ruegsegger *et al.*, 1996; Ruegsegger *et al.*, 1998), binds the mRNA upstream of the polyadenylation signal and facilitates CPSF binding (Brown and Gilmartin, 2003; Venkataraman *et al.*, 2005). The function of CFII is less clear and, beyond its requirement for cleavage, very few details have emerged (DeVries *et al.*, 2000).

Just prior to the genome-wide screen described in Chapter III, two of the CPSF subunits, CPSF-73 and Symplekin, were implicated in histone mRNA 3' end formation (Dominski *et al.*, 2005; Kolev and Steitz, 2005). UV crosslinking experiments in our lab utilizing a site-specific phosphothioate residue at the cleavage site resulted in the isolation of an 85 kDa protein that could be localized to the site of cleavage (Dominski *et al.*, 2005). Subsequently, immunoprecipitation experiments were performed that identified this protein as CPSF-73 (Dominski *et al.*, 2005). Interestingly, the same study found that CPSF-73 can

be UV crosslinked to the downstream cleavage product (DCP) of histone pre-mRNA, which is cleaved off of the histone pre-mRNA during maturation, suggesting that this protein may act as both an endonuclease and an exonuclease (Dominski *et al.*, 2005).

The other CPSF subunit to be identified as a component of the histone pre-mRNA processing machinery prior to the screen was Symplekin. Early experiments detected the presence of a heat labile factor (HLF) that is required for processing histone pre-mRNAs *in vitro*, though the identity of this factor remained unknown (Gick *et al.*, 1987). This factor was later identified through complementation studies as Symplekin (Kolev and Steitz, 2005). In these experiments, heat treated nuclear extract which was inactive for histone pre-mRNA processing was supplemented with Symplekin, which restored the activity of the extract (Kolev and Steitz, 2005). As is the case with the non-histone cleavage apparatus, it seems likely that this protein forms a scaffold that is responsible for organizing other components of the histone pre-mRNA processing machinery.

Following the genome-wide screen, described in Chapter III, a number of factors emerged as functional components of the processing machinery. In addition to identifying CPSF73 and Symplekin, two other CPSF subunits, CPSF-100 and Fip1, scored positively (Wagner *et al.*, 2007). While the role of Fip1 in processing is unclear, CPSF-100 has been better characterized and, as it does with non-histone pre-mRNAs, likely forms a heterodimer with its fellow metallo-β-lactamase domain-containing protein, CPSF-73. These two proteins, along with Symplekin, form a core histone pre-mRNA cleavage complex that is essential for processing histone pre-mRNA (Sullivan *et al.*, 2009b). In support of this, chromatin immunoprecipitation experiments reveal that both CPSF-73 and Symplekin associate with histone genes cotranscriptionally (Sullivan *et al.*, 2009b). In addition, RNAi-

mediated depletion of the other CPSF or CstF factors has no effect on histone pre-mRNA processing (Wagner *et al.*, 2007), indicating that not all of the cleavage and polyadenylation factors are required for processing histone mRNA. Further, while RNAi-mediated depletion of CPSF-160 or CstF-64 results in decreased levels of Symplekin, it does not affect the association of Symplekin with histone genes (Sullivan *et al.*, 2009b). Together, these results argue for the existence of two similar, but distinct cleavage factors that are separately responsible for processing either polyadenylated mRNAs or histone mRNAs.

DISSERTATION GOALS

Histone expression is highly regulated and is tightly linked to DNA replication and the cell cycle. Regulation of this expression occurs mainly at the mRNA level, with key controls on transcription, 3'end formation/maturation, and message stability. The primary target is 3' end processing of the histone pre-mRNA, which relies on a unique complement of cis elements and trans factors for efficient cleavage. Among these are the SLBP/stem loop complex, the histone downstream element and the U7 snRNP, including ZFP100 in mammals, and two components of the cleavage and polyadenylation complex, CPSF73 and Symplekin. As of the start of this dissertation project, this was the extent of our knowledge about the composition of the histone pre-mRNA processing complex. However, despite the identification of these key processing factors and other efforts to identify additional components of the histone pre-mRNA processing apparatus, the complete characterization of this complex machinery and its constituent parts remained unsolved. This deficiency argued

for a novel approach to identify these missing links and led to the creation of a histone pre-mRNA processing reporter that we used to screen the *Drosophila* genome.

In Chapter II, I will describe reporter-based methods for both small-scale and genome-wide screening and how they have been used to identify putative 3' end processing factors. In Chapter III, I will describe a genome-wide screen that we performed using one of these reporters and highlight a list of 24 proteins that we identified as being involved in processing histone pre-mRNA (Wagner *et al.*, 2007). In Chapter IV, I will describe another factor, FLASH, that escaped validation in the initial screen, but was subsequently picked up using a yeast-two-hybrid screen for factors that interact with the N-terminus of Lsm11 (Yang *et al.*, 2009). I will give evidence that supports the role of dFLASH as a histone pre-mRNA processing factor and will further characterize its interaction with Lsm11 and the implications of this interaction on histone pre-mRNA processing and localization of these two proteins to the histone locus body.

Figure 1-1. Histone genes are organized into gene clusters that are important for modulating histone expression. Diagram of the *Drosophila* histone gene cluster, which contains all five replication-dependent histone genes arranged in tandem and repeated about 100 times. The stemloop and HDE at the end of the histone H3 gene are shown and the cleavage site is indicated with an arrow. Downstream of the region encoding the normal mature histone mRNA transcript are cryptic polyadenylation signals that are used if normal processing of the histone mRNA is disrupted. These signals likely evolved as a means to prevent transcriptional readthrough.

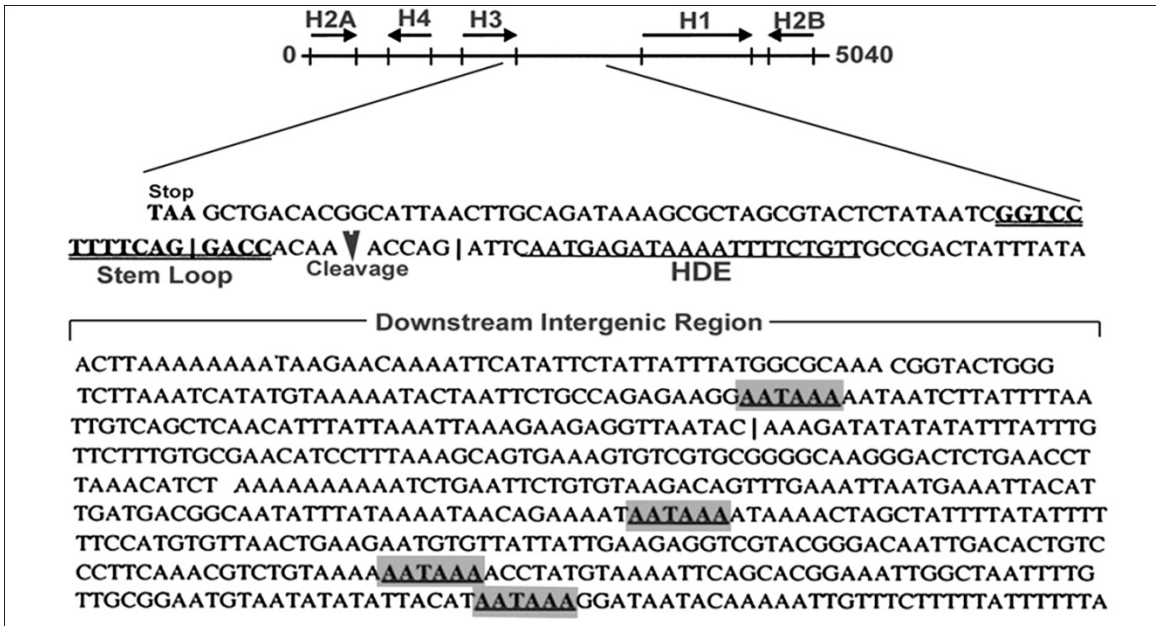

H2A
H4
H3
H1
H2B
0
5040
Stop
TAA GCTGACACGGCATTAACTTGCAGATAAAGCGCTAGCGTACTCTATAATCGGTCC
TTTTCAG|GACCACAA ACCAG|ATTCAATGAGATAAAATTTTCTGTTGCCGACTATTTATA
Stem Loop
Cleavage
HDE
Downstream Intergenic Region
ACTTAAAAAAAATAAGAACAAAATTCATATTCTATTATTTATGGCGCAAA CGGTACTGGG
TCTTAAATCATATGTAAAAATACTAATTCTGCCAGAGAAGGAATAAAAATAATCTTATTTTAA
TTGTCAGCTCAACATTTATTAAATTAAAGAAGAGGTTAATAC|AAAGATATATATATTTATTTG
TTCTTTGTGCGAACATCCTTTAAAGCAGTGAAAGTGTCGTGCGGGGCAAGGGACTCTGAACCT
TAAACATCT AAAAAAAAAATCTGAATTCTGTGTAAGACAGTTTGAAATTAATGAAATTACAT
TGATGACGGCAATATTTATAAAATAACAGAAAATAATAAAATAAAACTAGCTATTTTATATTTT
TTCCATGTGTTAACTGAAGAATGTGTTATTATTGAAGAGGTCGTACGGGACAATTGACACTGTC
CCTTCAAACGTCTGTAAAAAATAAAACCTATGTAAAATTCAGCACGGAAATTGGCTAATTTTG
TTGCGGAATGTAATATATATTACATAATAAAGGATAATACAAAAATTGTTTCTTTTTATTTTTTA

Figure 1-2. Histone mRNA expression is highly regulated and tightly linked to DNA replication and the cell cycle. Diagram of the normal cell cycle-regulated expression of mammlian histone mRNA and its response to inhibition of DNA synthesis. Levels of histone mRNA are low during G1, but rapidly increase upon entry into S phase. This increase requires activation of cyclinE/Cdk2 and its phosphorylation of NPAT, which localizes to the histone locus body and is required for histone mRNA expression. At the end of normal S phase or in response to inhibition of DNA replication, histone mRNA levels are rapidly reduced. SLBP protein levels are also linked to the cell cycle, increasing rapidly toward the end of G1 and decreasing rapidly at the end of S phase in response to the activity of cyclinA/Cdk1. The existence of a checkpoint during G1 that arrests cells with defects in the U7 snRNP has been suggested (Wagner and Marzluff, 2007; arrow). (Figure adapted from Marzluff *et al.*, 2009)

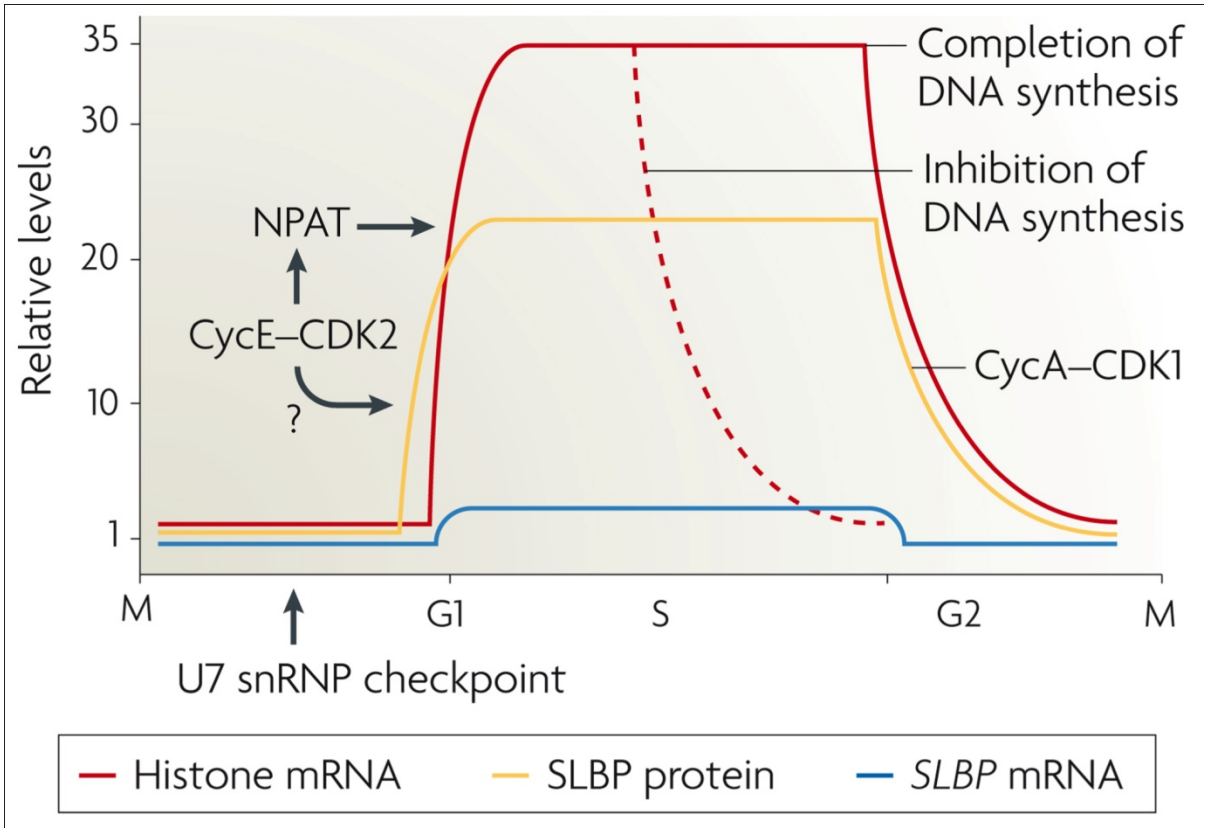

Relative levels
35
30
20
10
1
NPAT
CycE–CDK2
?
Completion of DNA synthesis
Inhibition of DNA synthesis
CycA–CDK1
M
G1
S
G2
M
U7 snRNP checkpoint
Histone mRNA
SLBP protein
SLBP mRNA

Figure 1-3. Metabolism of replication-dependent histone mRNAs in mammals. Histone expression begins with the activation of the histone locus through phosphorylation of NPAT and subsequent transcription of the histone mRNA. Histone pre-mRNAs are processed by a single endonucleolytic cleavage event requiring an assembly of several processing factors, including the stem loop binding protein (SLBP), U7 snRNP, and the cleavage factor complex of CPSF73, CPSF100, and Symplekin. Following cleavage, the mature histone mRNA, still bound by SLBP, is exported into the cytoplasm where it is circularized by a complex including SLBP, SLBP-interacting protein 1 (SLIP1), and eukaryotic inititation factor 4γ (eIF4G) and the histone mRNA is translated. At the end of S phase, the histone mRNA is oligouridylated by a terminal uridylyl transferase (TUTase). The oligo(U) tail is then bound by Lsm1-7, the decapping complex and exosome are recruited, and the histone mRNA is degraded. SLBP is targeted for degradation by cyclinA/Cdk1 phosphorylation. Abbreviations: CPSF = cleavage and polyadenylation specificity factor; DCP = mRNA decapping enzyme; HDE = histone downstream element; Lsm = Sm-like protein; NPAT = nuclear protein, ataxia-telangiectasia locus; RNA Pol II = DNA-dependent RNA polymease II; TUTase = terminal uridylyl transferase; XRN1 = 5'→3' exoribonuclease 1. (Adapted from Marzluff *et al.*, 2009)

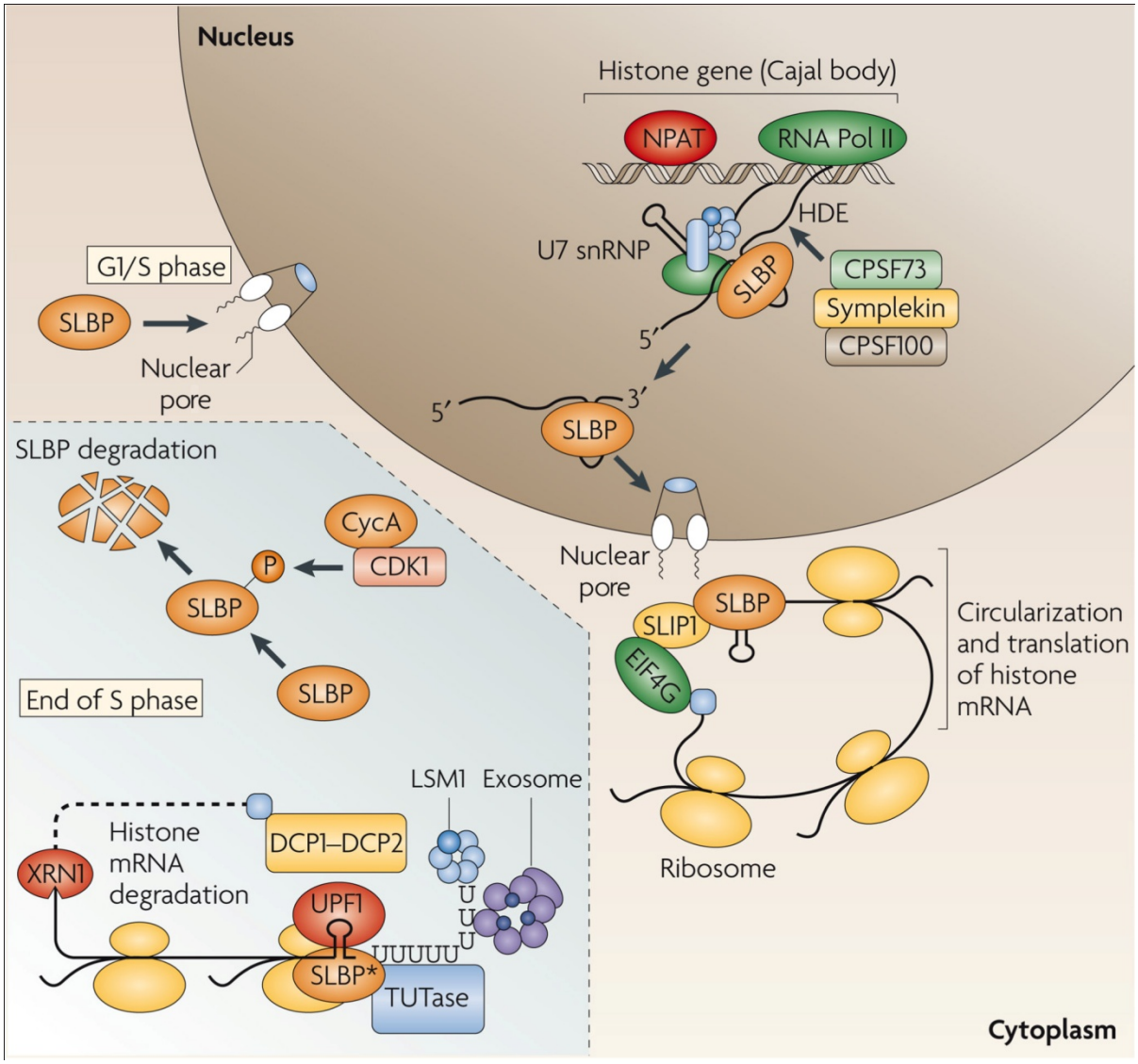

Nucleus
Histone gene (Cajal body)
NPAT
RNA Pol II
HDE
U7 snRNP
SLBP
CPSF73
Symplekin
CPSF100
5′
3′
5′
SLBP
G1/S phase
SLBP
Nuclear pore
SLBP degradation
CycA
CDK1
P
SLBP
SLBP
End of S phase
Nuclear pore
SLBP
SLIP1
EIF4G
Circularization and translation of histone mRNA
Ribosome
LSM1
Exosome
Histone mRNA degradation
DCP1–DCP2
XRN1
UPF1
U
U
U
UUUUU
SLBP*
TUTase
Cytoplasm

Figure 1-4. Histone mRNAs have a unique 3' end. Histone pre-mRNAs contain two cis elements, a stem loop and histone downstream element (HDE) that are bound by essential processing factors. The stem loop is bound by stem loop binding protein (SLBP) and the HDE is bound by the U7 snRNP through basepairing interactions with the U7 snRNA. Following cleavage, the stem loop remains bound by SLBP and the mature histone mRNA remains nonpolyadenylated. Because histone pre-mRNAs do not contain introns, this single cleavage event is all that is required to form the mature histone mRNA.

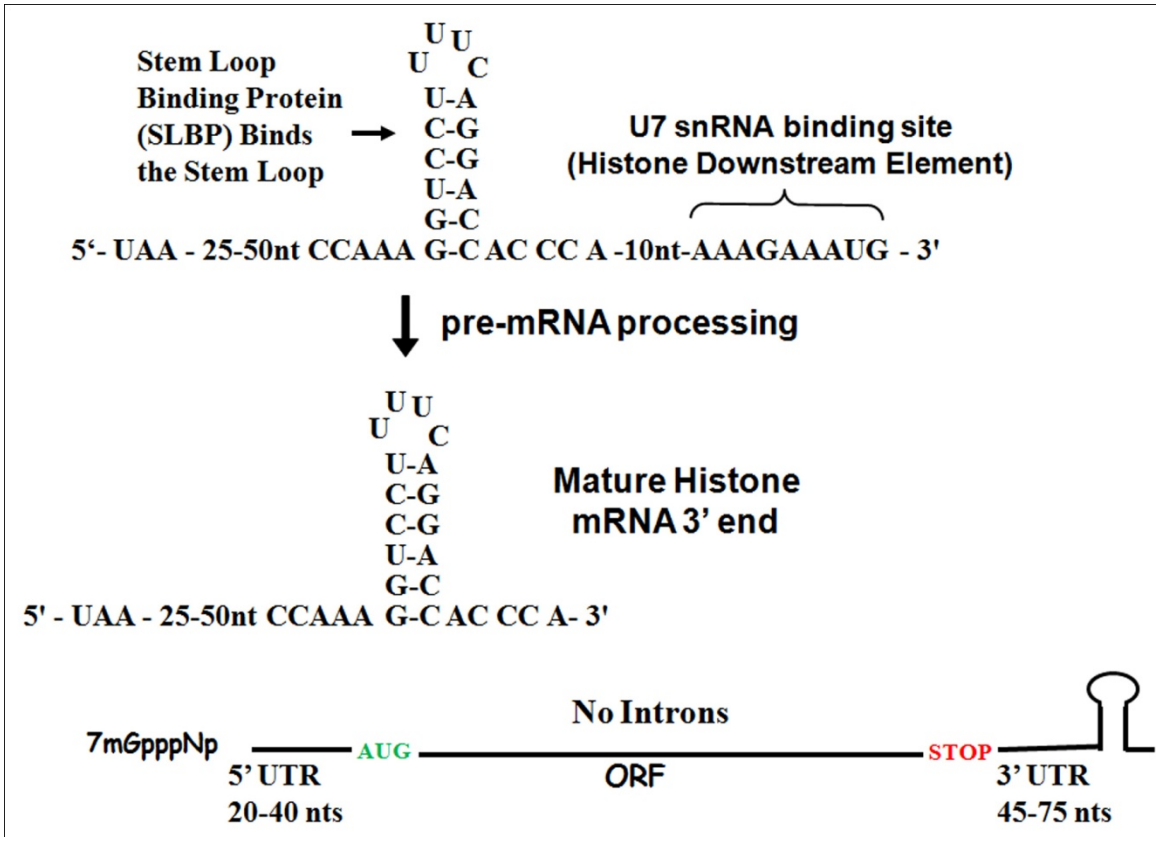

Stem Loop
Binding Protein
(SLBP) Binds
the Stem Loop
U U
U C
U-A
C-G
C-G
U-A
G-C
U7 snRNA binding site
(Histone Downstream Element)
5‘- UAA - 25-50nt CCAAA G-C AC CC A -10nt-AAAGAAAUG - 3'
pre-mRNA processing
U U
U C
U-A
C-G
C-G
U-A
G-C
Mature Histone
mRNA 3’ end
5' - UAA - 25-50nt CCAAA G-C AC CC A- 3'
No Introns
7mGpppNp
AUG
STOP
5’ UTR
20-40 nts
ORF
3’ UTR
45-75 nts

Figure 1-5. The mammalian histone pre-mRNA processing apparatus. (A) Model of known mammalian histone pre-mRNA processing factors at the beginning of my research (c. 2004). Factors that were known at the time include the stem loop binding protein (SLBP), U7 snRNA, the U7 snRNP components Lsm10 and Lsm11, and a zinc finger protein (ZFP100) that interacts with both SLBP and Lsm11 and forms bridge between these two proteins. (B) Our current model of the histone pre-mRNA processing machinery, which includes cleavage factor components CPSF73, CPSF100, and Symplekin and the pro-apoptotic factor FLASH, in addition to the factors listed in Panel A. It is still not known how the cleavage factor interacts with the processing complex, but possibly involves an unknown factor that links the two. This unknown factor is depicted in the model as factor "X."

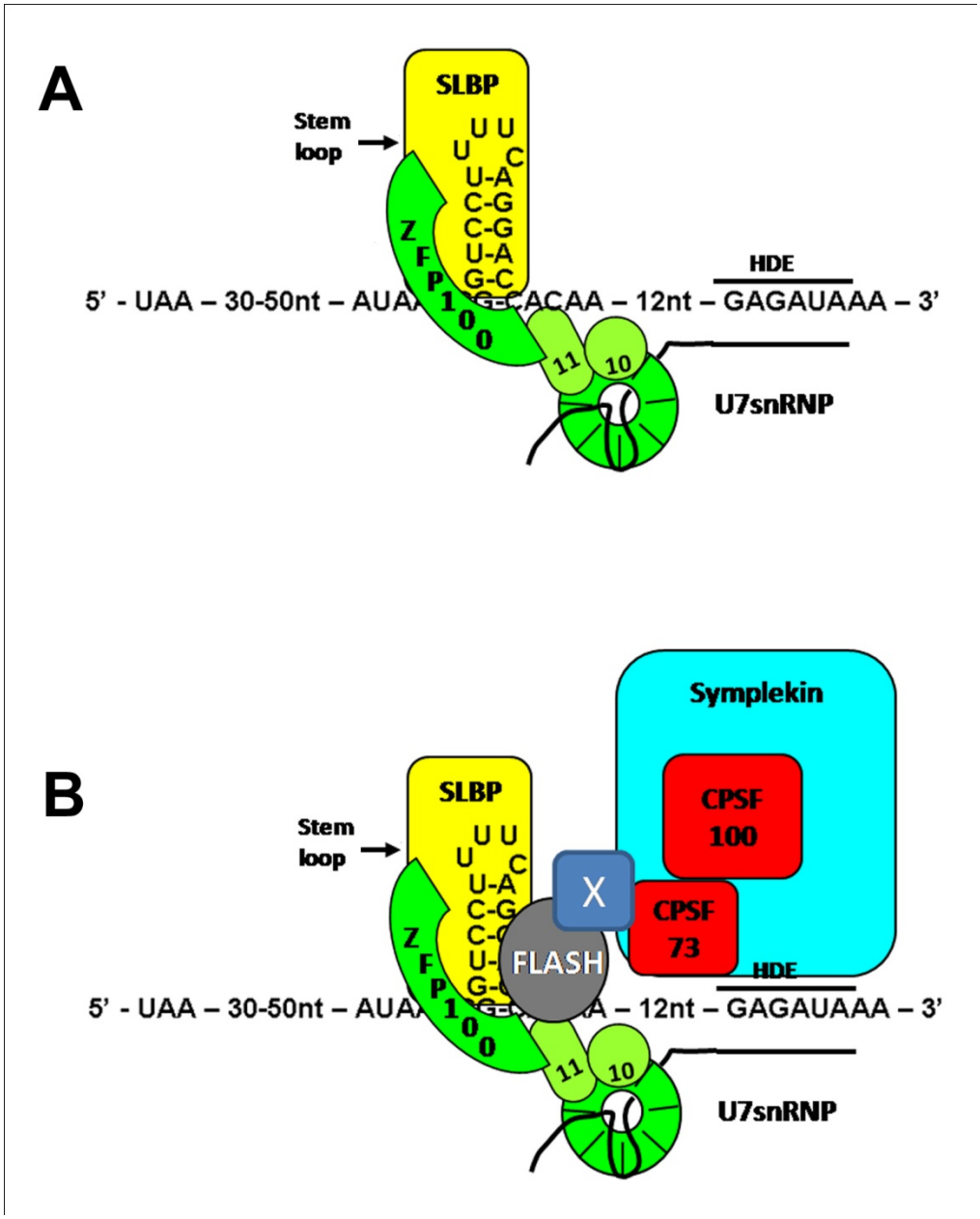
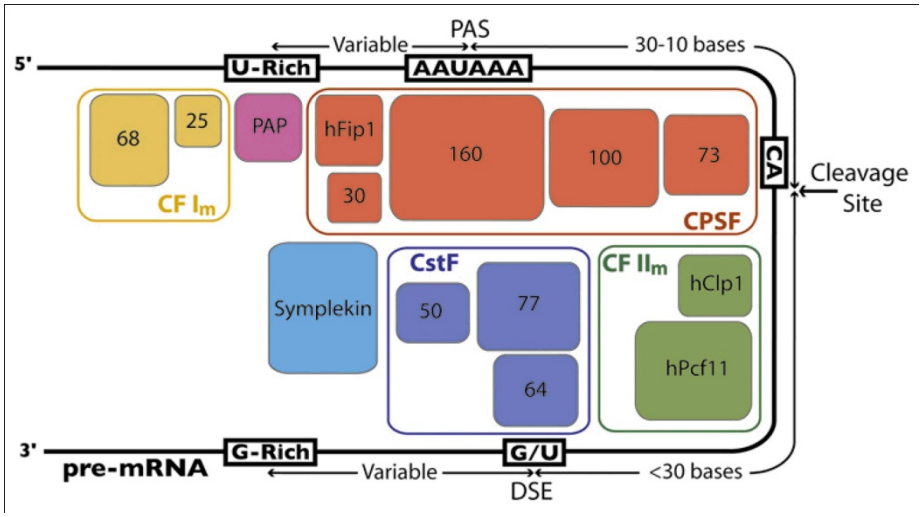

A
SLBP
Stem loop
U U
U C
U-A
C-G
C-G
U-A
G-C
ZFP100
HDE
5' - UAA – 30-50nt – AUA
CACAA – 12nt – GAGAUAAA – 3'
11
10
U7snRNP
B
Symplekin
SLBP
Stem loop
CPSF 100
X
CPSF 73
FLASH
HDE
5' - UAA – 30-50nt – AUA
A – 12nt – GAGAUAAA – 3'
11
10
U7snRNP

Figure 1-6. The mammlian cleavage and polyadenylation machinery used to process polyadenylated mRNAs. Model of the current understanding of the cleavage and polyadenylation machinery. Included in this model is the role of CPSF160 in binding the polyadenylation signal (AAUAAA), the endonucleolytic action of CPSF73 at the cleavage site just downstream of a CA dinucleotide, and the recognition of the G/U rich element by CstF64. The complete composition of this complex is unknown. (Adapted from Mandel *et al.*, 2008)


PAS
Variable
30-10 bases
5'
U-Rich
AAUAAA
68
25
CF Im
PAP
hFip1
30
160
100
73
CPSF
CA
Cleavage
Site
Symplekin
CstF
50
77
64
CF IIm
hClp1
hPcf11
3'
pre-mRNA
G-Rich
G/U
Variable
<30 bases
DSE

CHAPTER II

DEFINING RNA 3' END FORMATION USING REPORTERS

Introduction

Much success has been achieved over the years identifying and characterizing genes and gene products using traditional genetic and biochemical methods. Among the earliest studies were forward genetic screens employing mutagenic agents to introduce single changes in the genome of yeast strains and subsequent visual observation of altered phenotypes. By accumulating large numbers of mutants displaying the same or similar phenotypes and using the classical genetics tools available to them, researchers were able to identify genes that are linked to these phenotypes and often could map out the molecular pathways that controlled a given phenotype. Because of the success of these genetic screens and the availability of highly tractable model organisms, such techniques have also seen great use in a number of other eukaryotes, including flies and worms.

With the advent of the genomic age, large volumes of gene sequence data have become available, culminating in complete or nearly complete genome sequences for a wide number of organisms. This information, coupled with an ever expanding molecular toolbox, has allowed for novel methods of gene discovery to emerge. For example, instead of randomly mutating an organism and looking for phenotypes to identify candidate genes, as per the classical methods of gene discovery, researchers are now able to take a more targeted approach using sequence data readily available in online databases to identify candidate

genes *in silico*. Once genes are identified, the researcher can mutate the genes in the whole organism or use methods of gene silencing in cell culture to allow characterization of the factor of interest. A great advantage of this method is that putative homologues of a factor can often be identified through sequence similarity searches, and conserved domains can be uncovered that may give clues about the function of the factor, which would not have been possible before. However, though this method of highly targeted *in silico* screening is effective, unknown factors of interest may not be readily identifiable through these means. Fortunately, other avenues have opened up that allow researchers to use the power of the genome to quickly and efficiently identify factors involved in their process of interest.

One of the most powerful of these methods of gene discovery is the use of RNA interference (RNAi) in cultured cells. On a genome-wide scale, this method allows for the development of rapid and convenient reverse genetic screens to identify novel factors involved in a pathway of interest (Foley and O'Farrell, 2004; Paddison *et al.*, 2004; Armknecht *et al.*, 2005; Silva *et al.*, 2005; Echeverri and Perrimon, 2006; Ramadan *et al.*, 2007). Though genome-wide RNAi screens are a relatively new development, much progress has been made on this front over the past decade. In particular, recent years have seen the construction of various RNAi libraries that allow high-throughput, cell-based screening of the *Drosophila* genome by means of RNA interference (Foley and O'Farrell, 2004; Armknecht *et al.*, 2005). The increased availability of these resources has allowed more and more researchers the opportunity to perform genome-wide searches for factors that participate in their cellular process of interest. There are many possible designs for these screens, depending upon the information that the researcher wishes to gain, but most utilize a visual readout. These readouts can be as simple as which dsRNA targets result in growth and

viability defects (Boutros *et al.*, 2004) or how the cell morphologys (Kiger *et al.*, 2003). Other experimental designs may utilize antibody staining to determine factors that are required for localization or posttranslational modification of another protein, such as the MPM-2 antibody screen introduced in this chapter. Still others may employ reporter constructs to detect factors involved in a specific biochemical process or pathway, such as the genome-wide screen for factors involved in the histone pre-mRNA cleavage reaction described in Chapter III (Wagner *et al.*, 2007) or the targeted identification of dFLASH as a histone pre-mRNA processing factor (Yang *et al.*, 2009), as described in Chapter IV, as a result of biochemical experiments.

The basis of any screen that utilizes these technologies is the gene-silencing mechanisms already in place within the cell and commonly referred to as the RNAi pathway. Though RNAi mechanisms exist in most eukaryotes, for molecular biological studies there are advantages to choosing one organism over another. Certainly, the most appropriate choice for a system that is directly applicable to problems of human health is a human, or at least mammalian, cell line. However, mammalian cells present challenges that have been difficult or expensive to overcome, including the high cost of siRNAs and the limited availability of appropriate genome-wide libraries. The recent arrival of short hairpin RNA libraries (Paddison *et al.*, 2004; Silva *et al.*, 2005) has helped alleviate some of these challenges, but complete libraries are still difficult to come by and are oftentimes prohibitively expensive.

Another option for RNAi screening is the use of C. elegans genome-wide RNAi feeding libraries (Fraser *et al.*, 2000; Kamath and Ahringer, 2003), which consist of thousands of T7 RNA polymerase encoding, RNase-deficient bacterial strains containing

individual C. elegans ORFs or partial ORFs flanked by T7 promoters. An advantage of this is using bacteria to synthesize dsRNA instead of having to perform in vitro transcription reactions for every target. Because worms eat bacteria, this allows the introduction of siRNAs directly into the gut of the worm, where it can be delivered systemically into all the cells of the body (Fraser *et al.*, 2000; Kamath and Ahringer, 2003). Indeed, worms have the ability to amplify RNAi signal in a non-cell autonomous fashion that makes this method highly effective (Fire *et al.*, 1998).

However, if a cell-based system is desired, one of the most successful strategies in recent years has been the use of genome-wide dsRNA libraries on *Drosophila* cultured cells. There are many advantages to using a *Drosophila* system for study. From a biological perspective, many genes and molecular pathways found in human cells are very well conserved in flies, allowing researchers the ability to advance knowledge of their process of interest using a highly tractable and well characterized model system in a way that is highly applicable to human biology. From a technical standpoint, one highly useful property of the RNAi pathway in *Drosophila* is that, unlike mammalian cells which contain an interferon response to long dsRNAs that results in global translational inhibition (Williams, 1999), *Drosophila* cells can efficiently process these long dsRNAs into siRNAs that are able to downregulate expression of a single target gene (Caplen *et al.*, 2000; Hammond *et al.*, 2000). Thus, instead of having to design multiple siRNAs that target independent 21-23 nt regions of an mRNA in hopes of identifying one that is sufficiently effective at knocking down the message of interest, one is able to select a region within the mRNA of approximately 200-600 nt in length that will be targeted by a longer dsRNA. When introduced into the cell, these long dsRNAs are processed by Dicer into dozens of siRNAs that target various

sequences within the 200-600 nt region, essentially resulting in a library of siRNAs in a single cell (Hammond *et al.*, 2000). These long dsRNAs can be synthesized both conveniently and economically in-house using standard in vitro transcription protocols followed by annealing of the two strands of RNA together. A further advantage of using *Drosophila*, and specifically *Drosophila* cultured cells, is that, unlike mammalian cells, the dsRNAs do not have to be transfected into the cells. Instead, the dsRNAs are simply added to the culture medium (Clemens *et al.*, 2000) and the cells take up the dsRNAs by endocytosis (Saleh *et al.*, 2006). In addition, a major benefit of using *Drosophila* cultured cells is that the research can be seamlessly transferred over into an extremely well-characterized and genetically tractable whole organism for further study. This allows for more in-depth characterization of a factor of interest, including tissue-specific and developmental phenotypes, than would be possible in a cell culture system alone.

For the studies presented in this chapter, a reporter-based method of screening cells that have been RNAi-depleted for various factors has been employed. These reporters allow fast and efficient detection of factors of interest for a host of processes, including transcription of histone genes and 3' end processing of both histone pre-mRNAs and U7 pre-snRNA. Though the design of each construct is unique to its particular purpose, they all share several key features. First, each relies on a visual readout for detection of factors/outcomes of interest. To this end, an enhanced green fluorescent protein (EGFP) open reading frame (ORF) has been included in each of the constructs, allowing signal off the reporter to be detected by fluorescence microscopy. Another aspect of the reporters is that they are expressed efficiently in the cell. Most of the promoters that were used in these experiments represent promoters that exist within the *Drosophila* genome, such as the

histone H3 promoter and the Actin 5C promoter. Each promoter was empirically tested for efficacy in *Drosophila* cultured cells and the ultimate choice of promoter was made based upon this information and the dictates of the experimental design. Finally, each of the reporters contains a strong insect polyadenylation signal downstream of the reporter sequence, allowing for efficient expression of stable reporter mRNAs.

The processes being investigated in these experiments include histone transcription, histone pre-mRNA processing, and U7 snRNA 3' end formation. That each of these seemingly disparate processes can be illuminated using similar methods of inquiry, namely, the use of reporters to detect factors that are involved in each of these processes, speaks to the power of the reporter method and gives credence to the notion that these types of experiments, while not complete in and of themselves, can provide an important stepping stone for further study and allow the identification of candidate factors that otherwise may be nearly impossible to pinpoint.

Through these experiments, I find that the reporters I have used and tested are accurate representations of the cellular processes they model and further, that these assays are a highly effective method to test for transcription and mRNA processing factors. These assays have allowed me and others to identify several putative histone pre-mRNA processing factors (Wagner *et al.*, 2007; Yang *et al.*, 2009; White *et al.*, in preparation), as well as to gain new insight into the U7 snRNA processing reaction and the identity of U7 snRNA processing factors (Wagner *et al.*, in preparation). To this end, the experiments contained in this chapter are the building blocks of these many lines of research.

Materials and Methods

Creation of Reporter Constructs

The His_{proc} reporter was created by subcloning the histone H3 promoter, partial open reading frame, downstream sequence, and an EGFP coding sequence into a promoterless pIZ/V5/His vector backbone (Invitrogen). Briefly, the histone H3 5'UTR and amino acids 1-67 were amplified by PCR using forward primer 5' GGCCGAATTCCGACAAAAAC CCGAGAGAGTAC 3' and reverse primer 5' GGCCGGTACCTTAGGCAGCTTG CGGATTAGAAGC 3' and subcloned into pEGFPN1 (Clontech, Palo Alto, CA) using EcoRI and KpnI. The 3' end of the H3 gene starting immediately after the stop codon and continuing until 18 nt downstream of the HDE was amplified using forward primer 5' GGCCGGTACCACTTGCAGAT AAAGCGCTAGCG 3' and reverse primer 5' GGCCGGATCCTTGTTATAAATAG TCGGCAACA GAAAATTTTTTCTC 3' followed by ligation to the 5' product using KpnI and BamHI. The resulting construct contained the H3 promoter, amino acids 1-67 of histone H3 ORF, and H3 3' end containing a portion of the downstream intergenic region upstream of an EGFP ORF. Note that the H3 ORF is in frame with GFP and that we created a single mutation (U to A) within the HDE in order to disrupt a stop codon (red box in Fig. S1). The OpIE2 promoter was removed from pIZ-V5/His (Invitrogen, Carlsbad CA) vector by inverse PCR using forward primer 5' GGCCGCTAGCACAG CATCTGTTCGAATTTA 3' and reverse primer 5' GGCCGCTAGCAGACAT GATAAGATACATTGATGA 3' followed by digestion with NheI and religation. The reporter was then subcloned into the promoterless pIZ/V5/His vector using forward primer 5'

GGCCGAATTCCGACAAAAACCCGAGAGAG TAC 3' and reverse primer 5' GGCCTCTAGATTACTTGTACAGCTCGTCCAT GCC 3' followed by digestion with EcoRI and XbaI and subsequent ligation to form the pIZ/H3p/H3/GFP construct. Note that the polyA site used by the GFP ORF is the insect OpIE2 polyadenylation sequences 3' of the XbaI site in the pIZ/V5/His vector.

To create the Act_{proc} construct, the *Drosophila* Actin 5C promoter was amplified from genomic DNA by PCR using forward primer 5' GGCCAAGCTTTATGTATGTTTTGGCATACAATGAGTAGTTGG 3' and reverse primer 5' CCGGGAATTCGTGTCGGGAGGAGTATCCAC 3' and subcloned into the pIZ/promoterless/H3/GFP plasmid using HindIII and EcoRI. The resulting Act_{proc} reporter gene consisted of the Actin 5C promoter followed by a region encoding the first 67 amino acids of histone H3, a portion of the histone H3 3' UTR that included the SL and HDE, an in-frame GFP ORF, and a vector-encoded OpIE2 poly(A) signal.

The His_{trans} construct was created by amplifying the histone H3 promoter using forward primer 5' GGCCAAGCTTCGACAAAAACCCGAGAGAGTAC 3' and reverse primer 5' GGCCACCGGTCTCCGATTTGGGTTTCACTAAAGTTCACGTTC 3' and subcloning this piece into the pIZ/promoterless/H3/GFP plasmid using HindIII and AgeI. The resulting His_{trans} reporter consisted of the Histone H3 promoter followed by an EGFP ORF and a vector-encoded OpIE2 poly(A) signal.

The Act_{trans} construct was created by amplifying the Act 5C promoter with forward primer 5' GGCCAAGCTTTATGTATGTTTTGGCATACAATGAGTAGTTGG 3' and reverse primer 5' CCGGGAATTCGTGTCGGGAGGAGTATCCAC 3' and the EGFP ORF using forward primer 5' GGCCGAATTCATGGTGAGCAAGGGCGAGGAG 3' and reverse

primer 5’ GGCCTCTAGATTACTTGTACAGCTCGTCCATGCC 3’ and performing a three way ligation into the pIZ/promoterless vector using HindIII, EcoRI, and XbaI. The resulting Act_{trans} reporter consisted of the Actin 5C promoter followed by an EGFP ORF and a vector-encoded OpIE2 poly(A) signal.

The original pIZ/U7p/U7/GFP construct was created by Dr. Eric J. Wagner. A modular version of this reporter allowing simplified promoter and transcript swapping was created by amplifying the U7 snRNA promoter using forward primer 5’ GGCCAAGCTTGAAACTTACAAGTTAACTTAAAGCC 3’ and reverse primer 5’ GGCCGAATTCCTCAACAGAGGCTTGCAGAG 3’, the U7 snRNA transcript + EGFP using forward primer 5’ GGCCGAATTCATTGAAAATTTTTATTCTCTTTGAAATTTG 3’ and reverse primer 5’ GGCCTCTAGATTACTTGTACAGCTCGTCCATGCC 3’, and subcloning these into the pIZ/promoterless vector using HindIII, EcoRI, and XbaI. The resulting pIZ/U7p/U7/GFP reporter consisted of the U7 snRNA promoter followed by the U7 snRNA transcript, including downstream sequences, and and EGFP ORF, all followed by a vector-encoded OpIE2 poly(A) signal. To create the pIZ/Act5C/U7/GFP construct, the U7 snRNA transcript + EGFP was amplified off of the pIZ/U7p/U7/GFP reporter using forward primer 5’ GGCCGAATTCATTGAAAATTTTTATTCTCTTTGAAATTTG 3’ and reverse primer 5’ GGCCTCTAGATTACTTGTACAGCTCGTCCATGCC 3’ and subcloning this fragment into the pIZ/Act5Cp/H3/GFP vector using EcoRI and XbaI. The resulting pIZ/Act5Cp/U7/GFP reporter consisted of an Actin 5C promoter followed by the U7 snRNA transcript, including downstream sequences, and and EGFP ORF, all followed by a vector-encoded OpIE2 poly(A) signal. To create the pIZ/U7p/U7-H3insert/EGFP reporter, an XhoI restriction site was introduced into the U7 snRNA transcript by amplifying the the U7

promoter and 5’ portion of the U7 snRNA transcript off of the pIZ/U7p/U7/GFP reporter using forward primer 5’ GGCCAAGCTTGAAACTTACAAGTTAACTTAAAGCC 3’ and reverse primer 5’ GGCCCTCGAGAAAGAGAATAAAAATTTTCAATCTCAACAG 3’ and by amplifying the 3’ portion of the U7 snRNA transcript + EGFP off the pIZ/U7p/U7/GFP reporter using forward primer 5’ CCGGCTCGAGGAAATTTGTCTTGGTGGGAC 3’ and reverse primer 5’ GGCCTCTAGATTACTTGTACAGCTCGTCCATGCC 3’ and subcloning these pieces into the pIZ/promoterless vector using HindIII, XhoI, and XbaI, resulting in the pIZ/U7p/U7+XhoI/GFP construct. A portion of the histone H3 open reading frame was amplified using forward primer 5’ CGCGCTCGAGGCAAATCGACTGGTGGAAAG 3’ and reverse primer 5’ CCGGCTCGAGGCACACAAGTTGGTATCTTCGA 3’ and subcloned in the reverse orientation into the the pIZ/U7p/U7+XhoI/GFP construct using XhoI. Subcloning in the reverse orientation is essential to eliminate a cryptic start codon within the H3 insert sequence in the forward orientation that puts the ATG of the EGFP ORF out of frame. The resulting pIZ/U7p/U7-H3insert/EGFP reporter consisted of a U7 snRNA promoter, followed by a U7 snRNA transcript with an inserted histone H3 sequence, followed by downstream U7 snRNA sequence, an EGFP ORF, and a vector-encoded OpIE2 poly(A) signal.

Generation of dsRNAs and RNA interference

DNA templates for double stranded RNAs (dsRNAs) were generated by PCR using *Drosophila* Gene Collection library clones (Open Biosystems) or genomic DNA, if appropriate, and primers containing a T7 promoter (Table 2-1). Transcription was carried out in a total volume of 100 μl using T7 RNA polymerase and RiboMAX kit (Promega), as

recommended by the manufacturer. Following transcription, the RNA was treated with 5 units of RNase-free DNase Q (Promega), purified on G-50 micro columns (GE Healthcare), boiled for 5 minutes and cooled to room temperature to form dsRNA.

To induce RNAi-mediated knockdown of a factor of interest, dsRNAs targeting the factor were introduced into *Drosophila* D.Mel-2 cultured cells by adding the dsRNAs directly to the growth medium. For 384-well experiments, 5 ul diluted dsRNA containing 250 ng dsRNA in water was added to 10 µl SF-900II SFM (Gibco) containing $8x10^3$ cells in a black-walled, clear bottom tissue culture plate (Corning) in a final volume of 15 µl. For 96-well plates, $8x10^4$ cells were plated in 40 µl SF-900II SFM and supplemented with 800 ng dsRNA. For 24-well plates, $5x10^5$ cells were plated in 250 µl SF-900II SFM to which 5 µg dsRNA was added. For 6-well plate experiments, $3x10^6$ cells were plated in 1 ml SF-900II SFM and 30 µg dsRNA was added. For all experiments, the cells were allowed to knockdown at 27 °C in a humidified growth chamber for 3-6 days, depending upon the experiment and dsRNA target. If RNAi was to proceed longer than 3 days, an amount of dsRNA equal to the original application was added every 2 days to ensure continuous knockdown for all plating formats except 384-well format, for which dsRNA was added only on Day 0.

Cell Transfections and Reporter Assays

For the initial optimization experiments using the His_{proc} construct, cells were treated dsRNA and allowed to knock down for 3 days, followed by transfection of the reporter construct into the cells using Effectene reagent (Qiagen) on Day 3 and viewing of the cells by fluorescence microscopy on Day 5. For 384-well experiments, 20 µl fresh SF-900II SFM

was added on Day 3 to bring the total volume to approximately 35 µl. A master mix was created for all wells to be transfected such that each well contained the following: 50 ng of reporter plasmid was diluted into 8.75 µl Buffer EC per well. To this mixture, 0.2 µl of enhancer reagent was added, followed by incubation at room temperature for 5 min. Next, 0.3 µl of Effectene transfection reagent was added and the mixture incubated at room temperature for 10 min., after which, 10 µl was dispensed into each well. The cells were allowed to grow at 27 °C for 2 days, after which the cells were viewed with an inverted fluorescence light microscope using 100X magnification. For 6-well transfections, the same protocol was used, with the following volumes and amounts (per well): 400 ng plasmid DNA was combined with 3.2 µl enhancer in a total of 100 µl Buffer EC, followed by supplementation with 10 µl Effectene. The entire volume was then added to the cells, which were allowed to grow as above until viewing.

For later experiments involving the Act_{proc}, Act_{trans}, His_{proc}, His_{trans} and snRNA reporters, RNAi-depleted cells were transfected using the Amaxa Nucleofector electroporation system (Lonza) and the Nucleofector V kit (Lonza) according to the manufacturer's protocol for the *Drosophila* S2 cell line. Briefly, $5x10^6$ RNAi-depleted D.Mel-2 cells were isolated by centrifugation at 1000xg for 3 min., followed by aspiration of the media from the cell pellet and addition of 100 µl room-temperature supplemented Nucleofector solution. To this was added 2 µg reporter plasmid DNA and the cells were resuspended gently and transferred to a Nucleofector electrporation cuvette. The cuvette was capped and placed inside the Nucleofector device and the cells were subjected to electroporation using the G-030 Nucleofector program. After electroporation, 500 µl of room temperature growth medium (SF-900II Serum Free Medium supplemented with

antibiotic/antimycotic solution) was added to the cuvette and the cells were transferred with a plastic pipette into 6-well tissue culture plates containing 400 µl room temperature growth medium. The cells were then allowed to grow at 27 °C in a humidified growth chamber until viewing.

The Act_{proc} construct was ultimately integrated stably into *Drosophila* D.Mel-2 cells using the Amaxa transfection protocol above and selection of the cells from Day 3 through Day 10 with 500 µg/ml Zeocin selection reagent (Invitrogen). Note that most of the experiments using this reporter utilized the reporter stable line instead of the transiently transfected reporter. RNAi experiments were performed in this cell line the same way as was done for experiments in which the reporter was transiently transfected.

RT-PCR Analysis

Relative reporter and endogenous mRNA levels were measured using reverse transcriptase PCR (RT-PCR). D.Mel-2 cells were harvested after treatment with dsRNAs and total RNA was isolated using TRIzol reagent (Invitrogen). An aliquot of the isolated RNA (2 µg) was incubated with RNase-free DNase Q (Promega) followed by reverse transcription using MMLV-RT (Invitrogen) and random hexamers, as suggested by the manufacturer. PCR amplification was carried out with 0.5 µl of cDNA using the following forward (F) and reverse (R) primers written in the 5'-3' orientation: Misprocessed reporter (F within H3 ORF-GAGCACCGAGCTTCTAATCC, R within GFP ORF-GGCGGACTTGAAGAAGTCGTGC); Histone H2A misprocessed (F-GTGACAACAAGAAGACTAGAATTATTC, R-CTAATTACAACAAATTGCCAAGC); α1-Tubulin 84B (F-GCCCTACAACTCCATCCTGA, R-

GGTCACCAGAGGGAAGTGAA). PCR products were run on a 2 % agarose gel and detected by ethidium bromide staining.

S1 Nuclease Protection Assay

To create the probe construct for the assay, *Drosophila* histone H2A sequence was cloned into a TA vector. From a standard plasmid preparation (Qiagen), aproximately 3 μg of the plasmid was combined with 2 μl 10X bovine serum albumin (NEB), 2 μl 10X NEBuffer 3 (NEB), qs to 18 μl with dH_2O, and 2 μl BspE I restriction enzyme (NEB), for a total reaction volume of 20 μl. The digest was then incubated at 37 °C for 2 h. Next, the reaction was purified using a PCR purification kit (Qiagen) and eluted the digested DNA in Buffer EB (Qiagen) for a final volume of 28 μl. The result of the initial digest was a linearized construct with a 5'-CCGG-3' overhang at the 5' end of the histone H2A sequence.

To label the DNA, the purified digest was combined with 4 μl 10X dG,A,T deoxynucleotide mix, 4.5 μl 10X NEBuffer 2 (NEB), 4 μl α-P^{32} dCTP (MP Biomedicals), and 3 μl large fragment of DNA polI (Klenow) enzyme (NEB), for a total reaction volume of 43.5 μl. The reaction was incubated at 25 °C for 20 min. This reaction results in the filling-in of the recessed 3' end and 3'-end labeling of the histone H2A sequence with α-P^{32} dCTP.

After the labeling reaction, the reaction was spun through a ProbeQuant G-50 Micro Column (Amersham) to remove unincorporated deoxynucleotides. Next, to the cleaned-up, labeled construct 5 μl 10X BSA (NEB), 1 μl 10X NEBuffer2 (NEB), and 4 μl HindIII restriction enzyme (NEB) was added for a total reaction volume of 60 μl. The reaction was then incubated at 37 °C for 2 h. After incubation, the reaction was run on a 2 % agarose gel and the labeled fragment was purified using a gel extraction kit (Qiagen). These steps result

in a 650 nt, labeled histone H2A probe that is ready to be used in the S1 nuclease protection assay.

For the S1 nuclease protection assay, 5 μg total RNA from Dmel-2 cells was combined with 1 μl of the labeled histone H2A probe and the solution was dried using a Speedvac SC100 (Savant). The dried RNA-probe pellet was resuspended in 10 μl 1X hybridization buffer (40 mM Pipes, pH 6.4, 500 mM NaCl, 1 mM EDTA, 80 % deionized formamide) and boiled for 5 minutes. After boiling, the samples were incubated at 52 °C overnight to allow hybridization of the probe to its target RNA sequence.

After hybridization, 80 μl dH_2O, 9 μl 10X S1 buffer (Promega), and 0.6 μl S1 enzyme (Promega) was added to the samples followed by incubation at 25 °C for 1.5 h. After incubation, the DNA/RNA hybrids were precipitated by adding 300 μl 100 % ethanol and 1 μl GlycoBlue (Ambion) to each sample and placing the samples at – 80 °C for 30 minutes. After precipitation, the samples were centrifuged at 16000xg in an Eppendorf 5415C microcentrifuge for 30 minutes at 4 °C. After removing the supernatant, the pellet was washed with 80 % ethanol and spun the samples again at 16000xg for 10 min. at 4 °C. The ethanol was then removed and the samples were allowed to air dry. After the pellets had dried, they were resuspended in 40 μl formamide loading dye (>99 % formamide with bromophenol blue and xylene cyanole), boiled for 5 min. and loaded onto a 6 % Sequagel Sequencing System acrylamide gel (National Diagnostics) in 1X TBE (Tris-borate-EDTA) buffer.

After running the gel, the gel was dried on a SGD200 Slab Gel Drier (Savant) for 1.5 hours at 80 °C. After drying, the gel was placed either on a storage phosphor screen (Molecular Dynamics) or onto blue autoradiography film (Denville Scientific) for detection.

Phosphor screens were developed on a Storm Scanner (Amersham), and film was developed using a SRX-101A film developer (Konica).

Western Blot Analysis

D.Mel-2 cells were lysed in a buffer containing 50mM Tris-HCl (pH 8.3), 0.1% NP-40, and 50mM NaCl for 30 min. on ice. Lysates were separated with SDS-PAGE using standard techniques and were then transferred to immobilon-PVDF membrane (Biorad) and probed with either a 1:1000 dilution of an α-GFP JL8 monoclonal antibody (Clontech), a 1:1000 dilution of α-dSLBP antibody (*10*), a 1:1000 dilution of α-Lsm11 antibody, or a 1:1000 dilution of α-Lsm10 antibody (*24*). Blots were then probed using a secondary HRP-conjugated antibody and developed using chemiluminescence. Sufficient material could be obtained for Western blotting using cells from 6 wells of a 384 well plate.

Results

Creation, Validation, and Optimization of Histone mRNA Reporters

A reporter to detect misprocessing of histone pre-mRNA. To facilitate the identification of histone pre-mRNA processing factors, a novel screening reporter was created consisting of a histone H3 promoter driving expression of a partial histone open reading frame (ORF), without its stop codon, along with downstream sequence that includes the stem loop, normal cleavage site, and histone downstream element (HDE), all followed by an enhanced green fluorescent protein (EGFP) ORF and a strong, vector-encoded polyadenylation signal (His_{proc}) (Fig. 2-1A). The GFP was placed in frame with the histone ORF and a stop codon present in the intergenic region was mutated. Under normal conditions (Fig.2-1A, top), the reporter message is cleaved at its 3' end like a normal histone mRNA, and this cleavage is efficient based on the fact that the cells do not fluoresce. Because the transcript does not contain the EGFP coding region, the cells exhibit no fluorescence. However, if the normal cleavage reaction is perturbed (Fig. 2-1A, bottom) by depleting an essential processing factor, the reporter transcript is not processed normally, resulting in a reporter transcript that includes the EGFP coding region. When these transcripts are translated, histone H3-EGFP fusion protein is created, which is readily visualized under a fluorescence microscope.

To test the ability of the His_{proc} reporter to assess processing defects, dsRNAs targeting LacZ and pyrimidine tract binding protein (dPTB), as negative controls, and dLsm10, dLsm11, and dSLBP, as positive controls, were synthesized and added to *Drosophila* D.Mel-2 cultured cells. Additionally, a 2'-OCH_3 oligonucleotide antisense to the 5' end of U7 snRNA, which prevents the binding interaction between the U7 snRNP and the

HDE, was used as a positive control. After incubating the cells with their respective dsRNA/modified oligonucleotide for three days, the reporter construct was transfected into the cells and the cells were viewed 48 hours later. RNAi against the negative control targets resulted in only background levels of fluorescence, while RNAi targeting each of the positive controls or introduction of the U7 2'-OCH_3 oligonucleotide elicited much higher levels of fluorescence (Fig. 2-1B). Western blot analysis of these cells using an α-GFP antibody confirms this result (Fig. 2-1C), and allows one to quantify the amount of GFP produced. Further, western blotting reveals a marked decrease in the protein level of each of the processing factors that were targeted for knock down by their respective dsRNAs (Fig. 2-1C). Note that RNAi-mediated depletion of dLsm10 or dLsm11 results in depletion not only of the target, but also of the other Lsm protein (Fig.2-1C, lanes 3 and 4). This is likely attributable to the existence of these proteins together as a dimer in the cell and suggests that the stability of one protein relies on its binding with the other. Addition of the dsRNA against SLBP was quite effective at reducing SLBP levels in the cell, but had no effect on Lsm10 or Lsm11 levels (Fig. 2-1C, lane 5). Further, addition of the 2'-OMe oligonucleotide against the U7 snRNA did not affect levels of any of the other factors (Fig. 2-1C, lane 6), consistent with its function in disrupting the activity of the U7 snRNA by preventing its binding to the HDE rather than affecting its protein partner levels. Thus there is not a general co-regulation of multiple factors for histone pre-mRNA processing.

To determine whether the His_{proc} reporter was accurately reporting misprocessing of the endogenous histone message, the S1 nuclease protection assay was used. These assays reveal misprocessed histone mRNA species as bands that migrate more slowly on a gel than those corresponding to properly cleaved histone mRNA (Lanzotti *et al*., 2002; Godfrey *et al*.,

2006). For this assay, total RNA was isolated from mock-treated cells and polyadenylated mRNA was isolated from cells RNAi-depleted of dLsm11. The results of the S1 nuclease protection assay reveal that, in the absence of RNAi, only a single band appears on the gel (Fig. 2-1D, lane 2), corresponding to the correctly processed histone mRNA species. However, treating cells with a dsRNA that targets a processing factor, such as Lsm11, results in the appearance of multiple bands on the gel (Fig. 2-1D, lane 3), indicating that the histone pre-mRNA is not correctly processed under these circumstances.

In order to use the reporter for carrying out a genome-wide screen, optimization of the assay conditions was required, including the number of cells to plate, the amount of dsRNA to add, the amount of reporter DNA to be transfected, the timing of the transfections, and the appropriate amounts of transfection reagents to use. To this end, these conditions were systematically tested using the 384-well plating scheme shown in Fig. 2-2A. In this experiment, either 1 or 4 µl of dsRNA were added to each well, corresponding to 50 and 200 ng dsRNA, respectively. DNA concentration was assayed at 25, 50, 100, 150, and 200 ng per well. Effectene transfection reagent (Qiagen) was used to introduce the reporter DNA into the cells after the cells were grown in the presence of dsRNAs for 3 days. Using this system, the quantity of two reagents, an enhancer solution and Effectene, must be coordinated together for optimal results. For these experiments, enhancer amounts of 0.1, 0.3, 0.5, 0.8, and 1 µl per well and Effectene amounts of 0.3, 0.4, and 0.5 µl were used. On Day 5, these cells were imaged by fluorescence microscopy. The individual well images were combined into a 384-image collage, representative of the 384 wells of the assay plate, and the images were put through an algorithm using the Metamorph program (MDS Analytical Technologies Inc.) that calculated fluorescence intensity within each well. The

output of this algorithm is the same 384-well collage, but with each fluorescence dot converted into a vertical bar, with the height of the bar scaled to the intensity of the dot (Fig. 2-2B). As expected, the negative control dsRNA, targeting dPTB, gave no appreciable fluorescence under any condition assayed, while the positive control dsRNA, targeting dLsm11, and the U7 2'-Ome oligonucleotide gave higher levels of fluorescence for all treatments (Fig. 2-2B). Note, however, that the fluorescence signal elicited from the positive control cells varies widely and that maximal fluorescence output, still with very low noise, is achieved for only a subset of the tested conditions. In this case, the best signal was obtained for cells supplemented with a higher amount of dsRNA (200 ng), a lower amount of DNA (25-50 ng), lower volume of enhancer (0.1-0.3 µl), and 0.3 µl Effectene reagent (Fig. 2-2B).

A reporter that only targets histone pre-mRNA processing. Though the His_{proc} reporter is highly effective, it is likely that any factors that are required for both histone gene transcription and pre-mRNA processing will be missed as a result of the presence of the histone promoter driving the reporter construct. To alleviate this, I created another histone pre-mRNA processing reporter that utilizes a constitutively expressed Actin 5C promoter driving transcription of the reporter message (Act_{proc}) (Fig. 2-3A, top). RNAi experiments to test the effectiveness of this reporter reveal high levels of fluorescence when obligate processing factors are depleted and very little background under mock treatment (Fig. 2-3A, bottom). As is the case for the His_{proc} reporter, misprocessing of endogenous histone mRNA is accurately reflected by the Act_{proc} reporter. RT-PCR assays performed using cDNA created from cells depleted of SLBP reveal misprocessed reporter message that coincides with misprocessing histone H1, H2A, and H2B message (Fig. 2-3B). In contrast, when PTB is knocked down, there is only background misprocessing of the endogenous histone

message, which is mimicked by the reporter message (Fig. 2-3B). This reporter was ultimately used to create a stable cell line, eliminating variables introduced by having to transfect the reporter during the experimental timecourse. This stable reporter cell line works very well, although a stable cell line using the reporter driven by the histone promoter did not give sufficient fluorescence to be useful.

A reporter that targets histone transcription. In order to assay the effect of RNAi-mediated depletion of factors of interest on transcription from the histone promoter, a third reporter construct was made (His_{trans}). This reporter consists of a histone promoter driving an EGFP ORF (Fig. 2-3C, top) followed by a polyadenylation signal, allowing a direct visual readout of histone gene transcription. Importantly, this reporter is expressed efficiently in the cell (Fig. 2-3C bottom). Using this reporter, factors that reduce histone gene transcription will result in a reduction of the GFP signal, although this loss of fluorescence readout is much less sensitive than the gain of fluorescence of the processing reporters.

Use of Reporters to Help Characterize Other Aspects of Histone mRNA Metabolism

A genome-wide screen for factors that affect the presence in the HLB of a cell cycle regulated, MPM-2 antibody reactive phosphoepitope was performed by Anne White in Dr. Bob Duronio's lab here at UNC. Initial positive hits from this screen revealed a number of factors of interest that localize to the HLB, including multi-sex combs (Mxc), CG34415, FLASH and Spt6. These factors were tested in a secondary assay using the three reporters: His_{proc}, Act_{proc}, and H_{trans} to determine in which steps in histone mRNA metabolism they were involved. As negative controls I used (PTB) or the Cajal body-specific factor coilin and SLBP was used as a positive control for processing.

Figs. 2-4A-C show the results of testing the 3 reporters with the four genes from the screen and the controls. When cells were RNAi-depleted of CG34415 or Spt6, only background fluorescence from the His_{proc} reporter was observed, similar to the result of knocking down the negative controls. However, when Mxc was depleted, there was increased fluorescence, albeit considerably less than that observed for cells depleted of SLBP (Fig. 2-4A).

Using the Act_{proc} reporter to assay the same factors as in Fig. 2-4A, a similar result was obtained. Again, RNAi depletion of CG34415 or Spt6 gave only background levels of fluorescence, similar to coilin depletion, but Mxc now gave considerably higher levels of fluorescence relative to SLBP (Fig. 2-4B). Of note, the relative level of fluorescence off this reporter when Mxc is depleted compared to that arising in control cells is somewhat higher than the relative levels observed for the His_{proc} reporter, consistent with a possible role in transcription of the histone genes. The results for CG34415 and Spt6 in these experiments suggest these factors are not involved in processing histone mRNA, although they do not rule out their possible involvement in histone gene transcription.

To determine whether these factors are required for histone gene transcription, a reporter containing the histone promoter driving expression of an EGFP ORF was used. When this reporter is transiently transfected into cells depleted of Mxc or Spt6, fluorescence lower than that observed for control-treated cells is observed (Fig. Fig 2-4C), suggesting that they might be involved in histone transcription. In cells depleted of CG34415 a more significant reduction in fluorescence is observed (Fig. 2-4C) indicating that this factor is likely important for normal levels of histone transcription. As expected, RNAi depletion of PTB, SLBP, and coilin had no effect on transcription of the reporter (Fig. 2-4C).

These results were representative of only a few of the earliest factors identified in the MPM-2 genome-wide screen. Clearly this approach allows the rapid initial classification of the unknown factors for potential roles in histone mRNA metabolism. After the screen was completed, nearly 100 of the top hits were rescreened, in duplicate, using the Actin 5C promoter-driven histone pre-mRNA processing reporter stable line, to assess whether any of these genes might also be involved in histone pre-mRNA processing. While most of the targets scored negatively for involvement in histone pre-mRNA processing, FLASH scored very strongly and several other factors, including MCRS1, MBD-R2, Mxc, Dgt1, DMAP1, and a few unnamed proteins, including CG8142, CG9772, and CG31111, scored as weakly to moderately positive (Fig. 2-4D). Besides FLASH, whose involvement in histone pre-mRNA processing has been validated and partially characterized, the precise role, if any, of these factors in histone mRNA 3' end formation remains to be determined.

Use of Reporters to Define the Relationship between Histone mRNA Metabolism and DNA Replication

Since histone mRNA levels are tightly coupled with DNA replication, I tested whether knocking down factors required for DNA replication resulted in inhibition of histone pre-mRNA processing. To determine whether depletion of the origin recognition complex (ORC) subunits results in reduced transcription from the histone promoter, I transfected cells treated with dsRNAs targeting Orc1 and Orc3 with the His_{trans} reporter. In parallel, I also used a transcription reporter containing the Actin 5C promoter (Act_{trans}) in order to control against global effects on transcription. Depletion of these DNA replication factors has no effect on transcription from the constitutive promoter (Fig. 2-5A). However, there is an

effect on histone transcription, as reduced fluorescence is detected from the histone promoter when these factors are targeted by RNAi (Fig. 2-5B).

To discover whether ORC subunit depletion affects histone pre-mRNA 3' end formation, I used the Act_{proc} reporter. When the reporter is transfected into cells RNAi-depleted of Orc1 or Orc3, low level fluorescence is detected off the reporter (Fig. 2-5B). This experiment was repeated in the processing reporter stable line, this time using dsRNAs targeting three ORC subunits, Orc1, Orc2, and Orc5. Again, RNAi against the DNA replication factors resulted in low level fluorescence that is above background, but which is well below levels exhibited by the positive control cells (Fig. 2-5C).

Finally, dsRNAs targeting several other DNA replication factors, including several Mcm subunits and Doubleparked (DUP, dCdt1), gave rise to only background levels of fluorescence (Fig 2-5D and 2-5E). Of note, among the DNA replication factors tested, only RNAi-mediated depletion of DUP gave any appreciable effect on cell morphology and survival, giving rise to small cells (Fig. 2-5E, bottom) and lethality by day 2-3 of the experiment. RNAi-mediated depletion of geminin, an inhibitor of Cdt1 function, resulted in cells that became very large (Fig. F2-5E, bottom) and, like the other RNAi targets, did not result in any significant increase in fluorescence off the reporter (Fig. 2-5E, top). I conclude that there is a very small, if any, effect on histone pre-mRNA processing when DNA synthesis is blocked, but that there is likely a reduction in the rate of histone gene transcription.

Pre-snRNA Processing Reporters

The other class of RNA polymerase II transcripts that are not polyadenylated are the small nuclear RNAs. These are also generated by endonucleolytic cleavage of the nascent transcript. In order to investigate the cleavage reaction responsible for creating the mature U7 snRNA 3' end, a reporter was created by Dr. Eric J. Wagner that includes the entire U7 genomic sequence, from the U7 promoter through the end of the intron in which the U7 gene resides, fused at the 3' end to an EGFP ORF and all followed by the strong, vector-encoded polyadenylation signal (Fig. 2-6A). Much like the histone pre-mRNA processing reporters, failure to use the normal cleavage site of the U7 snRNA results in translation of a protein containing the EGFP ORF, causing the cells to fluoresce green. However, unlike the histone reporters, the only start codon occurs at the ATG of EGFP, essentially making the U7 transcript a 5' UTR. This reporter was initially used together with RNAi of specific factors to demonstrate that members of the integrator complex were required for snRNA processing in *Drosophila* and then was subsequently used by Dr. Wagner to perform a genome-wide screen for factors involved in processing U7 snRNA (Wagner *et al.*, in preparation).

When the U7 snRNA processing reporter is transfected into cells depleted of dPTB or dCPSF30, both negative control factors, there is only background fluorescence detectable from the reporter (Fig. 2-6B). However, when an essential processing factor, such as Int9, is depleted, robust GFP signal is observed (Fig. 2-6B). Western blot analysis using an α-GFP antibody confirms this (Fig. 2-6B, right panel).

A rotation student whom I mentored, Kirston Barton, and I used this reporter to probe some of the requirements for production of snRNAs. In particular, in mammalian cells formation of the 3' end of snRNAs requires that transcription initiate from an snRNA

reporter. To test whether this is the case in *Drosophila*, I created reporters with alternative promoters, including the Actin 5C and histone H3 promoters (Fig. 2-6C). I also created a reporter that contained an insertion of histone H3 sequence within the region coding the U7 snRNA transcript in order to test for any transcript length dependence of the processing reaction (Fig. 2-6C). Each of these requirements of snRNA processing has precedents from other mammalian snRNAs, but none have been defined for U7 snRNA in any species or for any *Drosophila* snRNAs.

Substitution of the Actin 5C promoter in place of the natural U7 snRNA promoter results in substantial misprocessing of the U7 snRNA processing reporter as evidenced by robust fluorescence arising in these cells in the absence of any other treatment (Fig. 2-6D, top). Addition of Int9 dsRNA to these cells to knock down this essential processing factor does not give increased fluorescence (Fig. 2-6D, bottom), indicating that the presence of the strong constitutive promoter results in nearly complete failure to process the snRNA. This contrasts markedly with the original U7 snRNA processing reporter containing the U7 snRNA promoter, which shows almost no fluorescence when treated with mock dsRNA, but presents robust fluorescence when dsRNA targeting Int9 is added (Fig. 2-6B and E). Thus processing of the U7 snRNA does not occur when transcription is initiated from the *Drosophila* actin promoter.

In order to test whether the promoter requirement is absolute, a construct was made with the *Drosophila* histone H3 promoter in place of the normal U7 promoter. When treated with mock dsRNA, these cells displayed fluorescence that was greater than that of cells transfected with the U7 promoter-driven reporter (Fig. 2-6E). However, unlike the reporter containing the Actin 5C promoter, addition of Int9 dsRNA resulted in a dramatic increase in

fluorescence (Fig. 2-6E). This suggests that the histone promoter allows some processing of the U7 snRNA, although processing is inefficient. Whether this is because the histone promoter normally transcribes a non-polyadenylated mRNA is not known. In mammalian cells and frog oocytes, snRNAs transcribed from histone promoters are not processed. In sea urchins the snRNA transcripts transcribed from the histone promoter can be inefficiently processed, although other promoters were not tested.

Another property that was tested was whether the length of the U7 snRNA transcript affects processing efficiency. Experiments with mammalian U1 snRNA suggested that processing of longer transcripts ending in snRNA 3' ends was less efficient. To test this, reporters were created in which pieces of the histone H3 gene of varying length were inserted in the middle of the U7 snRNA transcript. Note that these constructs were made such that they contained no cryptic start codons to ensure that the EGFP ORF remained translatable in frame. Introduction of an insert of 310 nt into the middle of the U7 transcript resulted in fluorescence off the reporter under mock RNAi conditions at levels similar to that of the histone H3 promoter-driven U7 snRNA reporter (Fig. 2-6E). Also like the histone H3 promoter-driven reporter, addition of Int9 to the reporter containing the extended U7 region elicited a robust fluorescence (Fig. 2-6E). Thus this 380 nt transcript (containing the U7 snRNA plus the histone fragment) was not processed efficiently, suggesting that there is a length dependence for maximally efficient processing of U7 snRNAs, and that longer transcripts were not processed efficiently. Insertion of smaller fragments of the histone gene did not result in inefficient processing, consistent with the fact that these transcripts were of the size range (< 250nts) of normal snRNAs.

Discussion

The use of RNA interference coupled with visual reporter-based screening strategies has been a great addition to our molecular toolbox (Wagner *et al.*, 2004; Levinson *et al.*, 2006) for studying transcription and alternative splicing. In this chapter, I report the creation and implementation of a number of novel reporters that have been quite effective in broadening our understanding of 3' end formation of RNA polymerase II transcripts, histone mRNAs and snRNAs. These reporters can be designed to identify cis elements involved in different steps of metabolism of an RNA. They can also help determine at what step a factors participates in metabolism of an RNA. Finally, they also can be used to identify novel factors involved in metabolism of the RNA *in vivo,* for example by using these reporters in genome-wide screens utilizing RNAi. Eric Wagner and I carried out a genome-wide RNAi screen using the histone processing reporter (described in Chapter III) and Eric has subsequently carried out a successful genome-wide RNAi screen with the snRNA reporter.

The histone promoter-driven histone pre-mRNA processing reporter allowed us to identify putative histone pre-mRNA processing factors on a genome-wide scale (Wagner *et al.*, 2007). However, this reporter would not detect any factors that are required for both transcription and processing. The creation of the constitutive Actin 5C promoter-driven histone pre-mRNA processing reporter allows the detection of any factors involved in processing whether or not they are involved in other steps in histone mRNA biosynthesis. An unexpected advantage of this reporter is that it is active enough to be used when integrated into the genome. The histone H3 promoter used in the His_{proc} reporter was not

active enough to use after integration into a chromosome. Thus the assay has evolved into a more sensitive and reproducible method that allows experimental time to be cut in half, since there is no transfection of the reporter. This reporter was very useful in assessing the functional role of dFLASH in histone pre-mRNA processing (Yang *et al.*, 2009), and is very easy to use to assay large numbers of factors, as it has in the secondary screening strategy for factors identified in the MPM-2 screen (White *et al.*, in preparation). Finally the histone promoter-driven EGFP construct allowed me to assess effects of factors on histone gene transcription. Together these three reporters allow me to distinguish in which of the multiple steps in histone mRNA metabolism a factor participates.

Design and Optimization of Histone mRNA Reporters

At the beginning of this work, several factors were known to be involved in histone pre-mRNA processing. These factors include the stem loop binding protein (SLBP), the U7 snRNP components Lsm10, Lsm11, and in mammals, ZFP100 and at least two proteins also involved in processing polyadenylated mRNAs, CPSF73 and Symplekin. However, it seemed likely that there were other factors, potentially many more, which remained to be identified.

In order to help in the identification of additional processing factors, a novel screening reporter was created. The design of this reporter relied on a property of *Drosophila* histone genes noted previously, that downstream of the normal cleavage site of each of the histone genes there exist cryptic polyadenylation sites that are used when normal cleavage is inhibited (Lanzotti *et al.*, 2002). The initial design concept for the reporter included the histone ORF and downstream sequence containing the stem loop and histone

downstream element followed by an in-frame fluorescent protein ORF. The idea behind this was to create an assay in which, under normal cellular conditions, the reporter will be transcribed and the 3' processing elements will be bound by their respective processing factors, resulting in cleavage of the reporter mRNA in a way that mimics a normal histone mRNA. In this case, the final reporter transcript will not contain the fluorescent protein coding region. Thus, when the reporter message is translated, there will be no fluorescent protein made and the cells will exhibit no fluorescence. However, if the histone pre-mRNA cleavage reaction is perturbed by depleting an essential processing factor, the reporter transcript will not be processed normally either. Instead, the mature reporter mRNA will include the fluorescent protein coding region and this misprocessed message will become polyadenylated through use of a vector-encoded polyadenylation signal. When these reporter messages are translated, a histone H3-fluorescent protein fusion will be created, which will cause the cells to appear green under a fluorescence microscope.

Once the initial reporter design was decided upon, much work remained to be done. Included in this were experiments to choose a promoter to drive expression of the reporter, modification of downstream sequences, and a determination of which fluorescent protein to include downstream. In the end, these experiments led to the creation of a construct that consisted of a histone H3 promoter driving a partial histone H3 ORF, without the stop codon, and downstream sequences modified to eliminate a cryptic stop codon in the HDE, followed by an enhanced green fluorescent protein (EGFP) coding region.

When this construct was tested using RNAi-mediated depletion of known processing factors, cells produced robust green fluorescence. Just as importantly, when factors unrelated to the histone pre-mRNA cleavage reaction were knocked down, very little signal was

observed. These observations suggested that this reporter could be an effective tool for discovering proteins required for processing histone message.

Though this reporter has been used with great success, most notably in the genome-wide screen presented in Chapter III, one concern about its design was whether the presence of the histone promoter driving its expression was essential. There is evidence that, at least for some non-histone messages, the maturation of their mRNA 3' ends is in some way coupled to transcription (Mifflin and Kellems, 1991; Cramer *et al.*, 2001; Bentley, 2002; Proudfoot *et al.*, 2002). If this is the case with histone messages, then it seems likely that a factor that is required for both transcription and processing of the histone message would not score positively using the histone promoter-driven reporter because this promoter would be silenced. To address this possibility, I designed several reporter constructs that incorporated a constitutive promoter in place of the histone promoter. Three different promoters were tested, including *Drosophila* Actin 5C and α-tubulin at 48B promoters and the lepidopteron OpIE2 promoter. In the end, the Actin 5C promoter was chosen, as it gave the best signal to noise ratio overall, although none of these reporters resulted in GFP synthesis in normal S2 cells. In an effort to make these reporter experiments as reproducible as possible, this construct was ultimately transfected into *Drosophila* D.Mel-2 cells, which were selected to create a stable cell line containing the reporter construct integrated into its genome. The final result was a reporter that is not sensitive to depletion of histone-specific transcription factors, gives more robust fluorescence signal than the original reporter, and allows for more reproducible results in nearly half the time since the transfection step has been eliminated. The latter is a key advantage for screening factors that are possible processing factors but that

may also result in lethality before Day 5, which was the standard experimental length for the histone-promoter-driven reporter.

Since its creation, the constitutive histone pre-mRNA reporter has been used to test a number of factors in targeted screens to determine whether they are required for processing histone pre-mRNA. One highly successful use of this reporter is the initial characterization of dFLASH as a histone pre-mRNA processing factor (Yang *et al.*, 2009), which is described in Chapter IV. Another example includes the use of this reporter in a secondary screen of positive hits from a genome-wide screen for factors required for the existence of a specific cell cycle regulated phosphoepitope at the HLB (White *et al.*, in preparation), as described in this chapter. In each of these cases, novel factors were identified as putative processing factors that were not revealed during the genome-wide screen.

In addition to detecting misprocessed histone message, I desired the ability to test factors for a role in transcription off the histone promoter. To this end, an additional reporter, with the histone H3 promoter driving expression of GFP, was created that can be used to measure downregulation of histone transcription upon depletion of a factor of interest. This reporter, while effectively expressed in the cell, does not offer as simple a readout as the processing reporters. The reason for this is that, unlike the processing reporters, which rely on a gain-in-fluorescence phenotype that is easy to detect over background, this reporter utilizes a loss-of-fluorescence readout that can be difficult to detect by eye. Though not employed in the experiments presented here, the use of quantification software or Western blotting for to quantify GFP levels should be a great help in this regard.

Application of the reporters: characterization of factors in the MPM-2 Screen

The constitutive histone pre-mRNA processing reporter was has been employed extensively since it was created in an effort to identify putative processing factors that were missed in the original genome-wide screen. In this role, it was used as part of a secondary screen of proteins identified as positive hits in a genome-wide antibody-based screen for factors that affect the existence of an MPM-2 reactive epitope at the HLB in *Drosophila* cultured cells (White *et al.*, in preparation). The MPM-2 antibody binds to a poorly defined phosphoepitope that in fly cells is cell cycle regulated and concentrates at the histone locus body. The purpose of the screen was to identify the protein containing this MPM-2 reactive epitope, and further, to identify other gene products that affect its appearance at the HLB. Once the initial screen was complete, around 100 of the top hits were chosen for secondary screening. Included in these secondary screens were additional antibody-based experiments to determine whether localization of other HLB factors is affected by depletion of the same proteins that cause the loss of the MPM-2 reactive epitope from the HLB.

Initial hits from this screen, including Mxc, CG34415, and FLASH, as well as Spt6, which was identified by mass spectrometry of proteins co-immunoprecipitated with the MPM-2 antibody, were tested for any role in histone pre-mRNA processing and histone transcription. Using the three reporters, I was able to determine that each of these factors likely participates in different ways in histone mRNA metabolism: FLASH only affected processing, Mxc affected processing and transcription, and CG34415 and Spt6 only affected transcription. Both Mxc and Spt6 partially reduced transcription, but CG34415 was essential for histone gene transcription.

At the conclusion of the MPM-2 antibody primary screen, the top 95 out of 140 hits were tested. Interestingly, three factors that have been independently implicated in histone pre-mRNA processing were identified in the primary screen. These factors were MBD-R2 and MCRS1, which were positive hits in the original histone pre-mRNA processing screen (Wagner *et al.*, 2007; Chapter III of this dissertation), and FLASH (Yang *et al.*, 2009; Chapter IV of this dissertation). Not surprisingly, these factors again scored positively in the secondary histone pre-mRNA processing screen. Though FLASH was the only strong hit of the group, MBD-R2, MCRS1, and Mxc qualify as moderate hits, making them interesting targets for further study. Though implicated as factors of interest, very little is currently known about any of the factors identified in this secondary screen beyond dFLASH (Yang *et al.*, 2009; Chapter IV of this dissertation).

Among the moderate hits from the secondary screen, MBD-R2, also known as TAM3, is represented by two protein species, one of 1169 amino acids and the other of 1081 amino acids. The sequences of these proteins are identical, with the exception of a THAP domain, which is a putative DNA binding domain, at the very N-terminus of the long form that is not found in the shorter protein. Both protein species contain a methyl-CpG binding domain (MBD), for which the protein is named, that may function as a DNA binding motif. In addition, the protein contains putative PHD-finger and Tudor domains, which represent protein interaction domains. Beyond the identification of these domains, very little is known about this protein in flies.

Drosophila MCRS1, alternatively known as Reduction in Cnn dots 5 (Rcd5), is 578 amino acids long and is the homolog to human microspherule protein 1 (MCRS1; also known as, MCRS2, p78, MSP58). In both species, this protein contains a forkhead-associated

(FHA) domain, which binds phosphopeptides. In humans, this protein is represented by several isoforms. At least one of these localizes, at least in part, to the nucleolus, where it participates in upregulating transcription of ribosomal genes (Shimono *et al.*, 2005). Many other roles have been suggested for the isoforms of this protein, including enhancement of transcriptional repression (Du *et al.*, 2006), participation in cell cycle regulation (Song *et al.*, 2004; Zhang *et al.*, 2007), translational regulation (Davidovic *et al.*, 2006), centrosome activity (Hirohashi *et al.*, 2006), and telomere maintenance (Song *et al.*, 2004). It remains to be seen exactly how this protein may function in histone mRNA expression, but it is clear that depletion of this factor results in direct effects on the HLB and the processes that occur there.

Mxc (multi sex combs) was identified as a moderate hit in my reporter assays. This protein is 1837 amino acids long and contains a LisH domain in its N-terminus. This domain is responsible for mediating protein-protein interactions, including dimerization and tetramerization. Importantly, this protein localizes to the HLB and, based upon the reporter experiments in this chapter, appears to have a role in both histone gene transcription and histone mRNA 3' end formation. Because the amount of readthrough of the histone pre-mRNA processing reporter is only moderate, it may be that this factor is more indirectly involved in the processing reaction, perhaps through some structural role at the HLB. Indeed, it has been observed that depletion of this protein in fly cells results in delocalization of other HLB-localizing factors, including dFLASH (White *et al.*, in preparation), which is reminiscent of the result of knocking down NPAT in human cells. Based upon this and other collected evidence, as well as the presence of a LisH domain that is 51 % identical between the two proteins, it almost certain that this protein is the *Drosophila* NPAT homolog.

Are DNA Replication and Histone mRNA Biosynthesis coupled?

Depletion of DNA replication factors in *Drosophila* cultured cells reduces transcription of histone mRNA, as assayed by the histone promoter-driven transcriptional reporter. This agrees with results from mammalian cells that show a 3-4 fold reduction in the rate of histone gene transcription, as well as rapid degradation of histone mRNA when DNA replication is inhibited. Combined with destabilization of histone message, as has been observed in mammals, transcriptional silencing is an effective strategy to prevent the toxic accumulation of free histone proteins under these conditions. The results of experiments using the constitutive 3' end processing reporter suggest that, while there is some amount of misprocessed histone message under conditions of DNA replication factor depletion, there is not nearly the effect as one would see for RNAi versus a dedicated processing factor. This seems to rule out the possibility of perturbations in DNA replication directly targeting a histone pre-mRNA processing factor. Instead, it seems to point to a scenario in which any effect on histone pre-mRNA processing is secondary and likely arises as a consequence of a more global response to DNA replication inhibition.

Of note from these experiments, treating cells with dsRNAs targeting the ORC subunits did not result in any appreciable effect on cell growth or viability, not matter which subunit was targeted. Indeed, use of dsRNAs targeting two independent regions within each of the ORC subunits resulted in nothing more aberrant than a slight slowing of cell proliferation, if any defect was observed at all. This may suggest insufficient knockdown of the ORC subunits or redundant functions within the cell. The result of knocking down Dup is more in line with what one might expect; that is, the cells show a distinct morphological phenotype, in this case very small cells, and then die within a few days. As might be

expected, depletion of geminin, which actively inhibits the action of Dup during the cell cycle to prevent rereplication, results in cells that are very large, presumably due to their existence in a perpetually active S phase during which the cell continues to grow and the genome is replicated again and again.

Using Reporters to Help Characterize U7 snRNA 3' End Formation

The snRNA reporters described in this chapter have been a great help to our understanding of the U7 snRNA processing reaction. The original screening reporter was used by Dr. Eric Wagner for his genome-wide screen for U7 snRNA processing factors (Wagner *et al.*, in preparation), and subsequent reporters made by me have given us greater insight into how the structure of the U7 snRNA gene affects the processing reaction. Together, all of these reporters have contributed to a more complete understanding of critical mechanisms within the cell.

In addition to their usefulness in detecting misprocessing of histone mRNA 3' ends, reporters can be used to identify factors involved in the 3' end processing reactions of other RNAs, including small nuclear RNAs (snRNAs). Such an approach to gene discovery again utilizes the basic elements of the original screening reporter, that is, a promoter to drive the reporter transcript and a way to detect readthrough of the normal cut site, such as inclusion of a fluorescent protein ORF downstream. In addition to identifying factors involved in processing these RNAs, other aspects of their metabolism can be tested using similar reporter strategies. For example the cis elements required for snRNA 3' end formation have not been thoroughly described in *Drosophila*, and introducing mutations into the reporter allows detection of sequences required for snRNA 3' end formation. In this way, properties such as

promoter, sequence element, and transcript length requirements can be identified and characterized in a way that may not be as easily detectable by other means.

There have been many studies that have addressed the link between transcription and 3' end formation and some of these point to the promoter as a crucial determinant of processing efficiency. It is known in vertebrates that recognition of an essential processing element at the 3' end of the snRNA transcript, known as the 3' box, only occurs when the snRNA is transcribed off of an snRNA promoter (Neuman de Vegvar *et al.*, 1986; Hernandez and Weiner, 1986). Indeed, use of a histone promoter to drive an snRNA transcript results in transcriptional readthrough and failure to form mature snRNAs (Pilch and Marzluff, 1991; Ramamurthy *et al.*, 1996). Only recently was the mechanism of 3' end formation of snRNAs elucidated with the discovery of the Integrator complex, a multi-subunit complex containing homologues of CPSF73 and CPSF100, that likely cleaves the pre-snRNA cotranscriptionally. While there is a conserved downstream element in snRNA genes, an AU rich 3' box, there is not a conserved upstream element and what sequences are required for snRNA 3' end formation are not known. Indeed substitution of a histone stem loop (or mutant stem loop with the stem reversed), allows efficient 3' end formation (Ramamurthy *et al.*, 1996).

In addition, phosphorylation of the CTD of RNA polymerase II on Ser7, and possibly Ser2, has been implicated in efficient snRNA 3' end formation (Medlin *et al.*, 2003; Medlin *et al.*, 2005; Jacobs *et al.*, 2004; Egloff *et al.*, 2007). This requirement suggests that the Integrator complex, which contains the snRNA cleavage factor (Baillat *et al.*, 2005), is recruited to the RNA pol II CTD only when the Ser7 residue is phosphorylated (Egloff *et al.*, 2007). It has been postulated that this interaction may involve the concerted action of an snRNA promoter-specific protein, proximal sequence element-binding transcription factor

(PTF), and the CTD of pol II to actively recruit processing factors, including the Integrator complex, to the snRNA 3' end (Egloff *et al.*, 2008).

Despite early work that determined promoter and sequence elements required for correct processing of spliceosomal snRNAs (Hernandez and Weiner, 1986; Neuman de Vegvar *et al.*, 1986; Ach and Weiner, 1987; Wendelburg and Marzluff, 1992a,b) and the identification of the Integrator complex as a mediator of snRNA 3' end formation for the RNA polymerase II-transcribed spliceosomal snRNAs U1 and U2 (Baillat *et al.*, 2005), very little is known about U7 snRNA 3' end formation. Since U7 snRNA is transcribed by RNA pol II and the Integrator complex directly interacts with the RNA pol II large subunit carboxy-terminal domain (CTD), one may reason that this complex is also required for processing U7 snRNA. However, experimental evidence that this is indeed the case is just now coming to light with the results of a very recent genome-wide RNAi screen carried out by Dr. Eric Wagner (Wagner *et al.*, in preparation). From this work, it has been determined that only a subset of the Integrator subunits are functionally required for U7 snRNA processing, which is reminiscent of the findings from the histone pre-mRNA processing screen in which only select cleavage and polyadenylation factors are required for processing histone mRNA. Interestingly, two of the Integrator factors are homologs of two of the core CPSF factors, with Int11 and Int9 sharing homology with CPSF73 and CPSF100, respectively (Dominski *et al.*, 2005). In addition to these key results, further work from this study has gone on to elucidate processing signals within the region downstream of the cut site that are required for proper processing of U7 snRNA. In collaboration, I devised a number of reporter constructs to help determine any promoter and length requirements for U7 snRNA processing.

By creating reporters in which the U7 promoter was substituted with other promoters of varying strengths and characteristics, I hoped to gain better insight into any promoter dependence of the U7 snRNA processing reaction. These experiments revealed that inclusion of the Actin 5C promoter, which is strong and constitutively active, resulted in very high levels of reporter readthrough. Indeed, even RNAi-mediated depletion of the Int9 subunit of the Integrator complex, which results in high levels of misprocessing, did not enhance the fluorescence signal from the reporter suggesting that there was essentially no snRNA 3' end formation when the actin promoter was used. However, when the histone H3 promoter was substituted, there was detectable fluorescence, but it was not nearly the amount that was seen for the actin promoter, and knockdown of int9 resulted in a large increase of fluorescence, consistent with the interpretation that the histone promoter could direct snRNA 3' end formation, although not efficiently. Interestingly, this is similar to results obtained for the U1 snRNA in sea urchin. In these experiments, Wendleburg and Marzluff found that, in contrast to vertebrate snRNAs, 3' end formation of the sea urchin U1 snRNA does not require the snRNA promoter and further, that substitution of a histone promoter still results in correct processing of U1 snRNA with about 50% efficiency (Wendelburg and Marzluff, 1992b). Together with my results from fly cells, this supports the idea that snRNA processing requirements vary between vertebrates and invertebrates, and that invertebrate histone promoters are compatible with snRNA 3' end formation at a reduced efficiency compared to snRNA promoters.

A dependence of vertebrate snRNA 3' end formation on the length of the transcript was previously reported (Ramamurthy *et al.*, 1996). In addition to testing the promoter requirement of U7 snRNA 3' end formation, I also tested whether there is a transcript length

requirement in *Drosophila*. To do this, I used a construct much like the original U7 snRNA screening reporter, but with a portion of the histone H3 gene inserted into the middle of the region encoding the U7 transcript. These constructs were designed such that no cryptic start codons were included in the insert sequence. When an insert of 310 nt was included in the U7 transcript, making the transcript length about 370 nts, misprocessing of the reporter was evident to a level similar to that of the histone promoter-driven construct, with the level of fluorescence well below that arising when Int9 is depleted from the cells.

These results illustrate that maximal U7 snRNA processing efficiency is dependent upon a U7 snRNA transcript that this sufficiently short, which agrees with data from a study of U1 snRNAs in which the efficiency of processing was dependent upon the length of the insert (Ramamurthy *et al.*, 1996). In this study, the researchers found that transcripts less than 350 nts resulted in processing efficiency above 90 %, while efficiency dropped to 50-70 % for transcripts of 450 nt and to 10-20 % when the insert length was extended above 650 nt (Ramamurthy *et al.*, 1996). The reason behind this may lie in the observation that the transcriptional machinery can pause or terminate with relative ease earlier in transcription, but sometime during elongation this attribute changes, making the polymerase resistant to termination (Kephart *et al.*, 1992; Marshall and Price, 1992). In this way, a sufficiently close 3' box may permit the transcriptional machinery to pause or terminate, allowing the processing factors associated with the polymerase to cleave the pre-snRNA. However, if the distance between the promoter and the 3' box is sufficiently large, the polymerase may enter a more pause-resistant state, reducing the amount of time the processing machinery has to cleave the snRNA and effectively decreasing processing efficiency.

The Possible Use of Reporters to Detect Inhibitory Factors

The use of reporters to detect factors that are required for a cellular function, such as histone pre-mRNA processing, is a highly effective method of gene discovery. However, determining the identity of these factors may only tell half the story. The mechanisms that occur within the cell are very complex and often changing, being both upregulated and downregulated at a given moment depending upon the needs of the cell and the dictates of the cell's internal and external environments. Therefore, just as important as those factors that allow a process to go forward are those factors that actively work to inhibit it. Expression of a factor that inhibits pre-mRNA or snRNA processing will also result in synthesis of GFP.

As shown in Chapter IV of this dissertation, introduction of a recombinant FLASH protein missing part of its N-terminus into a stable cell line containing a GFP-based, constitutively expressed histone pre-mRNA processing reporter results in misprocessing of the reporter and green fluorescence. In this case, this protein functions as a dominant negative, actively inhibiting processing of histone pre-mRNA 3' ends. A natural inhibitor of histone pre-mRNA processing would act in much the same way. Introduction of this protein in sufficient quantities into a cell harboring the reporter should prevent proper cleavage of the reporter message, and presumably the endogenous histone message, resulting in detectable green fluorescence. Thus, a candidate factor can be overexpressed in the reporter cell line and assessed for inhibitory activity based upon this readout. This should not be too difficult on a very small scale with a limited number of putative inhibitory factors to test. However, if one wishes to screen the genome for these factors, the task becomes a bit more daunting.

The challenge, then, is to find a way to individually overexpress every protein in the cell in a way that is conducive to high throughput screening.

One way to address this is to develop an overexpression screening library, which contains all of the protein-encoding sequences in the genome subcloned into an expression vector and driven by a strong promoter. These clones could be arrayed into plates, together with a transfection reagent, and cells added to initiate the transfection, in much the same was as dsRNA is added to the wells prior to performing an RNAi-based screen. Cells stably expressing the GFP-based reporter construct can then be plated into these wells and allowed to take in the DNA. After an appropriate incubation period, the cells can be viewed using fluorescence microscopy to determine any increase in GFP fluorescence over the negative control. The result of these experiments would be a list of possible factors that act as inhibitors of RNA 3' end formation. Once these factors have been identified, it may be possible, depending upon your system, to validate the hits biochemically. In the case of histone pre-mRNA processing, this would include *in vitro* processing assays. Further experiments, including crosslinking and immunoprecipitation assays, among others, may help define the mechanism of inhibition.

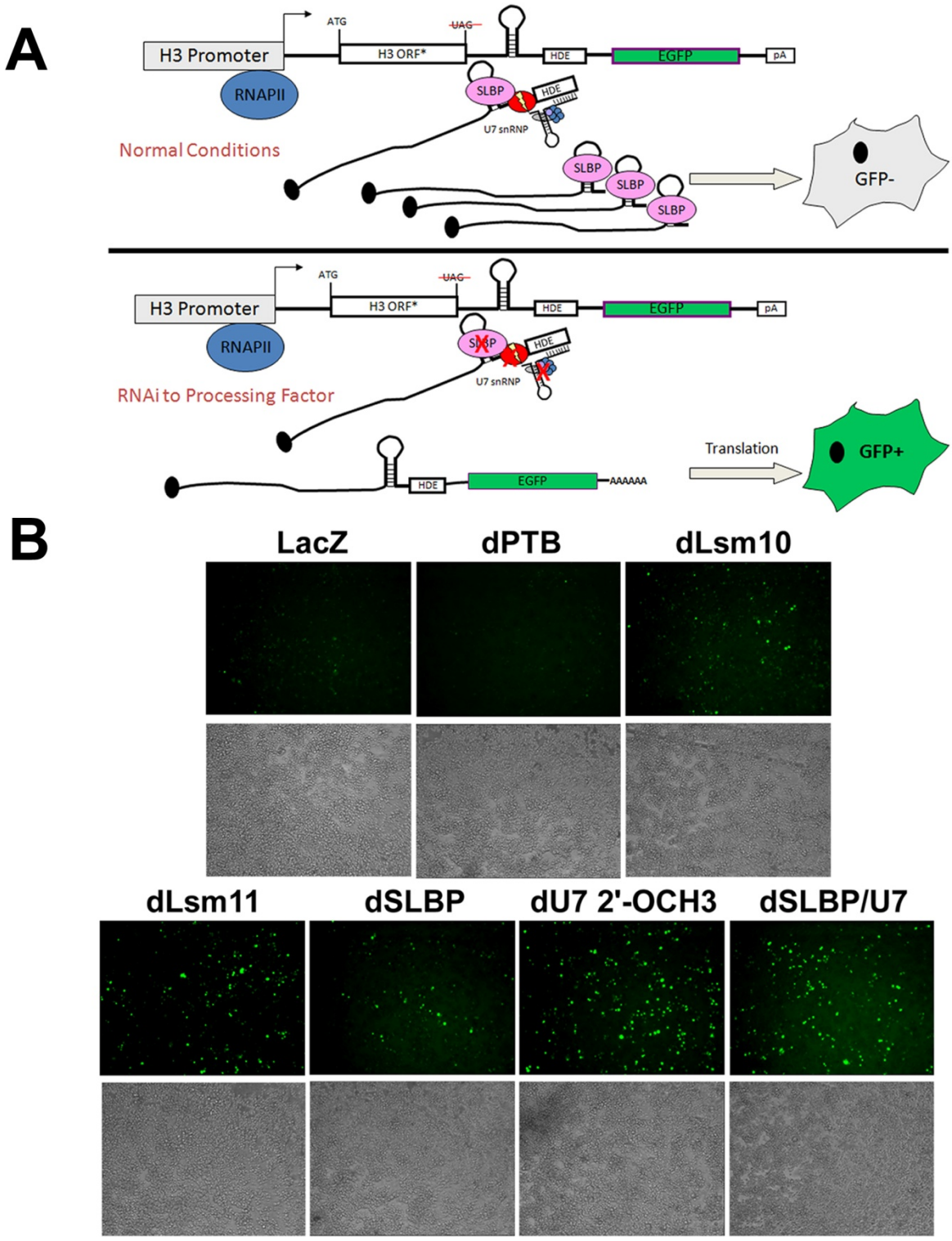
Though the reporters presented in this chapter have been very useful coupled with RNA interference to deplete factors of interest, this is not the limit of their applicability. These reporters can also be used to screen chemical compounds to find pharmaceutical inhibitors of transcription or RNA 3' end formation. The availability of small molecule libraries is increasing, examples of which exist locally at UNC Chapel Hill and the Biomanufacturing Research Institute and Technology Enterprise (BRITE) program at North Carolina Central University. These libraries allow high throughput screening of thousands of

chemical compounds, making it possible to discover compounds that one day might find therapeutic use in human health and/or provide valuable reagents that can be used by researchers to more effectively study key processes in the cell.

Table 2-1. Oligonucleotides used to create DNA template for transcription of dsRNAs. Shown is a list of oligonucleotides that were used to prime amplification of DNA to make template for transcription. The final amplicon contained T7 promoter sequences on each end, allowing for transcription off of both strands. "Target" indicates the gene being amplified, "Annotation ID" is the CG number for that gene, "Forward" and "Reverse" indicate primers that prime in these orientations with respect to the gene being amplified. Note that the T7 promoter sequence is abbreviated as "T7" for brevity. The T7 promoter sequence is GGTAATACGACTCACTATAGGG.

Target	Annotation ID	Forward	Reverse
PTB	CG31000	T7-GCCCATAGCGACTACAGC	T7-TGGAATGAATTGTTCTTTGTGAA
LSM10	CG12938	T7-ATGCAGCAGTTTAGTGCG	T7-TTAGGTTTCCTTATTCTT
LSM11	CG12924	T7-ATGGAATCGAGGGACCGGAAAAC	T7-CAACAGTTCACCCTCGACACTGCC
SLBP	CG11886	T7-TCCAGTTCCTTGAATAGCAG	T7-AGTCCGCTCGTCCTTTG
FLASH	CG4616	T7-CGAAAGTAAGCGTCCGAAAG	T7-ATTCCTGTGATGATCTCGCC
MXC	CG12124	T7-CCCATGAATCTGGCAAAGAA	T7-AACTGGGGCGTATCCATAAT
MUTE	CG34415	T7-TTCCCCTCAAGTCGACAAAC	T7-TTCCGATCTATCTTCGGTGG
SPT6	CG12225	T7-TGAAAATGAACGGGATGTCA	T7-GCGTGTATGGCGTATTGATG
COILIN	CG8710	T7-ATGTTCCGCGCATATTTCA	T7-AACATCCTTCGAATCAACGT
ORC1	CG10667	T7-GCCCAGGCGCAGTATTC	T7-CCAGTGCCCGGAACTCC
ORC2	CG3041	T7-AGCGATGCTGGCAACTC	T7-TATCCAGCATATCCTTGATGG
ORC3	CG4088	T7-AAAGAGGTAGTGCAGCAGCC	T7-AAAACCAATGTCTCAACGGC
ORC5	CG7833	T7-GAACAGTTCGCCCAGGATAA	T7-GCATAGTAGGGCAGCTCCAG
MCM3	CG4206	T7-AGCAAGGAGAGCAATTTGGA	T7-GGTCTCCTGATCCGTGGTAA
MCM5	CG4082	T7-TATCGCTGTCGACGTTCAAG	T7-TGAGTGTTTCCTGGTCCTCC
DUP	CG8171	T7-ATGTCGCCGCGTCACA	T7-GGATTCATGGACGTCGACT
GEMININ	CG3183	T7-ATCACGGCAGAGGATCTCAC	T7-TTAGACCAGCCGTTGTGTTG
INT9	CG5222	T7-GGTCTTTTGTGGCCATCCTA	T7-TAAATTCGATCCAGCTTCCG

Figure 2-1. Creation and validation of a histone pre-mRNA processing reporter. (A) Diagram of the design and action of the His_{proc} reporter under normal conditions (top) and when an essential processing factor has been depleted (bottom). (B) *Drosophila* D.Mel-2 cells were treated with dsRNAs targeting the indicated factors for 3 days before transfecting with the His_{proc} reporter. Two days later, cells were viewed under a fluorescence microscope to detect GFP signal. dsRNAs targeting LacZ and dPTB are used as negative controls, dsRNAs targeting dLsm10, dLsm11, and dSLBP are used as positive controls. A 2'-OCH_3 oligonucleotide targeting the U7 snRNA was also used as a positive control. (C) Western blot analysis of the cells from Panel B. Antibodies targeting dLsm11, dLsm10, and dSLBP were used to indicate knockdown of these factors. An antibody to GFP was used to detect translated misprocessed reporter. (D) S1 nuclease protection assay probing histone H2A and using total RNA from *Drosophila* S2 cells and oligo(dT)-selected poly(A) mRNA from S2 cells treated with dsRNA to dLsm11. A diagram of the S1 nuclease assay is shown below the figure. Abbreviations: His_{proc} = histone H3 promoter-driven histone pre-mRNA processing reporter; RNAPII = DNA dependent RNA polymerase II; H3 ORF = histone H3 open reading frame; SLBP = stem loop binding protein; HDE = histone downstream element; EGFP = green fluorescent protein; pA = polyadenylation signal; LacZ = β-galactosidase; dPTB = pyrimidine tract binding protein; dLsm = Sm-like protein; dU7 2'-OCH_3, U7, and α-U72'Ome = 2'-O-methyl oligonucleotide; α-dLsm11, α-dLsm10, α-SLBP, and α-GFP = antibodies targeting these factors.


A
ATG
UAG
H3 Promoter
H3 ORF*
HDE
EGFP
pA
RNAPII
SLBP
HDE
U7 snRNP
Normal Conditions
SLBP
SLBP
SLBP
GFP-
ATG
UAG
H3 Promoter
H3 ORF*
HDE
EGFP
pA
RNAPII
SLBP
HDE
U7 snRNP
RNAi to Processing Factor
HDE
EGFP
AAAAAA
Translation
GFP+
B
LacZ
dPTB
dLsm10
dLsm11
dSLBP
dU7 2'-OCH3
dSLBP/U7

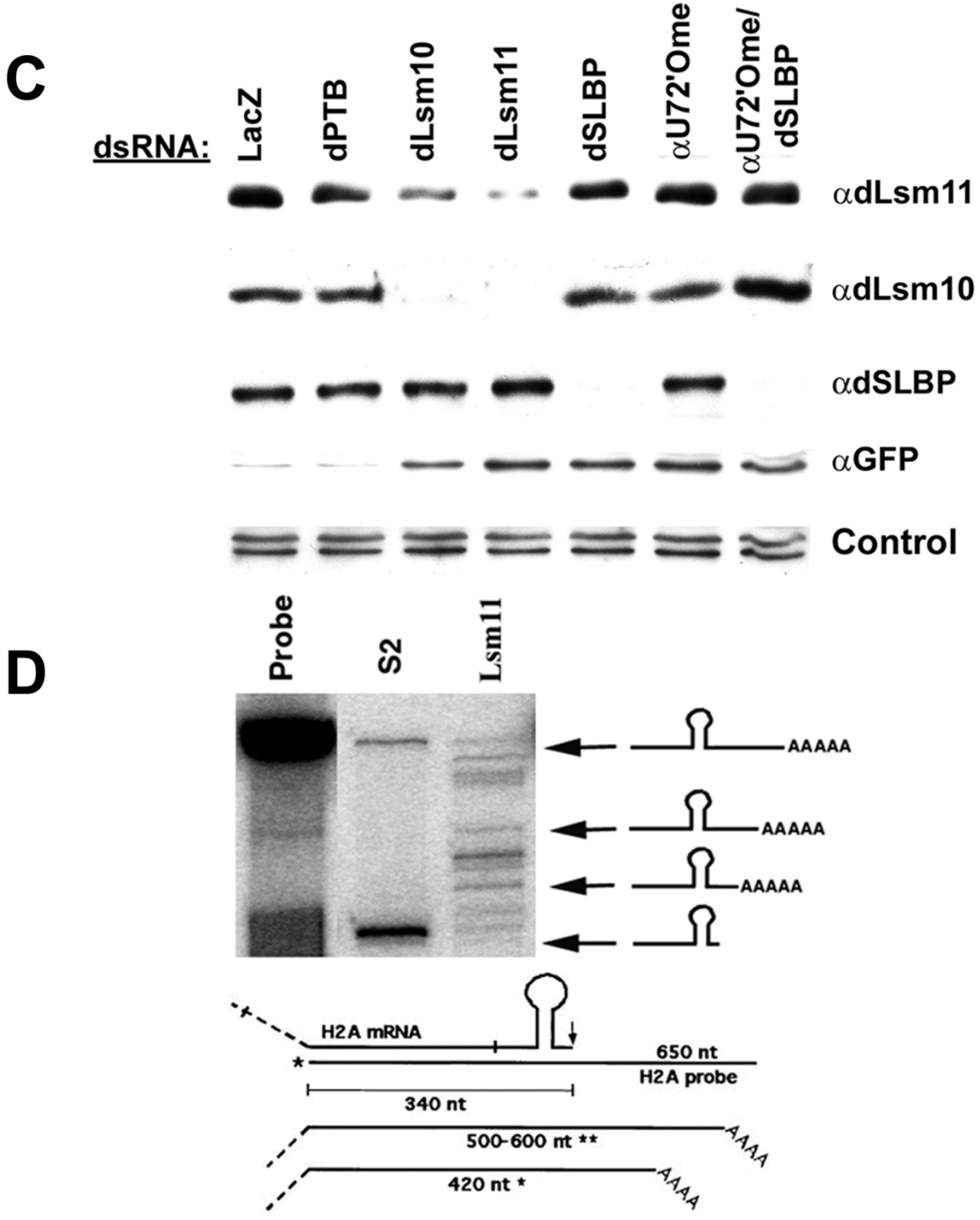
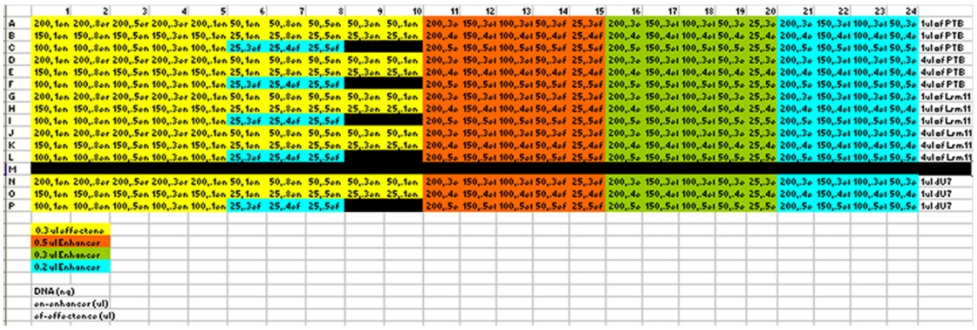

C
dsRNA:
LacZ
dPTB
dLsm10
dLsm11
dSLBP
αU72'Ome
αU72'Ome/ dSLBP
αdLsm11
αdLsm10
αdSLBP
αGFP
Control
D
Probe
S2
Lsm11
AAAAA
AAAAA
AAAAA
H2A mRNA
650 nt
H2A probe
340 nt
500-600 nt **
AAAA
420 nt *
AAAA

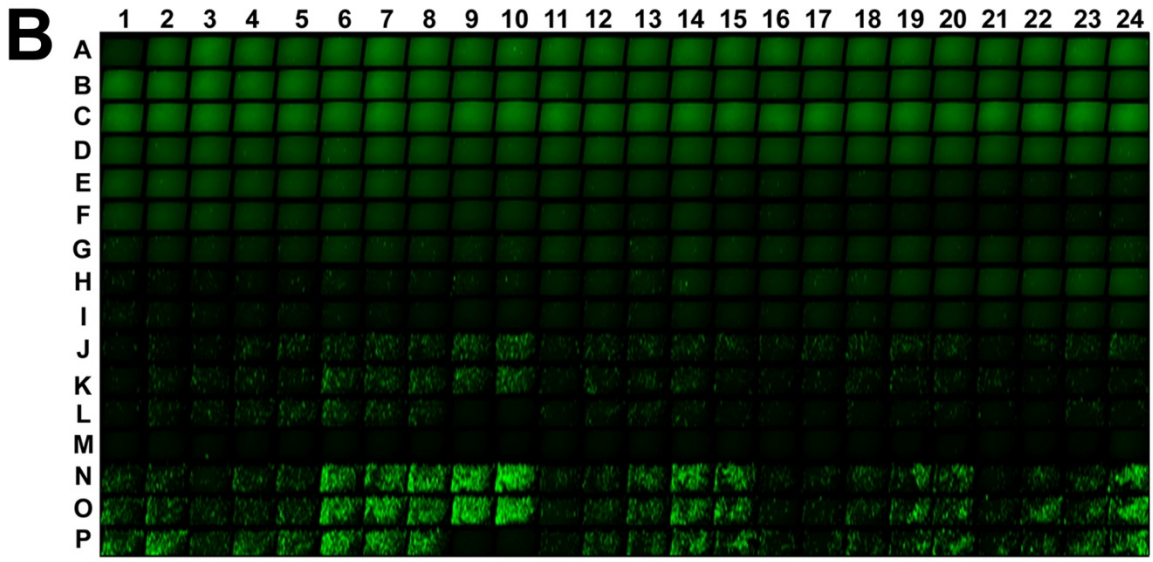
Figure 2-2. Optimization of the histone pre-mRNA processing reporter for use in high-throughput experiments. (A) Plating strategy for optimization of the conditions for screening for histone pre-mRNA processing factors in 384-well format. Cells were treated with varying amounts of reporter DNA, dsRNAs targeting dPTB (negative control) or dLsm11 (positive control), and Effectene reagents (enhancer and Effectene). Cells were treated with dsRNAs or 2'-OCH3 oligonucleotide for 3 days before transfection of the His_{proc} reporter with Effectene reagent. Cells were then incubated for 2 additional days before viewing on a fluorescence microscope. (B). Final image of the 384-well plate after running a Metamorph algorithm to convert the GFP signal into vertical bars with height corresponding to the intensity of the signal (top). Zoomed images of three matched wells from the experiment that were subjected to the same transfection conditions, but different dsRNAs. Abbreviations: His_{proc} = histone H3 promoter-driven histone pre-mRNA processing reporter; GFP = green fluorescent protein.

A



YELLOW	ORANGE	GREEN	BLUE
-5 DNA conc.	-5 DNA conc.	-5 DNA conc.	-5 DNA conc.
-5 enhancer conc.	-0.5 enhancer in all	-0.3 enhancer in all	-0.2 enhancer in all
-0.3ul effect in all	-3 effect conc.	-3 effect conc.	-3 effect conc.
-2 (–)dsRNA dose	-2 (–)dsRNA dose	-2 (–)dsRNA dose	-2 (–)dsRNA dose
-2 (+)dsRNA dose	-2 (+)dsRNA dose	-2 (+)dsRNA dose	-2 (+)dsRNA dose
-1 2'omethyl dose	-1 2'omethyl dose	-1 2'omethyl dose	-1 2'omethyl dose
25 conditions x 5 wells =125	15 conditions x 5 wells =75	15 conditions x 5 wells =75	15 conditions x 5 wells =75

B



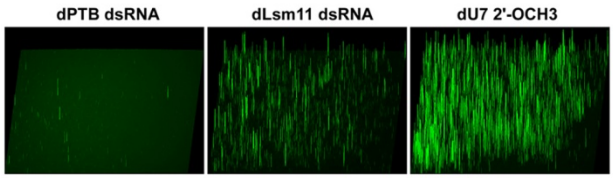
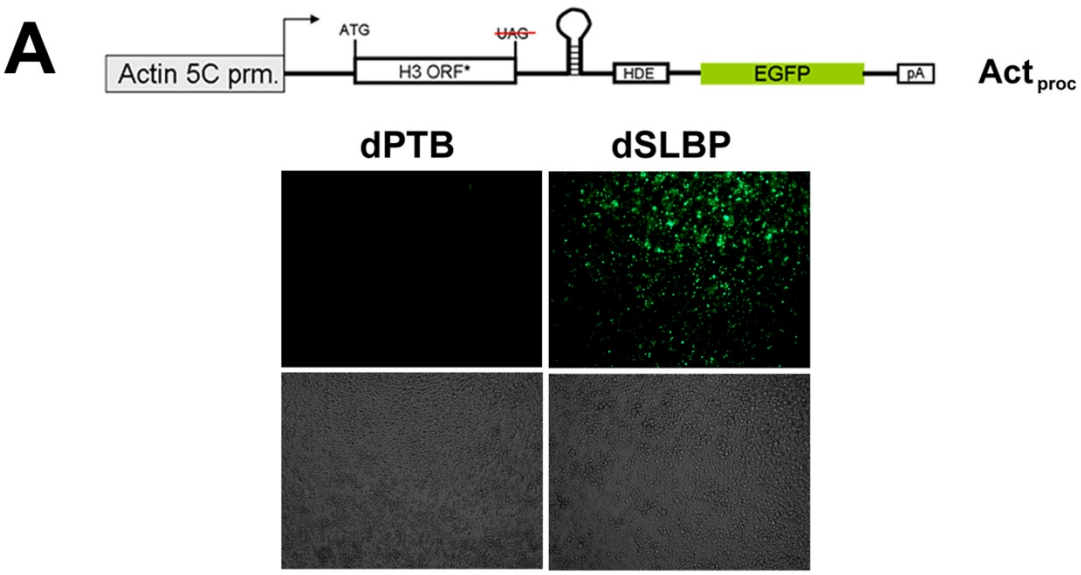
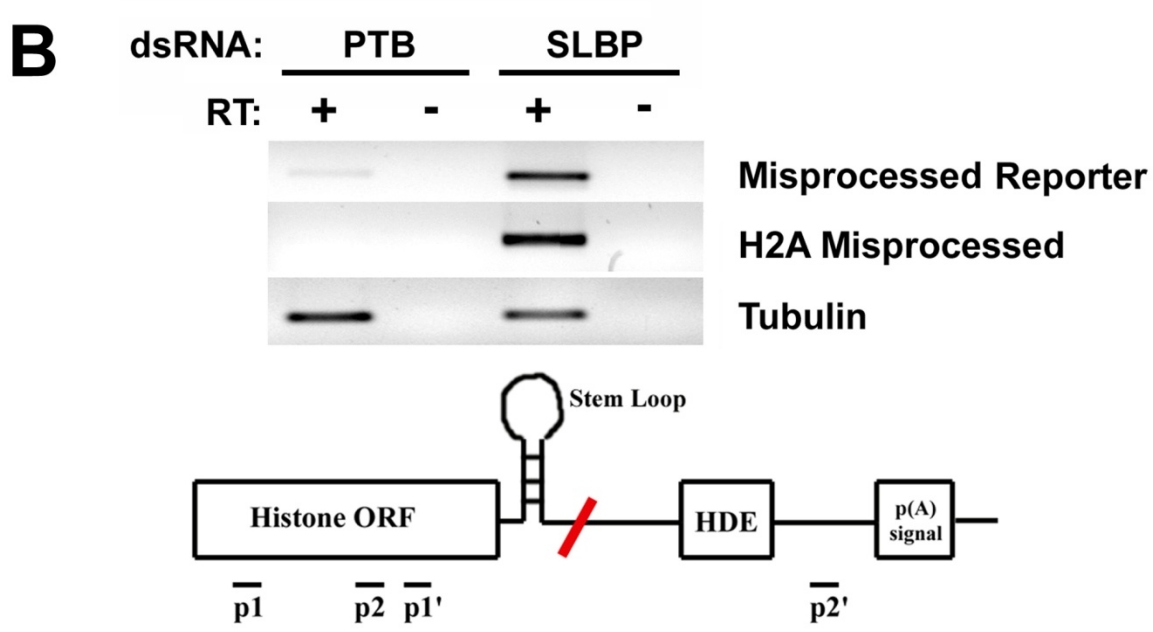


Figure 2-3. Creation and validation of a constitutively expressed histone pre-mRNA processing reporter and a histone transcription reporter. (A) Diagram showing the features of the Act_{proc} reporter (top). *Drosophila* D.Mel-2 cells stably expressing the Act_{proc} reporter were treated with dsRNAs targeting dPTB (negative control) or dSLBP (positive control) and incubated for 3 days before viewing on a fluorescence microscope to detect GFP signal (bottom). (B) Total RNA taken from the cells in Panel A were used as template for PCR either with (+) or without (-) prior reverse transcription to detect read-though histone mRNA transcripts. A diagram of the histone-specific primer targets is shown below. (C) *Drosophila* D.Mel-2 cells stably expressing the His_{trans} reporter were treated with dsRNAs targeting dPTB and dSLBP for 3 days and viewed using a fluorescence microscope to detect GFP. Abbreviations: Act_{proc} = actin 5C promoter-driven histone pre-mRNA processing reporter; His_{trans} = histone H3 promoter-driven transcription reporter; prm = promoter; ORF = open reading frame; EGFP = enhanced green fluorescent protein; pA = polyadenylation signal; PTB = pyrimidine tract binding protein; SLBP = stem loop binding protein.


A
Actin 5C prm.
ATG
UAG
H3 ORF*
HDE
EGFP
pA
Act proc
dPTB
dSLBP


B
dsRNA:
PTB
SLBP
RT:
+
-
+
-
Misprocessed Reporter
H2A Misprocessed
Tubulin
Stem Loop
Histone ORF
HDE
p(A) signal
p1
p2
p1'
p2'

C

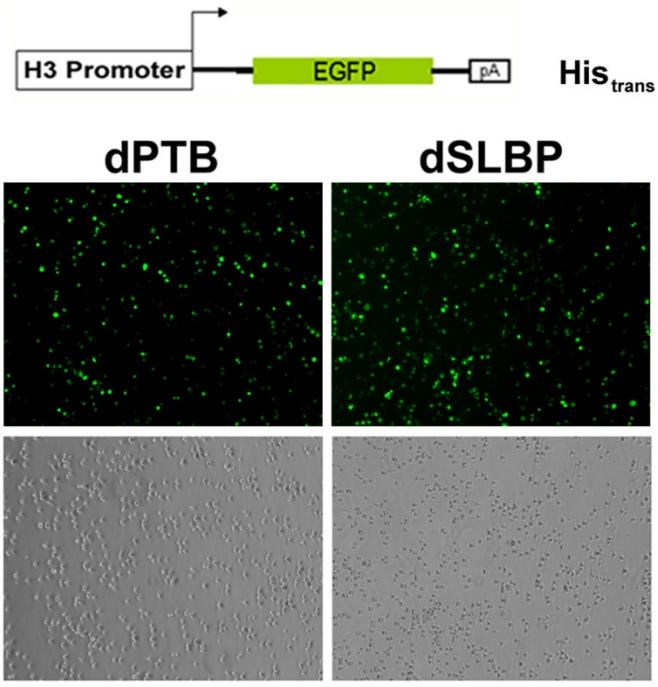
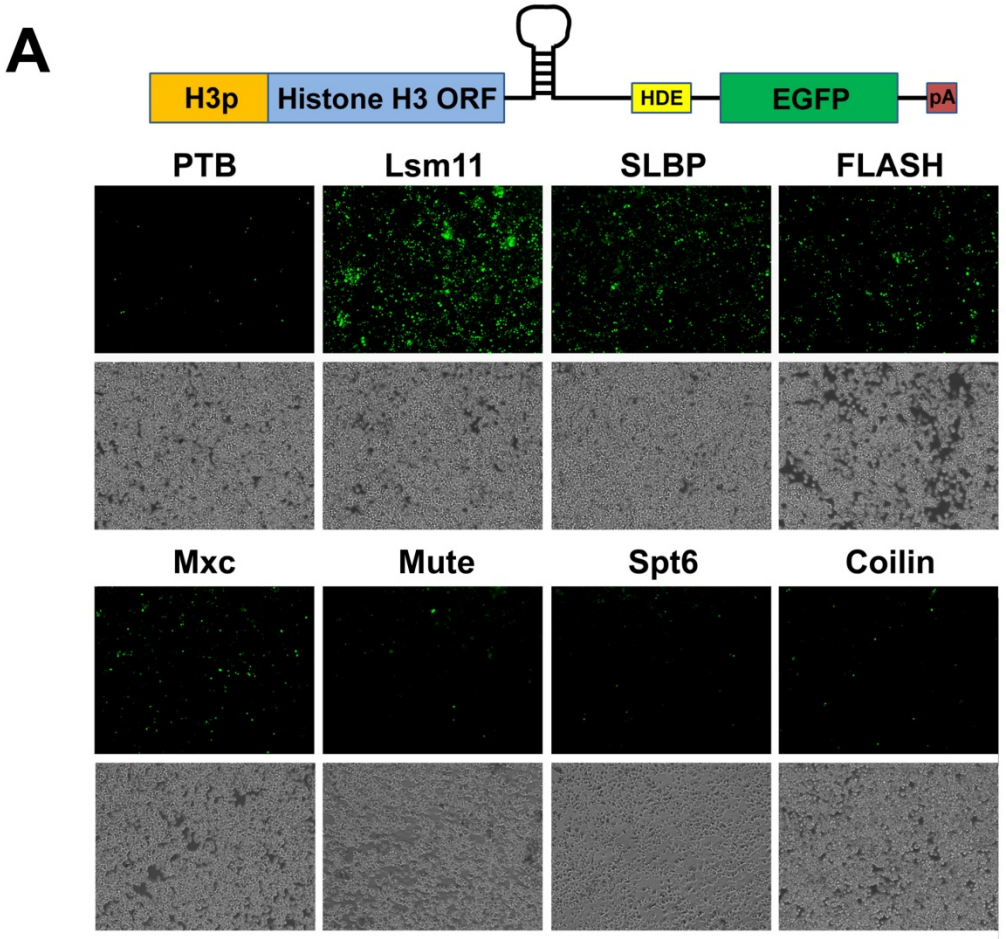
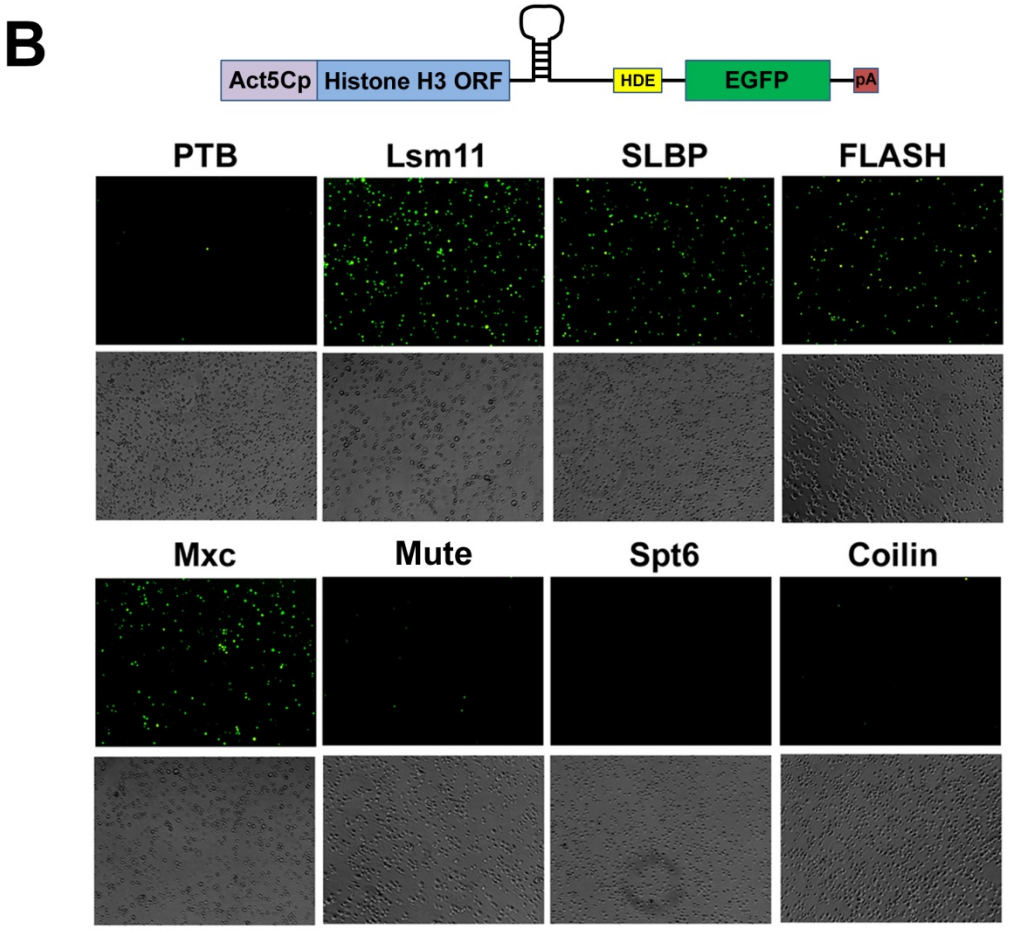
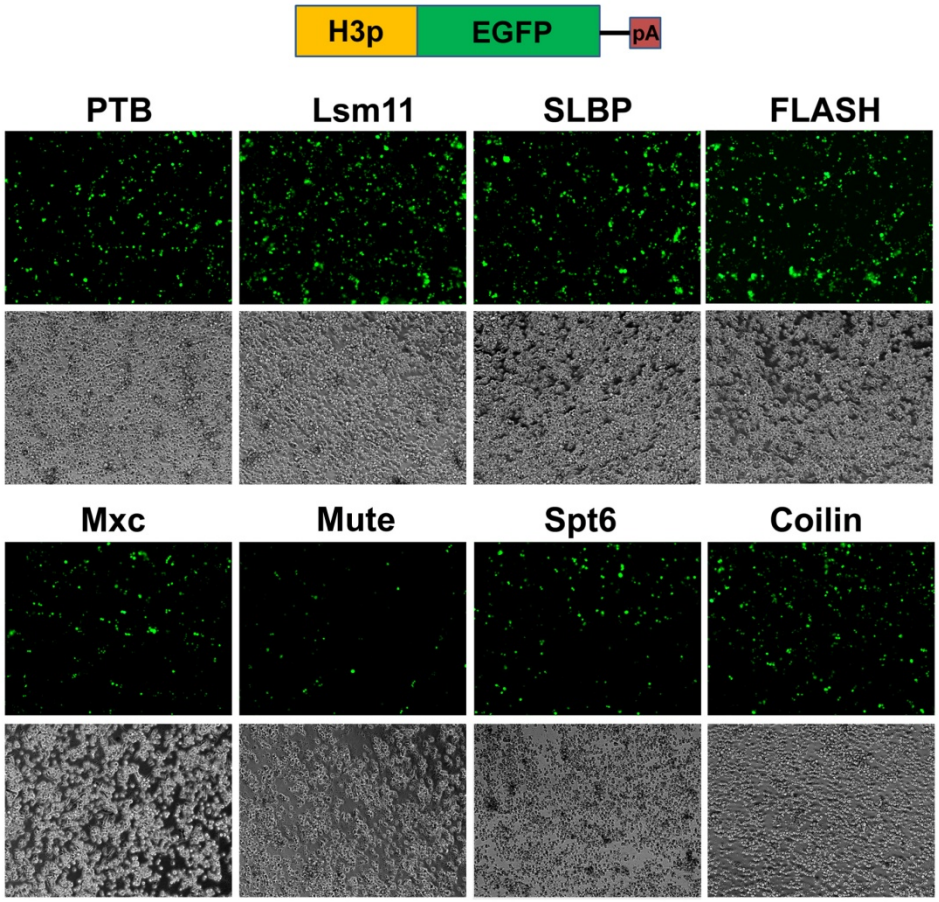


Figure 2-4. Using reporters to determine steps in histone mRNA metabolism in which specific proteins function. (A) *Drosophila* D.Mel-2 cells were treated with dsRNAs targeting the indicated factors and after 3 days, were transfected with the His_{proc} reporter. Two days later, the cells were viewed using a fluorescence microscope to detect GFP signal. (B) *Drosophila* D.Mel-2 cells containing the Act_{proc} reporter were treated with dsRNAs targeting the indicated factors and after 5 days, were viewed using a fluorescence microscope to detect GFP signal. (C) *Drosophila* D.Mel-2 cells were treated with dsRNAs targeting the indicated factors and after 3 days, were transfected with the His_{trans} reporter. Two days later, the cells were viewed using a fluorescence microscope to detect GFP signal. (D) *Drosophila* D.Mel-2 cells containing the Act_{proc} reporter were treated with dsRNAs targeting nearly 100 factors from the genome-wide MPM-2 antibody screen (Anne White). Shown is a representation of some of these factors. Note that Mute was negative, FLASH scored very strongly positive, and the remainder scored weakly (above background) to moderately positive. Abbreviations: His_{proc} = histone H3 promoter-driven histone pre-mRNA processing reporter; Act_{proc} = actin 5C promoter-driven histone pre-mRNA processing reporter; His_{trans} = histone H3 promoter-driven histone transcription reporter; GFP = green fluorescent protein.


A
H3p
Histone H3 ORF
HDE
EGFP
pA
PTB
Lsm11
SLBP
FLASH
Mxc
Mute
Spt6
Coilin


B
Act5Cp
Histone H3 ORF
HDE
EGFP
pA
PTB
Lsm11
SLBP
FLASH
Mxc
Mute
Spt6
Coilin

C



D

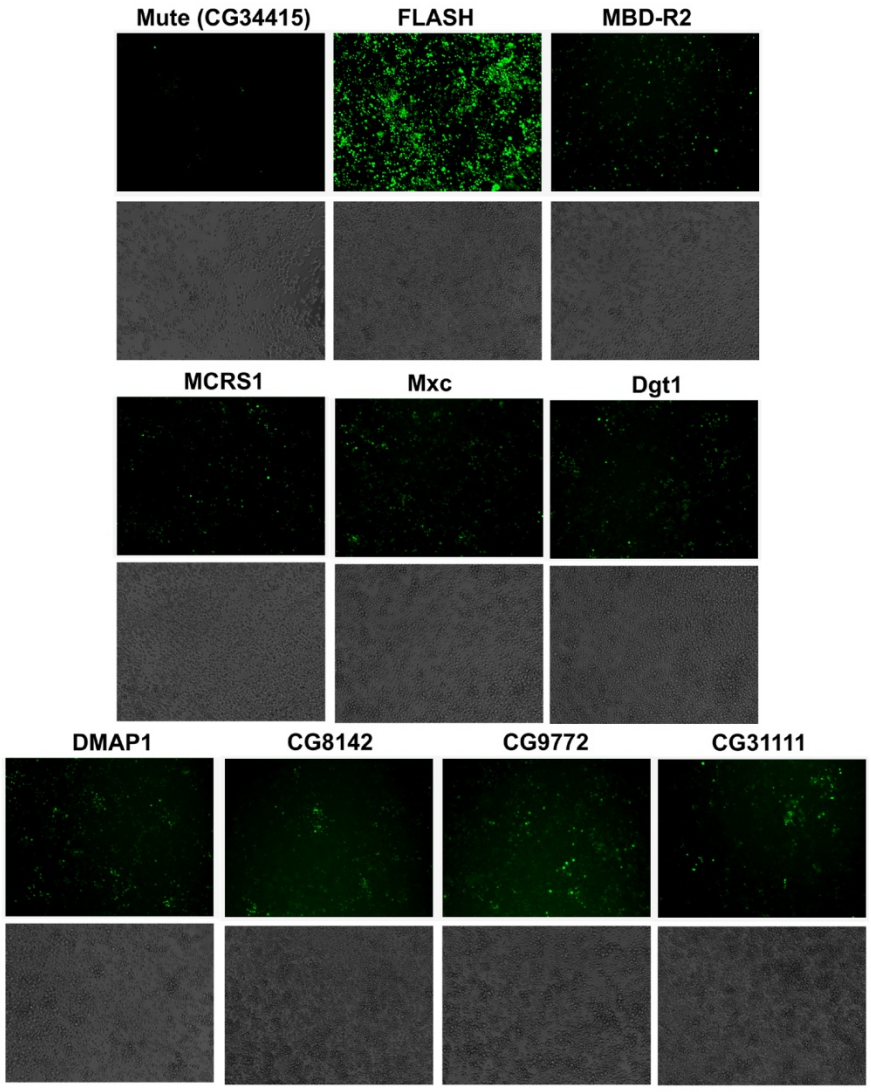
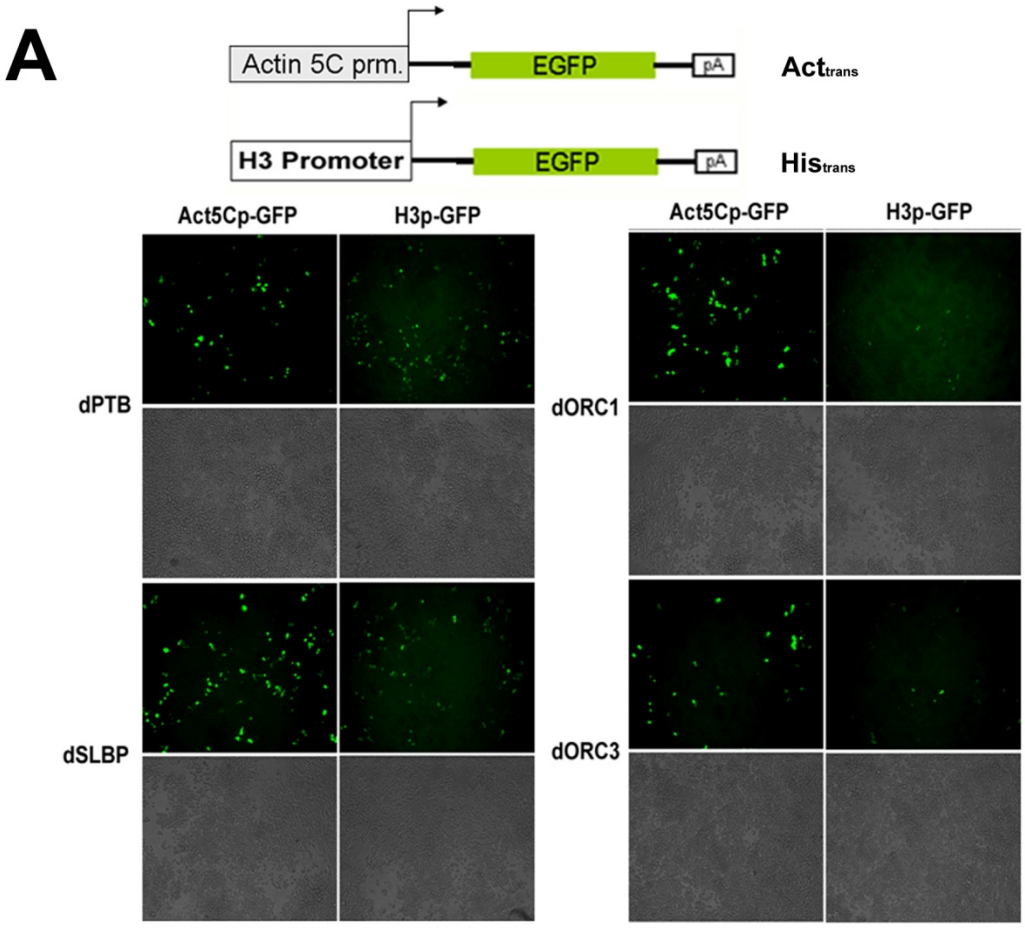
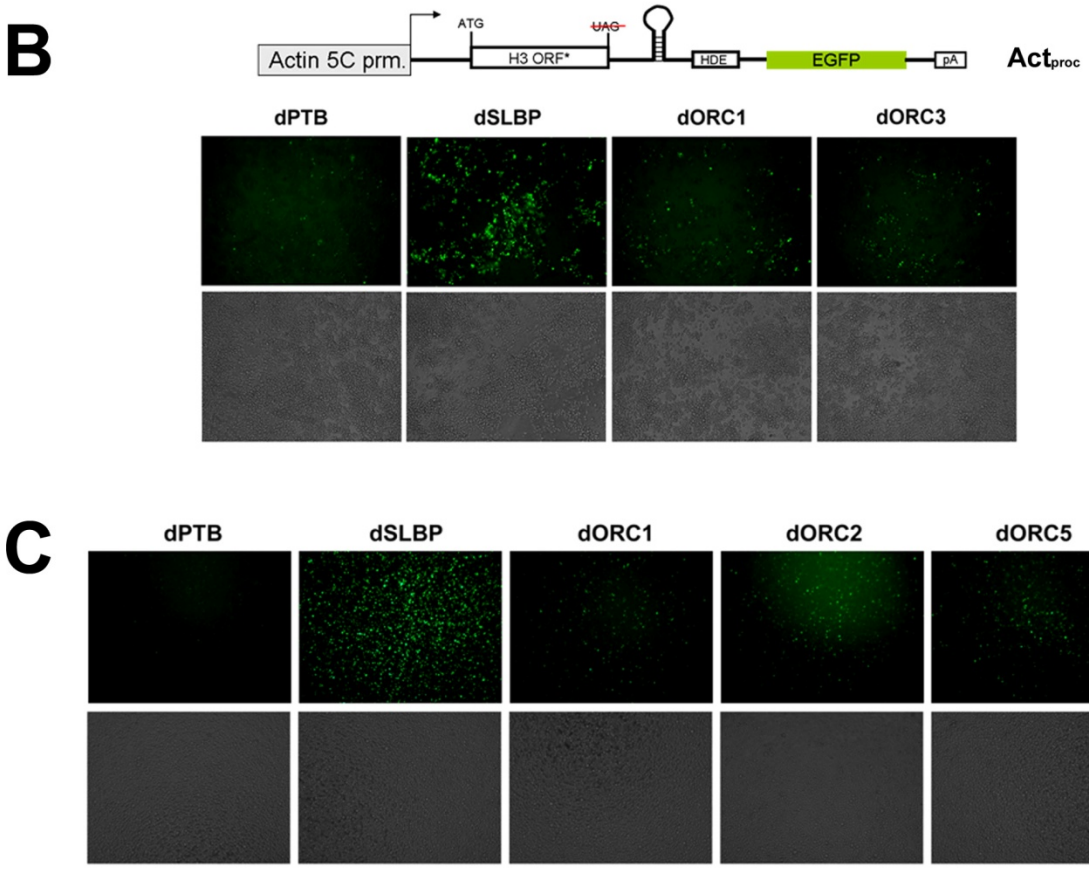
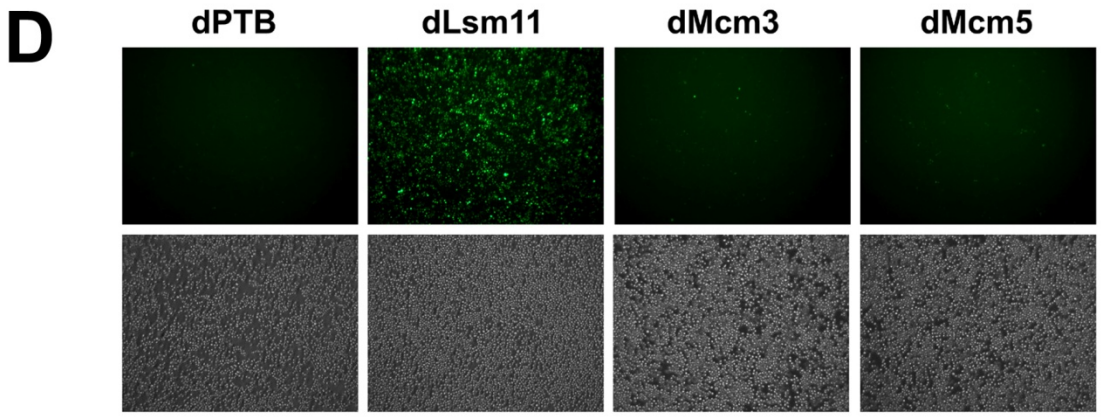

Mute (CG34415)
FLASH
MBD-R2
MCRS1
Mxc
Dgt1
DMAP1
CG8142
CG9772
CG31111

Figure 2-5. Effect of depletion of DNA replication factors on histone mRNA metabolism. (A) *Drosophila* D.Mel-2 cells were depleted of dPTB, dSLBP, dORC1, or dORC3, incubated 3 days and transfected with the His_{trans} reporter. As a control, a second reporter, containing an Actin 5C promoter driving GFP, was used. Cells were viewed after 5 days using a fluorescence microscope to detect GFP. (B) *Drosophila* D.Mel-2 cells were depleted of dPTB, dSLBP, dORC1, or dORC3, incubated 3 days and transfected with the Act_{proc} reporter. Cells were viewed 2 days later using a fluorescence microscope to detect GFP. (C) *Drosophila* D.Mel-2 cells stably expressing the Act_{proc} reporter were depleted of dPTB, dSLBP, dORC1, or dORC3. Cells were viewed after 5 days using a fluorescence microscope to detect GFP. (D) *Drosophila* D.Mel-2 cells stably expressing the Act_{proc} reporter were treated with dsRNAs targeting dPTB, dLsm11,dMcm3, or dMcm5 and viewed after 5 days using a fluorescence microscope to detect GFP. (E) *Drosophila* D.Mel-2 cells stably expressing the Act_{proc} reporter were treated with dsRNAs targeting dPTB,dSLBP, Dup, or dGem and viewed after 3 days using a fluorescence microscope to detect GFP signal (top). A zoomed image of cells treated with dsRNA targeting Dup or Gem is shown in comparison to mock treated cells (bottom). Note the small size of cells depleted of Dup and the large size of cells depleted of Gem relative to the control. All images are shown at 100X magnification. Abbreviations: His_{trans} = histone H3 promoter-driven histone transcription reporter; Act_{trans} = actin 5C promoter-driven transcription reporter; Act_{proc} = actin 5C promoter-driven histone pre-mRNA processing reporter; GFP = green fluorescent protein; dPTB = pyrimidine tract binding protein; SLBP = stem loop binding protein; ORC = origin recognition complex subunit; dLsm11 = Sm-like protein 11; Mcm = minichromosome maintenance protein subunit; Dup = doubleparked (*Drosophila* Cdt1); dGem = geminin.


A
Actin 5C prm.
EGFP
pA
Act trans
H3 Promoter
EGFP
pA
His trans
Act5Cp-GFP
H3p-GFP
dPTB
dSLBP
Act5Cp-GFP
H3p-GFP
dORC1
dORC3


B
Actin 5C prm.
ATG
UAG
H3 ORF*
HDE
EGFP
pA
Act proc
dPTB
dSLBP
dORC1
dORC3
C
dPTB
dSLBP
dORC1
dORC2
dORC5


D
dPTB
dLsm11
dMcm3
dMcm5

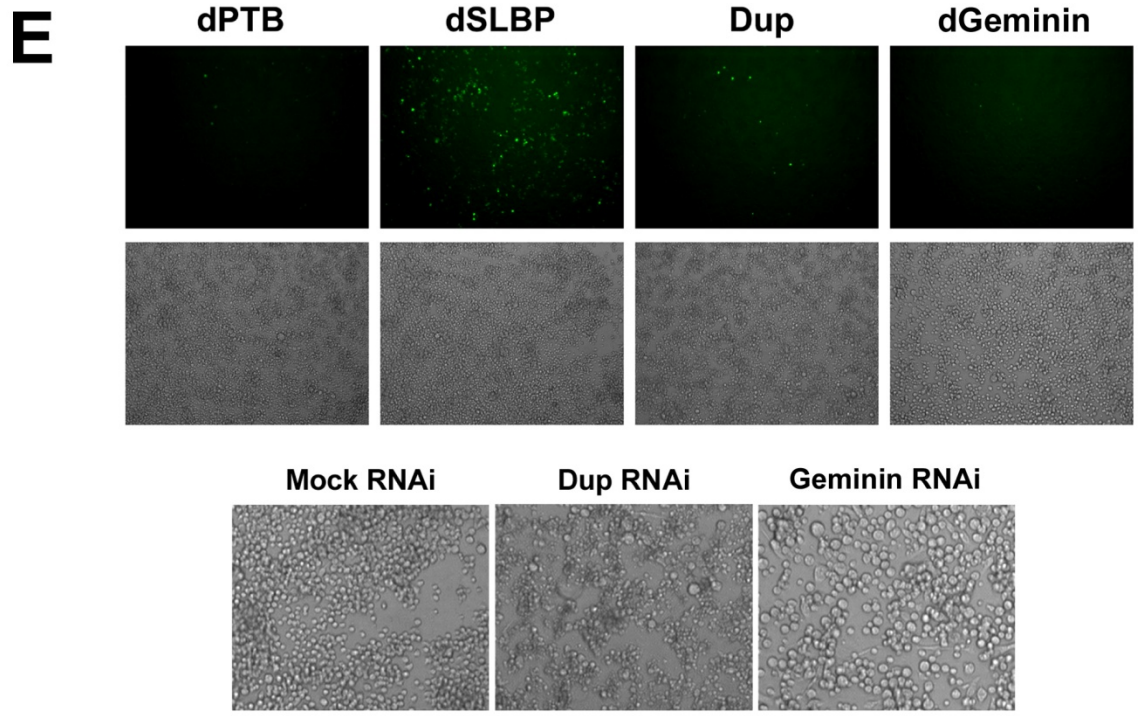
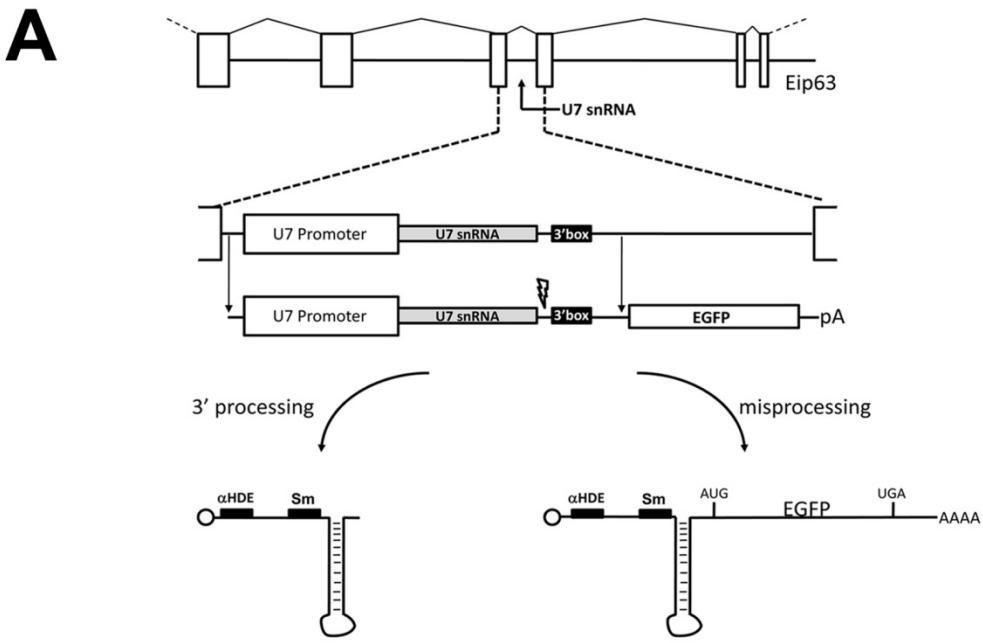
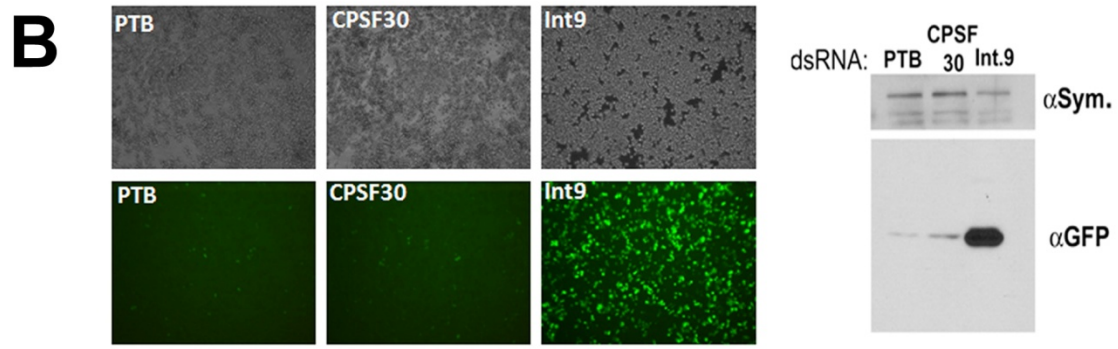
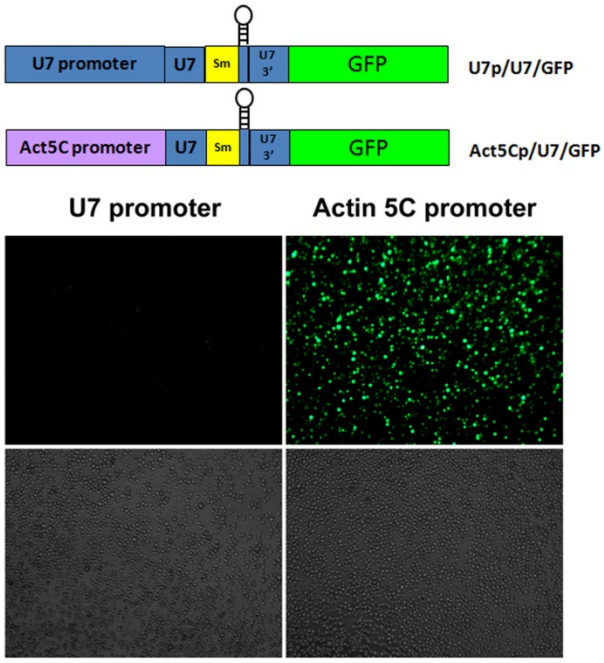

E
dPTB
dSLBP
Dup
dGeminin
Mock RNAi
Dup RNAi
Geminin RNAi

Figure 2-6. The use of novel reporters to characterize U7 pre-snRNA processing. (A) Diagram of the U7 pre-snRNA processing screening reporter. (B) *Drosophila* D.Mel-2 cells were treated with dsRNAs to PTB (negative control), CPSF30 (negative control), and Int9, incubated 3 days, transfected with the U7 pre-snRNA processing reporter and viewed using a fluorescence miscrosope to detect GFP 2 days later (left). Western blot analysis of protein lysates from these cells using antibodies targeting Symplekin (loading control) and GFP (to aid in quantification of fluorescence) (right). (C) Diagram of U7 snRNA processing reporters containing either a U7 promoter or Actin 5C promoter (top). *Drosophila* D.Mel-2 cells were transfected with the indicated reporter and viewed after 3 days using a fluorescence microscope to detect GFP (bottom). RNAi was not used. (D) Diagram of U7 snRNA processing reporters containing either a U7 snRNA or histone H3 promoter driving a wild type U7 transcript sequence or a U7 snRNA promoter driving a U7 snRNA sequence containing an insertion from the histone H3 gene (top). *Drosophila* D.Mel-2 cells were transfected with the indicated reporter and treated with mock dsRNA or dsRNA targeting Int9 after 3 days, then viewed on a fluorescence microscope to detect GFP after 2 more days (bottom). Abbreviations: U7 = U7 snRNA transcript; GFP = green fluorescent protein; Int9 = integrator subunit 9; Eip63 = ecdysone-induced protein 63; αHDE = U7 snRNA sequence complementary to the histone downstream element in histone mRNA; Sm = Sm binding region; AUG = start codon; EGFP = enhanced green fluorescent protein; UGA = stop codon; PTB = pyrimidine tract binding protein; CPSF30 = cleavage and polyadenylation specificity factor subunit 30; Int9 = integrator subunit 9; α-Sym = antibody targeting Symplekin; α-GFP = antibody targeting GFP. These experiments were carried out with rotation student, Kirston Barton. (Panels A, B courtesy of Eric Wagner)

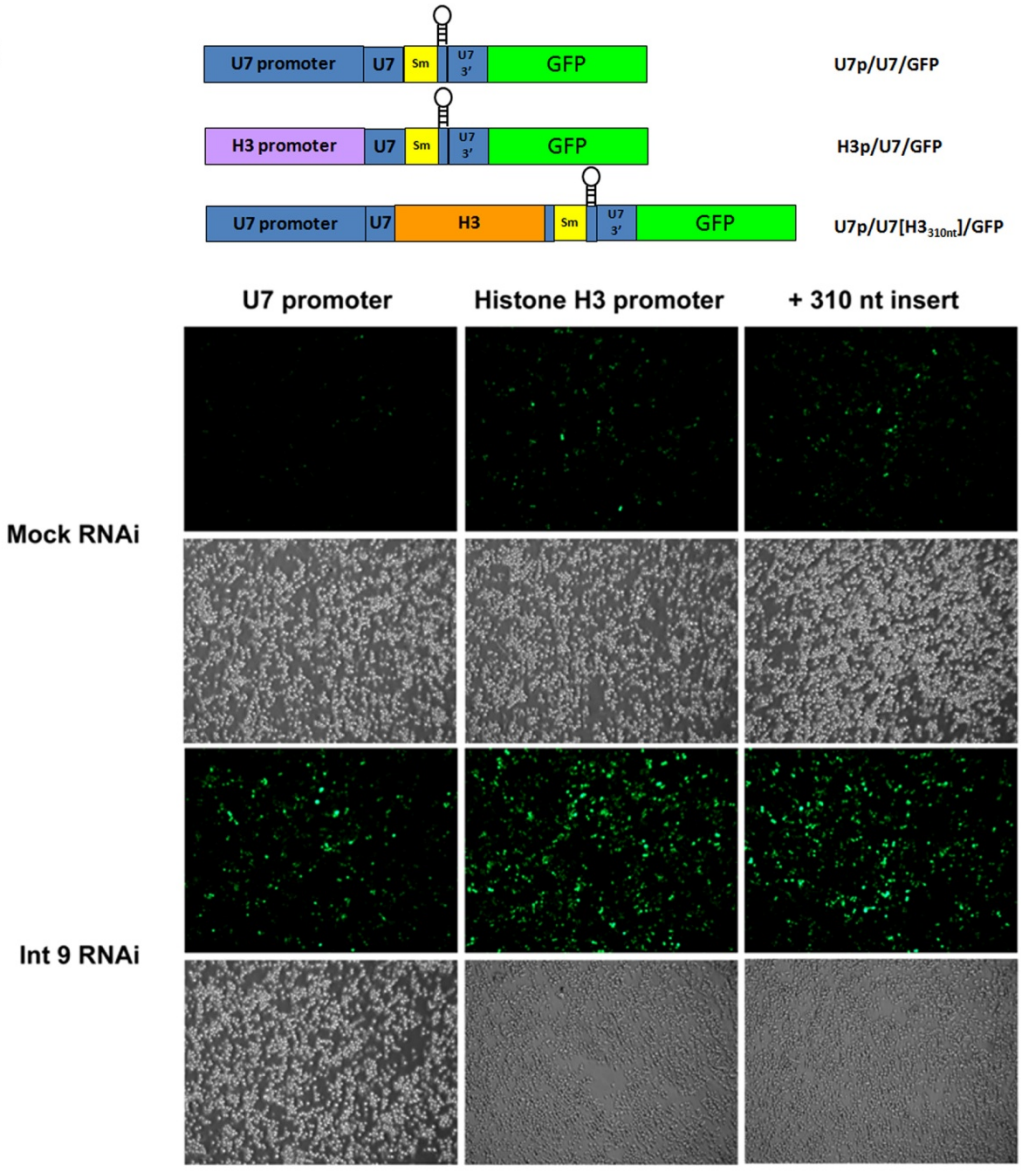

A
Eip63
U7 snRNA
U7 Promoter
U7 snRNA
3'box
U7 Promoter
U7 snRNA
3'box
EGFP
pA
3' processing
misprocessing
αHDE
Sm
αHDE
Sm
AUG
EGFP
UGA
AAAA


B
PTB
CPSF30
Int9
PTB
CPSF30
Int9
CPSF
dsRNA: PTB 30 Int.9
αSym.
αGFP

C


U7 promoter
U7
Sm
U7 3′
GFP
U7p/U7/GFP
Act5C promoter
U7
Sm
U7 3′
GFP
Act5Cp/U7/GFP
U7 promoter
Actin 5C promoter

D


U7 promoter
U7
Sm
U7 3′
GFP
U7p/U7/GFP
H3 promoter
U7
Sm
U7 3′
GFP
H3p/U7/GFP
U7 promoter
U7
H3
Sm
U7 3′
GFP
U7p/U7[H3310nt]/GFP
U7 promoter
Histone H3 promoter
+ 310 nt insert
Mock RNAi
Int 9 RNAi

CHAPTER III

A GENOME-WIDE RNA INTERFERENCE SCREEN FOR HISTONE PRE-MRNA PROCESSING FACTORS

Introduction

Metazoan histone pre-mRNAs lack introns and require only a single 3' end processing event to form mature mRNA, which terminates in an evolutionarily conserved stem loop (SL) rather than a polyA tail. Processing occurs by endonucleolytic cleavage downstream of the SL and 5' of a purine-rich sequence, the histone downstream element (HDE) (Fig. 1A). Cleavage requires a protein that binds the SL (the Stem Loop Binding Protein or SLBP) (Wang *et al.*, 1996), and U7 snRNP, which interacts with the HDE via base-pairing with U7 snRNA (Mowry and Steitz, 1987). The 3' SL remains bound by SLBP and is necessary for the export, translation, and eventual decay of histone mRNA. Thus, accurate pre-mRNA processing is essential for the expression of histones during S phase.

A number of histone pre-mRNA processing factors have been identified from mammals, *Drosophila*, and *Xenopus*, including SLBP, U7 snRNA, and U7 snRNP-specific components Lsm11, Lsm10, and ZFP100 (Dominski *et al.*, 2002; Pillai *et al.*, 2001; Pillai *et al.*, 2003). Recent in vitro experiments have unexpectedly found that factors involved in the canonical cleavage/polyadenylation reaction, such as CPSF73 and Symplekin, also participate in histone pre-mRNA processing (Dominski *et al.*, 2005; Kolev and Steitz, 2005). However, it is unclear whether all or only a subset of the cleavage/polyadenylation factors are necessary for processing histone pre-mRNAs. To address this and to identify additional factors necessary for histone pre-mRNA processing in intact cells, we carried out a genome

wide RNA interference screen in cultured *Drosophila* cells to identify proteins necessary for production of mature histone mRNA.

Materials and Methods

Cloning of GFP reporter

The histone pre-mRNA processing reporter was created by subcloning the histone H3 promoter, partial open reading frame, downstream sequence, and an EGFP coding sequence into a promoterless pIZ/V5/His vector backbone (Invitrogen). Briefly, the histone H3 5'UTR and amino acids 1-67 were amplified by PCR using forward primer 5' GGCCGAATTCCGACAAAAAC CCGAGAGAGTAC 3' and reverse primer 5' GGCCGGTACCTTAGGCAGCTTG CGGATTAGAAGC 3' and subcloned into pEGFPN1 (Clontech, Palo Alto CA) using EcoRI and KpnI. The 3' end of the H3 gene starting immediately after the stop codon and continuing until 18 nt downstream of the HDE was amplified using forward primer 5' GGCCGGTACCACTTGCAGAT AAAGCGCTAGCG 3' and reverse primer 5' GGCCGGATCCTTGTTATAAATAG TCGGCAACA GAAAATTTTTTCTC 3' followed by ligation to the 5' product using KpnI and BamHI. The resulting construct contained the H3 promoter, amino acids 1-67 of histone H3 ORF, and H3 3' end containing a portion of the downstream intergenic region upstream of an EGFP ORF. Note that the H3 ORF is in frame with GFP and that we created a single mutation (U to A) within the HDE in order to disrupt a stop codon (red box in Fig. S1). The OpIE2 promoter was removed from pIZ-V5/His (Invitrogen, Carlsbad CA) vector by inverse PCR using forward primer 5' GGCCGCTAGCACAG CATCTGTTCGAATTTA 3' and reverse primer 5' GGCCGCTAGCAGACAT GATAAGATACATTGATGA 3' followed by digestion with NheI and religation. The reporter was then subcloned into the promoterless pIZ/V5/His vector using forward primer 5'

GGCCGAATTCCGACAAAAACCCGAGAGAG TAC 3' and reverse primer 5' GGCCTCTAGATTACTTGTACAGCTCGTCCAT GCC 3' followed by digestion with EcoRI and XbaI and subsequent ligation to form the pIZ/H3p/H3/GFP construct. Note that the polyA site used by the GFP ORF is the insect OpIE2 polyadenylation sequences 3' of the XbaI site in the pIZ/V5/His vector.

Construction of transgenic *Drosophila* expressing the histone/GFP reporter

The GFP reporter was subcloned into pCaSpeR-4 to generate transgenic *Drosophila*. Brains dissected from wandering third instar larvae of the genotype w^{1118}; P [GFP reporter]/P [GFP reporter]; $U7^{EY11305}$/ $U7^{EY11305}$ or w^{1118}; P [GFP reporter]/P [GFP reporter]; $Slbp^{15}$ / $Slbp^{15}$ were fixed for 20 minutes in 4% formaldehyde/PBS, incubated with 1^{o} antibodies rabbit α-GFP (1:1000; Upstate) and mouse α -phosphotyrosine (1:500; Abcam) followed by 2^{o} antibodies goat α -rabbit-Cy2 (1:500; Abcam) and goat α-mouse-Cy3 (1:500; Abcam), mounted in fluoromount-G (Southern Biotech), and analyzed with a Zeiss 510 Laser scanning confocal microscope.

Immunofluorescence of Polytene Nuclei and Brains

Salivary glands from early wandering 3rd instar larvae were dissected into 1X PBS containing 0.1% Triton-X-100. Glands were fixed for 2 minutes in 3.7% paraformaldehyde in 1X PBS-Triton X-100, and then for 2 minutes in 3.7% paraformaldehyde/50% acetic acid. Glands were immediately squashed and gently tapped under a coverslip onto a polylysine Q coated slide. Slides were washed 3X in PBS containing 0.1% Tween 20. Slides were incubated in the following primary

antibodies diluted in PBS-Tween 20 overnight at 37 degrees: HP1 (mouse C1A9 1:100 from Developmental Studies Hybridoma Bank, rabbit Lsm11 1:200 generously provided by Dr. Joe Gall (Carnegie Institute)). Slides were washed 3X in PBS-Tween 20, and incubated at room temperature for 1 hour with mouse α-Cy3 (1:500) or rabbit α-Cy5 (1:500) (Jackson Immunolabs). Slides were washed 1X in PBS-Tween 20 and stained with DAPI (1mg/ml) 1:1000 for 1 minute. Slides were washed in the dark 3X in PBS-Tween 20 and mounted in Fluoromount medium. Brains from wandering 3rd instar larvae were dissected into 0.1% Triton X-100 in PBS. They were then fixed in 4.5% formaldehyde in 0.1% Triton X-100 in PBS for 22 minutes. Brains were washed 3X in 0.1% Triton X-100 in PBS and subsequently treated for 1 hour in 1% Triton X-100 in PBS. Samples were blocked for 1 hour at RT in 5% Normal Goat Serum in 0.1% Triton X-100 in PBS. After blocking, samples were washed 3X in 0.1% Triton X-100 in PBS. Samples were incubated with primary antibodies overnight at 37 °C (rabbit α-Lsm11 (1:1000), mouse α-MPM-2 (1:1000) (Upstate)). Slides were washed 3X in 0.1% Triton X-100 in PBS, and incubated at room temperature for 1 hour with mouse α-Cy3 (1:500) or rabbit α-Cy5 (1:500). Slides were washed three times in 0.1% Triton X-100 in PBS and mounted with Fluoromount media.

Cell culture and transfection

D.Mel-2 cells (Invitrogen) were grown in serum free conditions in SF900II-SFM media (Invitrogen). For RNA interference in 384-well plates, 8000 cells were plated in each well in 10 μl of serum free media and incubated with 250ng of dsRNA resuspended in 5 μl of water. Cells were allowed to grow unperturbed for 72 hours. On the third day, 20 μls of serum free media was added to each well. Transfections were performed using Effectene

reagent (Qiagen) with a master mix using the following conditions per well: 50 ng of GFP reporter was incubated in 8.75 µl of EC reagent at RT for 5 minutes. 0.2 µl of Enhancer reagent was added to the DNA and incubated at RT for 5 minutes. 0.3 µl of Effectene reagent was added to the mixture and incubated for 10 minutes at RT. 10 µl of the mixture was pipetted into each well containing cells growing in 35 µl of serum free media. For hit validation, the RNA interference was done identically with the exception that cells were supplemented with an additional 500ng of dsRNA 12 hours post transfection and hence received two doses of dsRNA. For 6 well transfections, 2 $X10^6$ cells were plated in 2 mls of serum free media, followed by addition of 10µg of dsRNA. Each of the following 2 days an additional 10µg of dsRNA was added. The next day cells were transfected with 400ng of reporter using the manufacturer's protocols.

Genome-wide RNA interference

The screen was performed at the *Drosophila* RNAi Screening Center (DRSC) at Harvard University. 250ng of each dsRNA included in the library are pre-aliquotted into 384-well, black-walled, clear-bottom plates at a concentration of 50ng/ul. The library of 22,000 dsRNAs was distributed into a total of 63 384-well plates and each plate had an empty well in grid position B1 in which the 2'-OCH_3 oligonucleotide to *Drosophila* U7 snRNA was added as a positive control. In grid position B2 of each plate there is a dsRNA against *thread* which kills the Dml-2 cells to control for the effectiveness of the RNAi.

The screen was performed as follows: On Day 0, 8000 D.Mel-2 cells in a total of 10 µl of *Drosophila* SFM-II media (Gibco) plus antibiotic/antimycotic solution (1X) were plated into each well using a Wellmate microplate dispenser (Matrix Technologies). After plating,

the cells were incubated in a 24^{o}C humidified incubator (Percival). On Day 3, 20 μl of SFM-II media were added to the cells, followed by transfection of the reporter construct into the cells according to the transfection protocol outline above using the microplate dispenser. Following transfection, the cells were incubated at 24 ºC for two more days. They were imaged using the Discovery-1 high content screen system (Molecular Devices) equipped with a CataLyst Express (Thermo Scientific) robotic arm. Each plate was imaged using a 4x objective lens using a FITC filter set (ex. 470 nm). Using the Metamorph software suite (Molecular Devices) included with the imaging system, the images from each plate were combined into a 384-image collage and converted into fluorescence intensity plots using the software within the Metamorph program.

We compared the fluorescence signal in each well visually to that of surrounding wells and to the well containing the U7 2'-OCH_3 oligonucleotide. Positive hits were independently scored by two people (BDB and EJW). Our compiled lists were combined and each hit subsequently scored on a scale from 1-3, with "1" being a weak hit, "2" a moderate hit, and "3" a strong hit. To validate the hits, we synthesized templates and dsRNAs to 90 hits and used these dsRNAs in repeated RNA interference/reporter transfection experiments (N>6). dsRNAs reproducibly resulting in increased fluorescence were included in the final list.

S1 Nuclease Protection Assay

To create the probe construct for the assay, *Drosophila* histone H2A sequence was cloned into a TA vector. From a standard plasmid preparation (Qiagen), aproximately 3 μg of the plasmid was combined with 2 μl 10X bovine serum albumin (NEB), 2 μl 10X

NEBuffer 3 (NEB), qs to 18 µl with dH_2O, and 2 µl BspE I restriction enzyme (NEB), for a total reaction volume of 20 µl. The digest was then incubated at 37 °C for 2 h. Next, the reaction was purified using a PCR purification kit (Qiagen) and eluted the digested DNA in Buffer EB (Qiagen) for a final volume of 28 µl. The result of the initial digest was a linearized construct with a 5'-CCGG-3' overhang at the 5' end of the histone H2A sequence.

To label the DNA, the purified digest was combined with 4 µl 10X dG,A,T deoxynucleotide mix, 4.5 µl 10X NEBuffer 2 (NEB), 4 µl α-P^{32} dCTP (MP Biomedicals), and 3 µl large fragment of DNA polI (Klenow) enzyme (NEB), for a total reaction volume of 43.5 µl. The reaction was incubated at 25 °C for 20 min. This reaction results in the filling-in of the recessed 3' end and 3'-end labeling of the histone H2A sequence with α-P^{32} dCTP.

After the labeling reaction, the reaction was spun through a ProbeQuant G-50 Micro Column (Amersham) to remove unincorporated deoxynucleotides. Next, to the cleaned-up, labeled construct 5 µl 10X BSA (NEB), 1 µl 10X NEBuffer2 (NEB), and 4 µl HindIII restriction enzyme (NEB) was added for a total reaction volume of 60 µl. The reaction was then incubated at 37 °C for 2 h. After incubation, the reaction was run on a 2 % agarose gel and the labeled fragment was purified using a gel extraction kit (Qiagen). These steps result in a 650 nt, labeled histone H2A probe that is ready to be used in the S1 nuclease protection assay.

For the S1 nuclease protection assay, 5 µg total RNA from D.Mel-2 cells was combined with 1 µl of the labeled histone H2A probe and the solution was dried using a Speedvac SC100 (Savant). The dried RNA-probe pellet was resuspended in 10 µl 1X hybridization buffer (40 mM Pipes, pH 6.4, 500 mM NaCl, 1 mM EDTA, 80 % deionized

formamide) and boiled for 5 minutes. After boiling, the samples were incubated at 52 °C overnight to allow hybridization of the probe to its target RNA sequence.

After hybridization, 80 µl dH_2O, 9 µl 10X S1 buffer (Promega), and 0.6 µl S1 enzyme (Promega) was added to the samples followed by incubation at 25 °C for 1.5 h. After incubation, the DNA/RNA hybrids were precipitated by adding 300 µl 100 % ethanol and 1 µl GlycoBlue (Ambion) to each sample and placing the samples at – 80 °C for 30 minutes. After precipitation, the samples were centrifuged at 16000xg in an Eppendorf 5415C microcentrifuge for 30 minutes at 4 °C. After removing the supernatant, the pellet was washed with 80 % ethanol and spun the samples again at 16000xg for 10 min. at 4 °C. The ethanol was then removed and the samples were allowed to air dry. After the pellets had dried, they were resuspended in 40 µl formamide loading dye (>99 % formamide with bromophenol blue and xylene cyanole), boiled for 5 min. and loaded onto a 6 % Sequagel Sequencing System acrylamide gel (National Diagnostics) in 1X TBE (Tris-borate-EDTA) buffer.

After running the gel, the gel was dried on a SGD200 Slab Gel Drier (Savant) for 1.5 hours at 80 °C. After drying, the gel was placed either on a storage phosphor screen (Molecular Dynamics) or onto blue autoradiography film (Denville Scientific) for detection. Phosphor screens were developed on a Storm Scanner (Amersham), and film was developed using a SRX-101A film developer (Konica).

RT-PCR Analysis

D.Mel-2 cells were treated with dsRNA for three consecutive days and then allowed to grow for two more days. The cells were then harvested and total RNA was isolated

using Trizol reagent (Invitrogen). 4 μg of RNA was used for a reverse transcription reaction using random hexamers as primers according to manufacturer's protocols (MMLV, Invitrogen). *Drosophila* PTB was amplified using standard PCR techniques using primers F 5'GTTTCTTTAAAGTGGGTAGCGACC 3' and R 5'TGGTGGCCAGGTCCTGATTGTC 3' at 35 cycles (55 °C annealing temp) and resolved on a 3% agarose gel using ethidium staining to visualize PCR products. Analysis of the CPSFs and CstFs were done using the following primer sets and cycling conditions: CPSF30 primers F 5'GCATCCGCACTTCGAGCTGCCC 3' and R 5' TTGACGAAGCCGGGACCATGC 3' using 63°C as an annealing temp. and 32 cycles. CPSF73 primers F 5'GTATGATGCAGTCGGGATTGTCGC 3' and R 5' TGCTCGCCGTGGACGAGCACGAC 3' using 63°C as an annealing temp and 28 cycles. CPSF100 primers F 5' AGAGAGCAGTTCCGAGTCCGAGG 3' and R 5' AACTCATAGCCTGTCGCATCCGC 3' at an annealing temp or 60°C and 28 cycles. CPSF160 primers F 5'TGTCGGTAATCTCGGACAGCAGC 3' and R 5' CCTTTCACTGTGAATAGTCCTG 3' using 49°C as an annealing temp. and 40 cycles. CstF50 primers F 5' CCCGGAACCACATTCCTACGAAACC 3' and R 5' TCGCGCGAGGCGGAGGCCAGGA 3' using 55°C as an annealing temp and 37 cycles. CstF 64 primers F 5' GGCCAATGTCCATCCGAACGATATCG 3' and R 5' TGCTGAGGACCAGGTCCTGCCC3' using 61°C as an annealing temp and 36 cycles. CstF77 primers F 5' ACGAGTCGCTAGTTAATGTGTTTCC 3' and R 5' CTGAACGAGTGTAGATCCATGCCG 3' using 55°C as an annealing temp and 36 cycles. Symplekin primers F 5' CCTTCAGTCGAGAGCCGCCAATGC 3' and R 5' AGTTGCGTGGCAGAGGCTTGGTG 3' using 63°C as an annealing temp and 36

cycles. Fip1 primers F 5’ GAGGAACCCTTCTTCCACGAGCC 3’ and R 5’ GTTGGGTGGCATAATGCCGCGCAA 3’ using 55°C as an annealing temp and 36 cycles.

To the analyze the expression of the domino gene, 2μg of total RNA from D.Mel-2 cells was incubated with 200ng of random hexamer primers (GE Healthcare Bio-Sciences AB) at 95°C for 1 minute and cooled to room temperature. Reverse transcription reactions were done in a total of volume of 20 μl according to manufacturer’s protocols (Invitrogen). 2 μl of each RT reaction was used as template for a PCR reaction using the cycling conditions: 95ºC/30sec, 57.5ºC/30sec, and 72ºC/40sec for a total of 32 cycles. The domino B isoform was amplified using the primer set: F 5’ GCCAAAGC TGCCGAAGAAAGAAG 3’ and R 5’ CACCACTGACTGC TGCTGATGAG ‘3. The domino A isoform was amplified using the primer set: F 5‘ATCAGCTCAAGCCCT GGCTGCGG 3’, and R 5’ CCAGCGGTTG GACCCGCAATACTT 3’. All RT-PCR products were sequenced to confirm identity and in all cases there were no products in the absence of RT.

Western Blot Analysis

D.Mel-2 cells were lysed in a buffer containing 50mM Tris-HCl (pH 8.3), 0.1% NP-40, and 50mM NaCl for 30’ on ice. Lysates were separated with SDS-PAGE using standard techniques and were then transferred to immobilon-PVDF membrane (Biorad, Hercules CA) and probed with either a 1:1000 dilution of an α-GFP JL8 monoclonal antibody (Clontech), a 1:1000 dilution of α-dSLBP antibody (*10*), a 1:1000 dilution of α-Lsm11 antibody, or a 1:1000 dilution of α-Lsm10 antibody (*24*). Blots were then probed using a secondary HRP-conjugated antibody and developed using chemiluminescence.

Sufficient material could be obtained for Western blotting using cells from 6 wells of a 384 well plate.

Northern Blot

RNA was isolated from wandering third instar *H2Av* mutant larvae of the genotype *w1118*; +/+; *H2Av810*/*H2Av810* (identified using a TM6B balancer) using TRIzol Reagent (Gibco/Invitrogen). 2 µg of total RNA from each sample was resolved by formaldehyde agarose gel electrophoresis, and probed with probes to either the histone H3 or histone H2a coding regions.

Nuclear extracts from RNAi treated cells

10^7 D.Mel-2 cells were plated in 10 mls of SFII-900M media and incubated with 100µg of dsRNA. The same amount of dsRNA was added to the culture each day for 2 more days. The cells were allowed to grow for 2 more days and then were harvested. Nuclear extracts were prepared as described previously (Dominski *et al.*, 2002) with the exception that the cells were lysed in Buffer A using 40 strokes through a 27.5 gauge needle. Lysates were dialyzed overnight in Buffer D using a slide-a-lyzer cassette (Pierce). Processing reactions were done using 12.5µl of nuclear extract and in conditions described previously (Dominski *et al*., 2002).

Results

Creation and validation of the histone pre-mRNA processing reporter

To facilitate screening, we developed a visual reporter assay for histone pre-mRNA processing. The mini-gene reporter contains a GFP ORF and polyadenylation signal downstream of the SL and HDE of the histone H3 gene (Fig. 1B). Normal histone pre-mRNA processing results in mRNA lacking the GFP ORF, while misprocessing leads to read-through transcription and production of mRNA encoding GFP. Transgenic flies containing the reporter display robust GFP expression in *U7* snRNA or *Slbp* null mutant brains compared to wild type controls (Fig. 3-1C). Thus, the reporter accurately reflects the requirements for endogenous histone mRNA biosynthesis.

Transfection of the reporter into D.Mel-2 cells depleted of SLBP using dsRNA, or with U7 snRNA inhibited using a 2'OCH_3 oligonucleotide complementary to the 5' end of U7 snRNA (αU7), resulted in GFP expression (Fig. 3-1D). Little GFP signal was seen in cells treated with control dsRNA or a 2'OCH_3 oligonucleotide complementary to human U7 snRNA (Fig. 3-1E). GFP expression occurred after SLBP depletion despite only a low amount of transcriptional readthrough of the endogenous histone genes (Fig. 3-1F, lane 3 and 5; Fig. S2) and a modest (10-20%) readthrough of the reporter gene (Fig. 3-1F, lane 7), underscoring the sensitivity of the reporter to small amounts of misprocessing.

Genome-wide RNAi screen for factors that participate in histone pre-mRNA processing

Using this reporter we performed a genome-wide RNAi screen at the *Drosophila* RNAi Screening Center (www.flyrnai.org) (Boutros *et al.*, 2004). D.Mel-2 cells were

incubated with dsRNA for 3 days, transfected with the reporter, and analyzed by fluorescence microscopy 48 hours later. About 22,000 dsRNAs were tested in duplicate, resulting in 63 collages of 384-well plates. Plate 18 is shown as an example (Fig.2- 2A). The two positive wells (Fig. 3-2A, inset) contained dsRNA targeting the *Drosophila* orthologs of the human Lsm10 and Lsm11 proteins, specific components of the U7snRNP (Azzouz and Schumperli, 2003). We synthesized these dsRNAs and found they knocked down endogenous Lsm10 and Lsm11 proteins (Fig 3-2B).

By visual inspection, we initially scored 354 genes as potentially positive (Appendix A). We synthesized dsRNA targeting each of the top 90 genes and assayed them under identical conditions. Of these, 24 were found to score repeatedly above background (Fig. 3-3A). We designed dsRNAs targeting two independent regions of these factors (Fig. 3-3B). Each of the identified factors scored positively using the first set of dsRNAs (Fig. 3-3C) and were confirmed with dsRNA targeting a second site of the mRNA (Fig. 3-3D). Western blot analysis of protein taken from RNAi-treated cells transfected with the reporter verifies enhanced expression of GFP when these factors are depleted (Fig. 3-3E).

Only four of the polyadenylation factors are required for histone pre-mRNA processing *in vivo*.

While it has been suggested that many of the proteins involved in cleavage and polyadenylation might have some role in histone pre-mRNA processing (Kolev and Steitz, 2005), only four of these 9 factors (CPSF73, CPSF100, Symplekin, and Fip1) scored in our screen. We depleted each of the 9 known factors involved in cleavage/polyadenylation and only the same four factors activated the reporter (Fig. 3-4A, B). All of the dsRNA-treated

cells grew at near normal rates, suggesting that polyadenylation factors are present in excess in D.Mel-2 cells. Indeed, the reporter requires polyadenylation to score positively. To determine whether the dsRNA treatment for each of the polyadenylation factors caused a functional reduction in polyadenylation, we tested the usage of polyadenylation sites in *domino* mRNA, which has short and long isoforms resulting from utilization of two different polyadenylation sites within distinct 3' terminal exons (Ruhf *et al.*, 2001). Depletion of 7 of the 9 polyadenylation factors we tested resulted in increased usage of the distal *domino* polyadenylation site (Fig. 3-4C). Depletion of CPSF30, CstF64, or CstF50 resulted in the greatest usage of the distal polyadenylation site, although there was no effect on histone pre-mRNA processing. In contrast, knockdown of Symplekin scored strongly for histone pre-mRNA processing but did not affect polyadenylation of *domino* mRNA. We conclude that only a subset of cleavage/polyadenylation factors is necessary for histone pre-mRNA processing in *Drosophila* cultured cells.

Symplekin is concentrated in the histone locus body.

A nuclear structure termed the histone locus body (HLB) is associated with the *Drosophila* histone gene cluster (Liu *et al.*, 2006). The HLB is distinct from the *Drosophila* Cajal body, which contains SMN and the U85 snRNA (Liu *et al.*, 2006). U7 snRNP is enriched in the HLB and is visualized with antibodies against Lsm11 (Fig. 3-4D). Other components of the HLB are likely involved in histone mRNA biosynthesis. We transiently or stably expressed Myc-tagged versions of the proteins identified in the screen and analyzed their localization with anti-Myc antibodies. Most of these proteins localized to the nucleus but did not concentrate in subnuclear foci. An exception was Myc-tagged Symplekin, which

concentrated in the HLB (Fig. 3-4D). Myc-MCRS1 was detected in several nuclear foci, one of which often overlapped with the HLB (Fig. 3-4D). In HeLa cells, the MCRS1 orthologue localizes to several discrete nuclear foci (Davidovic *et al.*, 2006), some of which are coincident with Cajal bodies (EJW, WFM unpublished). Myc-tagged CPSF100 and CPSF73 did not specifically concentrate in the HLB. All known mammalian U7 snRNP proteins (Lsm10, Lsm11 and ZFP100) localize to Cajal bodies. The failure to find additional proteins in the HLB other than Symplekin, Lsm10, and Lsm11, suggests that we did not identify any U7 snRNP-specific proteins.

Depletion of the histone variant H2Av prevents localization of U7 snRNP to the HLB and histone pre-mRNA processing

Our screen unexpectedly identified the variant histone proteins H3.3 and H2Av, which are expressed from polyadenylated mRNAs whose synthesis is not replication-coupled. The role of H2Av in histone pre-mRNA processing was confirmed *in vivo* by transgenic reporter gene expression in *H2Av* null mutant larvae (Fig. 3-5A). Northern blot analysis demonstrated read-through of both endogenous histone H3 and H2A mRNA (Fig. 5B). Loss of H2Av expression might cause misprocessing of histone pre-mRNAs because of increased histone gene transcription, which itself might reduce processing efficiency. This is unlikely as the RNAi-mediated depletion of H2Av also results in misprocessing of a strong actin promoter-driven reporter construct (Fig. 3-5C).

To determine how H2Av contributes to histone pre-mRNA processing, we assessed the localization and activity of processing factors in H2Av-depleted cells. In wild type salivary gland nuclei, Lsm11 antibodies detect the HLB as a prominent focus adjacent to the

chromocenter, visualized by HP1 staining (Fig. 3-5D). In contrast, there was no detectable Lsm11 focus in *H2Av* mutant salivary gland nuclei (Fig. 3-5D). To determine whether the HLB had assembled, Mpm-2 monoclonal antibodies, which stain a phosphoprotein present in the HLB during S-phase (White *et al.*, 2007), were used to analyze both brains from *H2Av* mutant larvae (Fig. 3-5E) and cultured cells with H2Av knocked down (Fig. 3-4D). All cells in the wild-type larval brain contained an HLB stained by Lsm11, which co-localized with Mpm-2 staining in cells positive for Mpm-2. In contrast, many cells in *H2Av* mutant larval brains contained foci stained by Mpm-2 and lacked foci stained by Lsm11 (Fig. 3-5E, bottom). Similarly, numerous H2Av-depeleted D.Mel-2 cells contained Mpm-2 foci but not Lsm11 foci (Fig. 3-5F). Thus, H2Av depletion results in a loss of Lsm11 from the HLB.

Western blot analysis of lysates from H2Av dsRNA treated cells demonstrated that, in addition to a reduction in H2Av protein levels, there was also a slight decrease in the level of Lsm11 protein (Fig. 3-5G, top three panels). There is a similar reduction in the level of U7snRNA (Fig. 3-5G bottom panel). The Lsm11 mRNA levels were unchanged in H2Av depleted cells relative to a control mRNA (Fig. 3-5G lower panels). Therefore, the misprocessing seen in *H2AV* null flies could result from either mislocalization of the processing machinery or a consequence of reduced levels of U7 snRNP. To address this question, we prepared nuclear extracts from control cells (PTB dsRNA), and cells treated with H2Av, Lsm11, or SLBP dsRNA and asked if they were competent to process a synthetic histone pre-mRNA substrate. We analyzed processing after a 1hr incubation of a labeled histone pre-mRNA substrate in equal amounts of the various nuclear extracts (Fig. 3-5H). Depleting Lsm11 or SLBP greatly reduced processing (Fig. 3-5H). Mixing the extracts from Lsm11 and SLBP depleted cells restored activity (Fig. 3-5H, lane 5), indicating that the

knockdown removed only the expected component required for processing. Strikingly, extracts from cells with H2Av knocked down had the same activity as control extracts (Fig. 3-5H, lanes 1 vs. 3), and could rescue processing when mixed 1:1 with either SLBP- or Lsm11-depleted extracts (lanes 6 and 7). The failure to accumulate U7 snRNP at the HLB resulted in a slight decrease in the overall level of U7snRNP that was not sufficient to reduce the processing activity in H2Av-depleted cells. More importantly, these data suggest- that the defect in processing in *H2Av* mutants *in vivo* is a result of a failure to localize U7 snRNP to the HLB, and not a defect in any processing factor.

Discussion

Here we report the use of a novel reporter assay to identify factors that participate in histone pre-mRNA processing. The assay was based on our observation that mutation of either the *Slbp* or *U7* snRNA genes in *Drosophila* results in expression of incorrectly processed histone mRNAs (Godfrey *et al.*, 2006; Sullivan *et al.*, 2001). These misprocessed mRNAs are polyadenylated due to transcriptional read-through and subsequent usage of canonical polyadenylation signals downstream of the HDE. We postulated that read-through to a downstream polyadenylation site would be a general phenotype associated with reduced expression of any factor required for histone pre-mRNA processing, and designed a mini-gene capable of reporting histone pre-mRNA misprocessing.

Using the histone pre-mRNA processing reporter to screen a genome-wide *Drosophila* RNAi library, we discovered 24 factors that play some role in histone pre-mRNA processing. We identified 5 previously known components of the histone pre-mRNA processing machinery (SLBP, Lsm11, Lsm10, Symplekin, and CPSF73). Among the other 19 genes are 2 cleavage/polyadenylation factors (CPSF100, Fip1), 4 proteins involved in chromatin structure and assembly (H2a.V, H3.3A, H3.3B, Asf1), 4 zinc finger proteins (MBD-R2, CG17361, Lola, CG9684), and 4 known/putative signaling molecules (RACK1, CG8866, CKIIα-iI, Cdk2). Many of these proteins (all of which have potential mammalian orthologues) are largely unstudied, except for the mammalian SR protein 9G8, implicated in alternative splicing (Cramer *et al.*, 2001) and histone mRNA nuclear export (Huang *et al.*, 2003); the cyclin-dependent kinase Cdk2, essential for progression through S-phase

(Knoblich *et al.*, 1994), and NELF-E, a component of the negative transcription elongation factor (Wu *et al.*, 2005). We did not identify proteins required for U7 snRNP biosynthesis, such as SMN, snRNA transcription factors, or Integrator factors required for snRNA 3' end formation (Baillat *et al.*, 2005), suggesting that the screen detected factors directly involved in histone pre-mRNA processing. None of these dsRNAs had a large effect on cell growth. In contrast, dsRNAs against spliceosomal Sm proteins, some of which are also components of U7 snRNP, all caused cell death, accounting for our failure to identify these proteins in the screen.

Polyadenylation factors and histone pre-mRNA processing

Biochemical fractionation of histone pre-mRNA processing factors from human cells suggested many of the proteins involved in cleavage and polyadenylation might be involved in histone pre-mRNA processing (Kolev and Steitz, 2005). However, only four of these, CPSF73, CPSF100, Symplekin, and Fip1, were identified in our genome-wide screen. CPSF73 (Dominski *et al.*, 2005) and Symplekin (Kolev and Steitz, 2005) were previously identified as histone pre-mRNA processing factors, and CPSF100 interacts with CPSF73. CPSF73 and CPSF100 each contain a nucleolytic β-lactamase domain, although this domain may be inactive in CPSF100 (Mandel *et al.*, 2006). Based upon crosslinking experiments, it appears likely that CPSF73 is the nuclease responsible for cleaving the histone pre-mRNA to form mature histone mRNA (Dominski *et al.*, 2005). Symplekin, first identified as a component of tight junctions (Keon *et al.*, 1996), likely acts as a scaffold for assembly of other processing factors, while the role of Fip1 in processing is less clear.

The role of histone variants in histone pre-mRNA processing

From the screen, we identified three variant histone proteins, H2Av, H3.3A, and H3.3B, suggesting that these proteins have a role in processing replication-dependent histone pre-mRNAs. H3.3 is assembled into chromatin preferentially at active genes, and can be incorporated in the absence of DNA replication (Ahmad and Henikoff, 2002). *Drosophila* H2Av is a functional orthologue of both human H2A.X and H2A.Z and plays an important role in defining the boundary between euchromatin and heterochromatin and in the DNA damage response (Swaminathan *et al.*, 2005). H2Av is present throughout the genome, including at the histone locus (Swaminathan *et al.*, 2005; HRS and RJB, unpublished). We find that depletion of H2Av in flies results in a failure to accumulate Lsm11 at the HLB, but that nuclear extract from cells RNAi depleted of H2Av exhibit no decrease in processing efficiency relative to that of control extract. Further, mixing H2Av extract with extract from SLBP or Lsm11-depleted extract, which are inactive for processing, restored processing activity. Together these data suggest that H2Av has some role in recruitment of U7 snRNP to the site of processing rather than a direct effect on levels of essential processing factors.

Conclusions

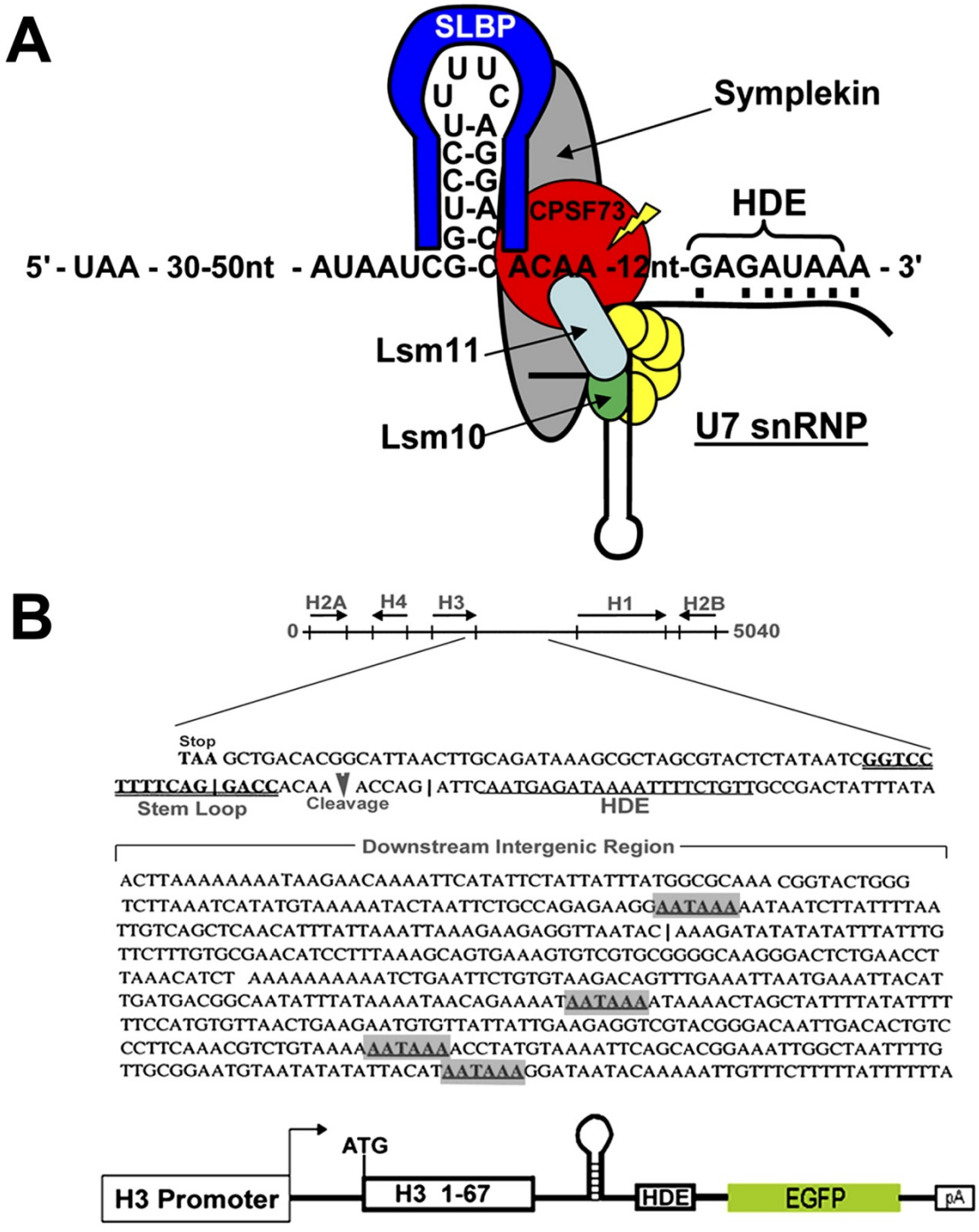
Here we present *in vivo* evidence that a subset of proteins involved in mRNA polyadenylation are also involved in the production of histone mRNAs, which are not polyadenylated, thus demonstrating a remarkable conservation in the machinery needed to generate different mRNA 3' ends in animal cells. We postulate that the histone pre-mRNA cleavage factor contains CPSF73/CPSF100, Symplekin and Fip1 and other polypeptides that may be among the uncharacterized proteins identified in the screen. Of these, only

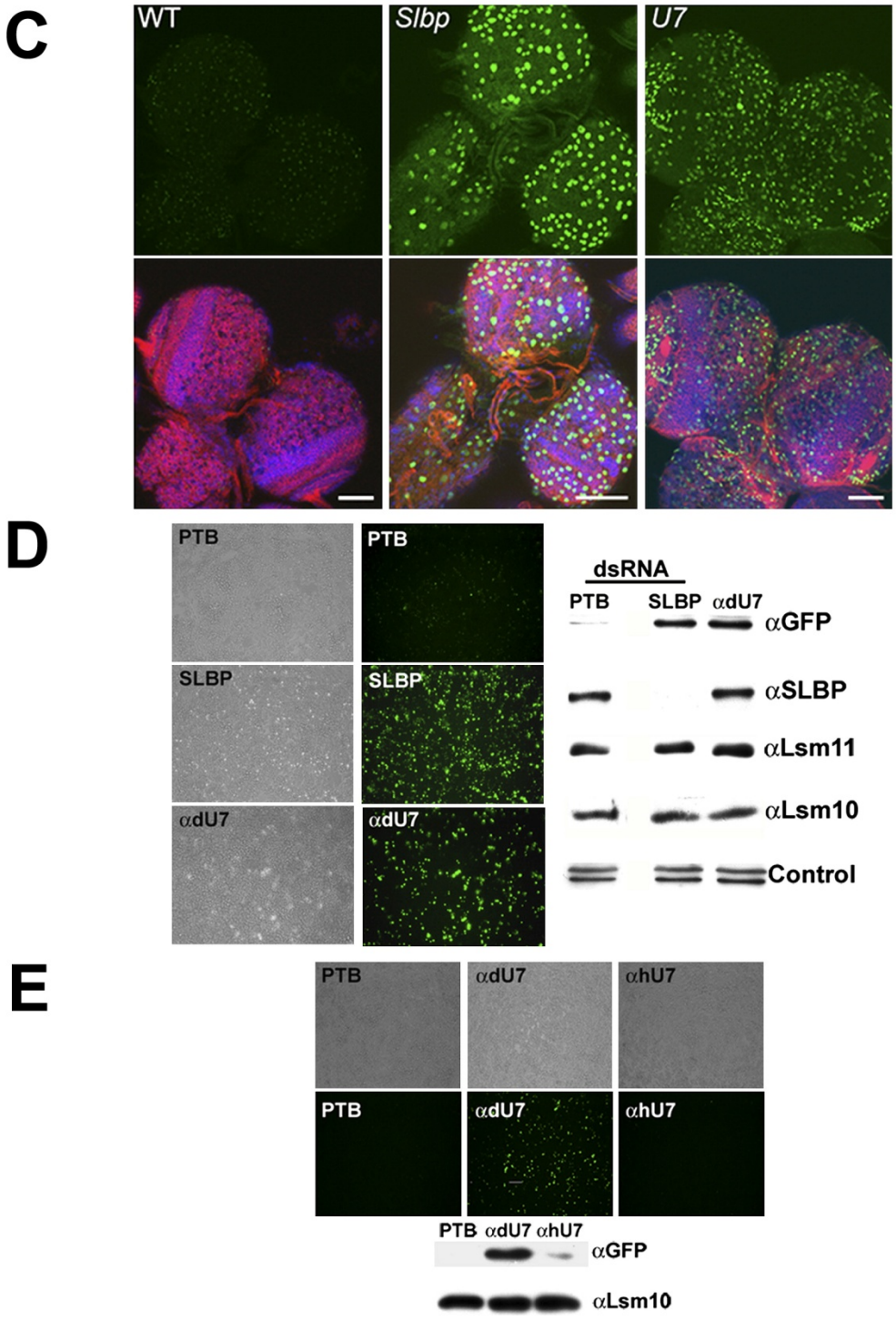
Symplekin may concentrate in the HLB, consistent with its proposed role in organizing the active cleavage factor (Kolev and Steitz, 2005).

Since we did not identify known factors in snRNA or snRNP biosynthesis, many of the proteins identified in our screen are likely directly involved in histone pre-mRNA processing. In addition, we identified factors that may regulate histone pre-mRNA processing, such as Cdk2, or serve to couple transcription and processing *in vivo*, such as NELF-E, recently shown to be required for efficient histone pre-mRNA processing in mammalian cells (Narita *et al.*, 2007), and H2Av, necessary for concentration of the U7 snRNP in the HLB.

This latter result suggests that cells balance assembly of variant and canonical histones by regulating pre-mRNA processing of the canonical histones. Cells deficient in H2Av do not assemble the HLB properly, and the U7 snRNP particle is not concentrated in the HLB, although it is active. The failure to properly localize the U7 snRNP to the HLB results in inefficient processing of the histone mRNA *in vivo*. Future experiments will help determine whether the failure to localize U7 snRNP is due to a specific defect in the amount of H2Av at the histone locus, or to a general H2Av deficiency.

Figure 3-1. A reporter for *Drosophila* histone pre-mRNA processing. (A) Schematic of *Drosophila* histone pre-mRNA processing machinery. (B) Diagram of *Drosophila* histone gene cluster showing the H3-H1 intergenic region containing the cryptic cleavage and polyadenylation sites (gray boxes) (Lanzotti *et al*., 2002), and a schematic of the reporter. (C) Confocal images of third instar larval brains containing the transgenic reporter and stained with anti-GFP (green), anti-phosphotyrosine (for cell cortex labeling; red) and DAPI (blue). Scale bar = 50 µm. (D) Left: Brightfield (left) and fluorescence images (right) of D.Mel-2 cells treated with dsRNA to PTB (control) or SLBP or with the αU7 oligonucleotide. Right: Corresponding Western blot analysis. (E) Bright field and fluorescence images and corresponding western blot analysis of cells treated with PTB dsRNA or a 2'O-CH_3 oligonucleotide targeting *Drosophila* or human U7 snRNA. (F) S1 nuclease protection assay (schematic at bottom) of endogenous H2A mRNA (lanes 2-5) or the reporter mRNA (lanes 6,7) isolated from D.Mel-2 cells treated with PTB dsRNA (lanes 2, 4, 6) or a dsRNA that activated read through (lanes 3,5,7). Note that lanes 4 and 5 are a darker exposure of lanes 2 and 3. Lane 1 contains input probe. Abbreviations: GFP = green fluorescent protein; PTB = pyrimidine tract binding protein; SLBP = stem loop binding protein. (panel C courtesy of Ashley Godfrey)


A
SLBP
U U
U C
U-A
C-G
C-G
U-A
G-C
Symplekin
CPSF73
HDE
5' - UAA - 30-50nt - AUAAUCG-C ACAA -12nt-GAGAUAAA - 3'
Lsm11
Lsm10
U7 snRNP
B
H2A
H4
H3
H1
H2B
0
5040
Stop
TAA GCTGACACGGCATTAACTTGCAGATAAAGCGCTAGCGTACTCTATAATCGGTCC
TTTTCAG | GACCACAA ACCAG | ATTCAATGAGATAAAATTTTCTGTTGCCGACTATTTATA
Stem Loop
Cleavage
HDE
Downstream Intergenic Region
ACTTAAAAAAAATAAGAACAAAATTCATATTCTATTATTTATGGCGCAAA CGGTACTGGG
TCTTAAATCATATGTAAAAATACTAATTCTGCCAGAGAAGGAATAAAAATAATCTTATTTTAA
TTGTCAGCTCAACATTTATTAAATTAAAGAAGAGGTTAATAC | AAAGATATATATATTTATTTG
TTCTTTGTGCGAACATCCTTTAAAGCAGTGAAAGTGTCGTGCGGGGCAAGGGACTCTGAACCT
TAAACATCT AAAAAAAAAATCTGAATTCTGTGTAAGACAGTTTGAAATTAATGAAATTACAT
TGATGACGGCAATATTTATAAAATAACAGAAAATAATAAAATAAAACTAGCTATTTTATATTTT
TTCCATGTGTTAACTGAAGAATGTGTTATTATTGAAGAGGTCGTACGGGACAATTGACACTGTC
CCTTCAAACGTCTGTAAAAAATAAAACCTATGTAAAATTCAGCACGGAAATTGGCTAATTTTG
TTGCGGAATGTAATATATATTACATAATAAAGGATAATACAAAAATTGTTTCTTTTTATTTTTTA
ATG
H3 Promoter
H3 1-67
HDE
EGFP
pA


C
WT
Slbp
U7
D
PTB
PTB
SLBP
SLBP
αdU7
αdU7
dsRNA
PTB
SLBP
αdU7
αGFP
αSLBP
αLsm11
αLsm10
Control
E
PTB
αdU7
αhU7
PTB
αdU7
αhU7
PTB
αdU7
αhU7
αGFP
αLsm10

F

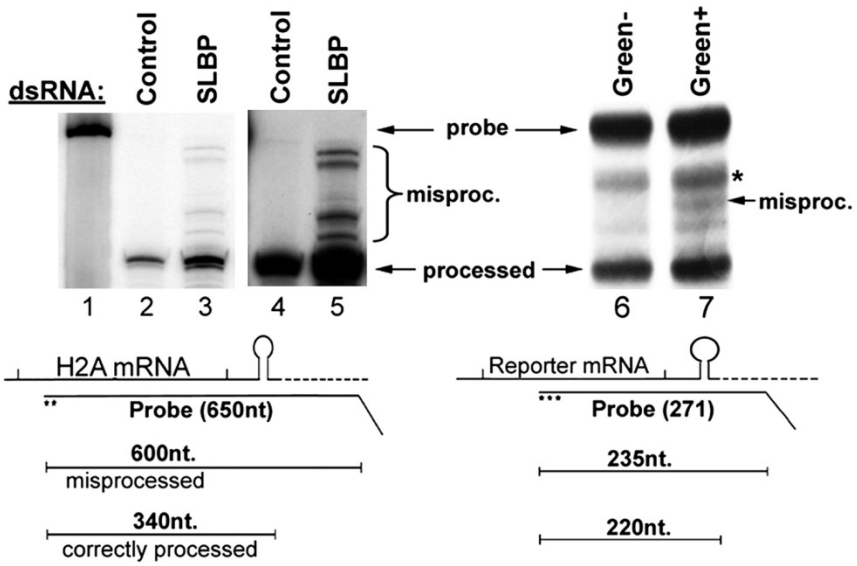

dsRNA:
Control
SLBP
Control
SLBP
Green-
Green+
probe
misproc.
processed
*
misproc.
1
2
3
4
5
6
7
H2A mRNA
Probe (650nt)
600nt.
misprocessed
340nt.
correctly processed
Reporter mRNA
Probe (271)
235nt.
220nt.

Figure 3-2. Example plate from the screen. (A) Fluorescence images of D.Mel-2 cells grown on plate 18 of dsRNA library and transfected with reporter. Cells in column 1, row 2 were treated with αU7 oligonucleotide. The inset is a higher magnification view of 16 wells, two of which contain dsRNA for Lsm10 or Lsm11, from replicate experiments. (B) Western blot analysis of lysates from D.Mel-2 cells treated with dsRNA targeting PTB (control), Lsm10, or Lsm11. Abbreviations: PTB = pyrimidine tract binding protein; Lsm = Sm-like.

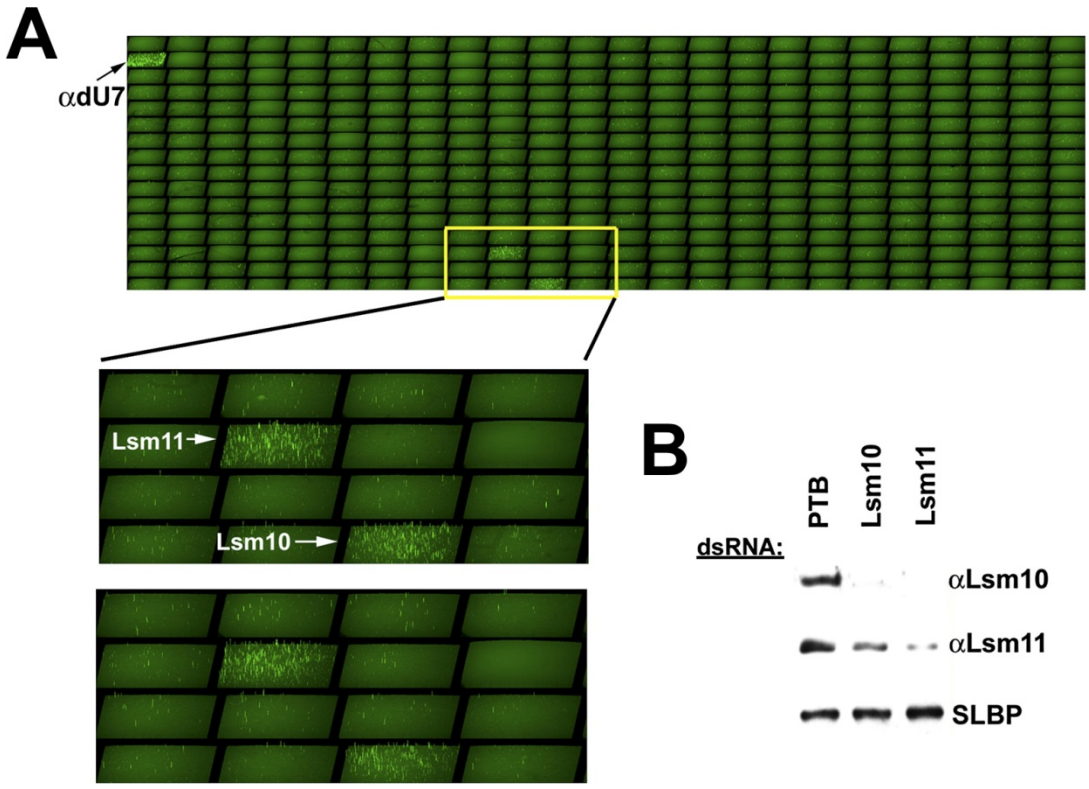

A
αdU7
Lsm11
Lsm10
B
PTB
Lsm10
Lsm11
dsRNA:
αLsm10
αLsm11
SLBP

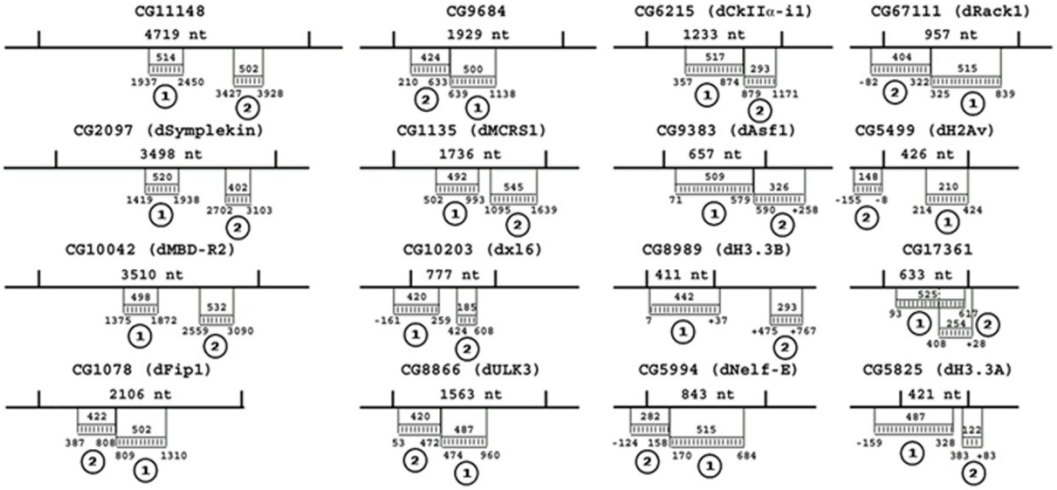
Figure 3-3. Results of the screen. (A) Table of hits from the genome-wide screen categorized numerically by qualitative strength of reporter signal and by color for relevant domains or putative/known functions. (B) Schematic of nucleotide positions of all first and second site targeting with dsRNA. The first targeting sites (labeled 1) were chosen by the DRSC and we designed the second sites not to overlap with the first, with the exception of CG17361 (due to its small size). (C) Fluorescence microscopy of first site targeting relative to negative control dsRNA (PTB) and positive control dsRNA (CPSF73). The lower panels are brightfield images. (D) Fluorescence microscopy of second site targeting relative to negative control dsRNA (PTB) and positive control dsRNA (CPSF73). Note that second site targeting was performed in serum-dependent S2 cells, and that the amount of fluorescence in the CPSF73 control (with site 1 oligonucleotide) was lower than in panel C. (E) Lysates from D.Mel-2 cells treated with dsRNA targeting genes identified in the screen were analyzed by Western Blot. Lysates were probed with α-GFP antibodies to detect the amount of readthrough from the reporter. The asterisk indicates a background band that served as a loading control. Abbreviations: DRSC = Drosophila RNAi Screening Center; dsRNA = double stranded ribonucleic acid; CPSF = cleavage and polyadenylation specificity factor; S2 = Schneider 2; GFP = green fluorescent protein.

A

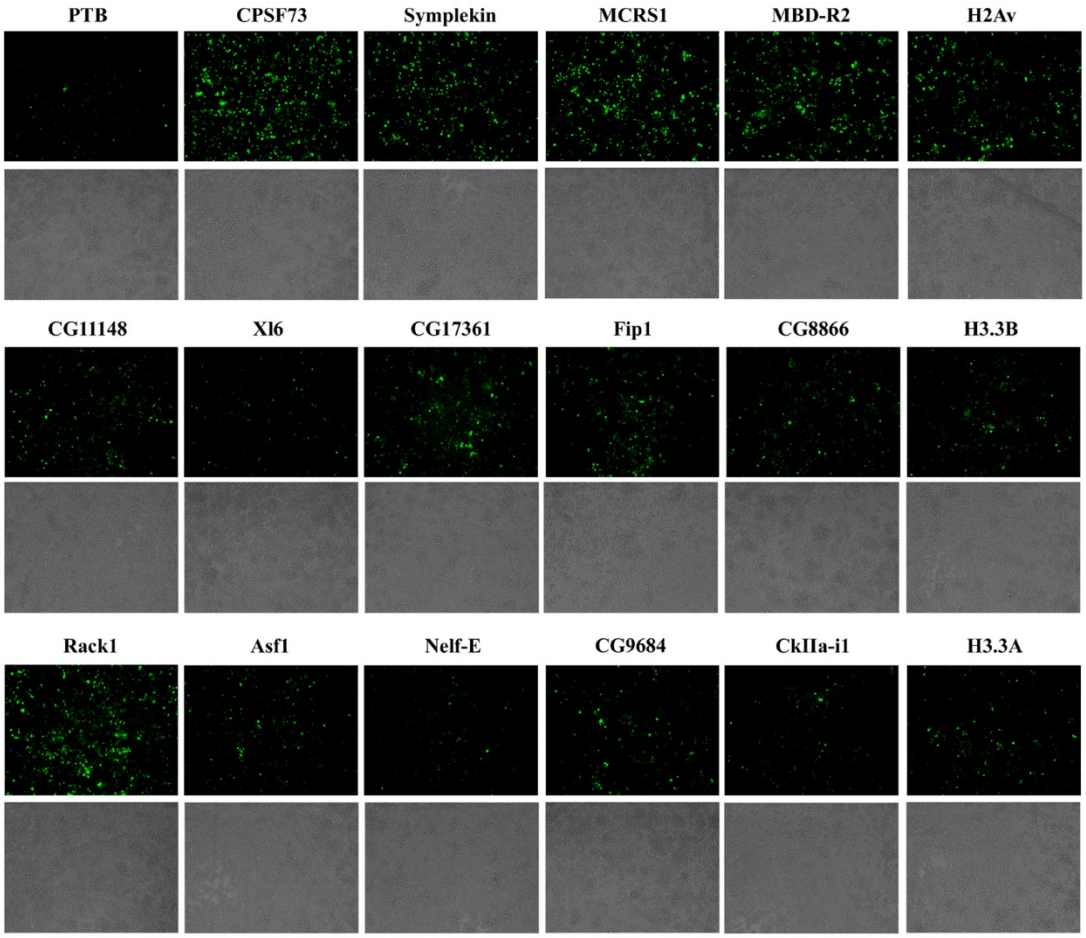
5 (Very Strong)	4 (Strong)	3 (medium)	2 (weak)
-SLBP	-His2Av	-CG8866	-CG9684
-Lsm10	-CG17361	-Lola	-CKIIα-I1
-Lsm11	-Fip1	-H3.3B	-H3.3A
-CPSF73		-CG11148	-XL6 (9G8)
-CPSF100		-Asf1	-CDC2C (CDK2)
-Symplekin		-Yeti	
-CG1135 (MCRS1)		-Nelf-E	
-MBD-R2			
-Rack1			

Zinc Finger proteins
Putative protein kinase
Histone proteins, chromatin assembly
Polyadenylation factors
U7 snRNP proteins
RNA metabolism

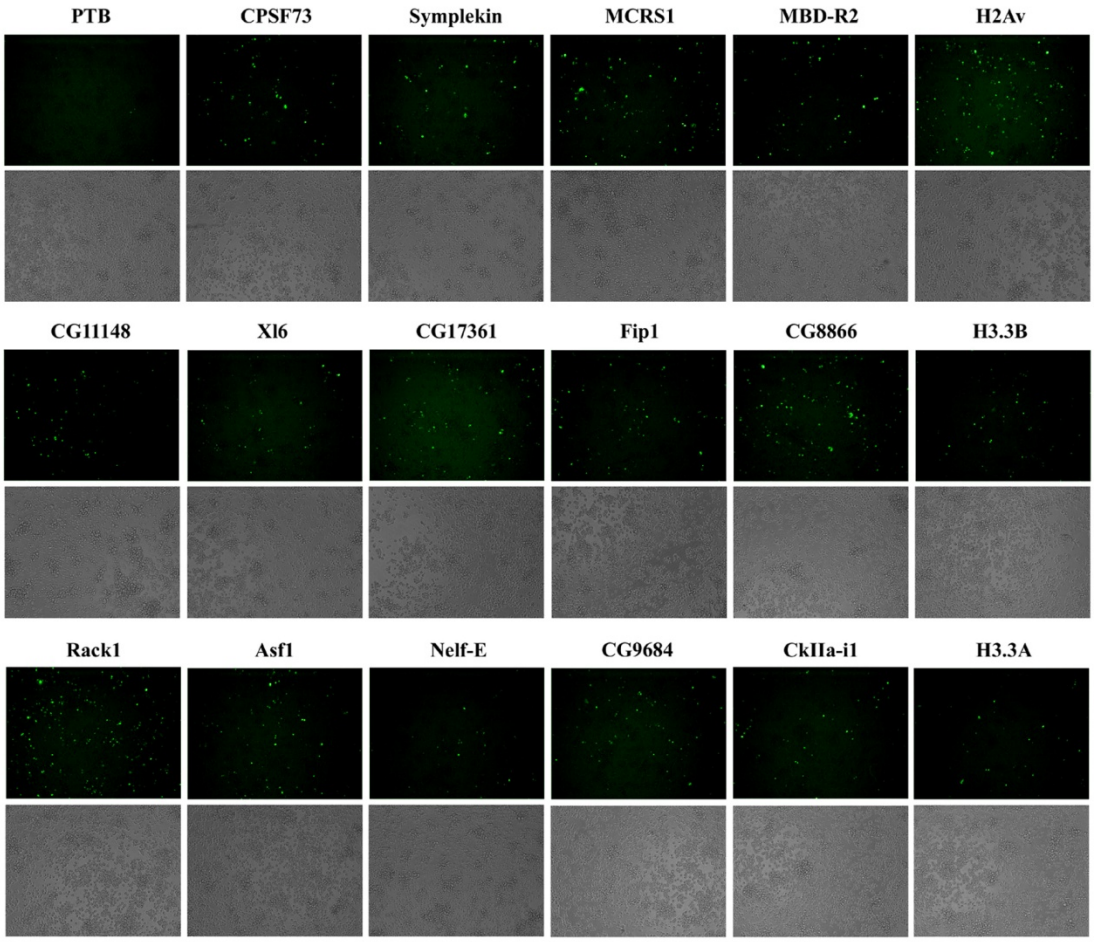
B



C


PTB
CPSF73
Symplekin
MCRS1
MBD-R2
H2Av
CG11148
Xl6
CG17361
Fip1
CG8866
H3.3B
Rack1
Asf1
Nelf-E
CG9684
CkIIa-i1
H3.3A

D


PTB
CPSF73
Symplekin
MCRS1
MBD-R2
H2Av
CG11148
Xl6
CG17361
Fip1
CG8866
H3.3B
Rack1
Asf1
Nelf-E
CG9684
CkIIa-i1
H3.3A

E

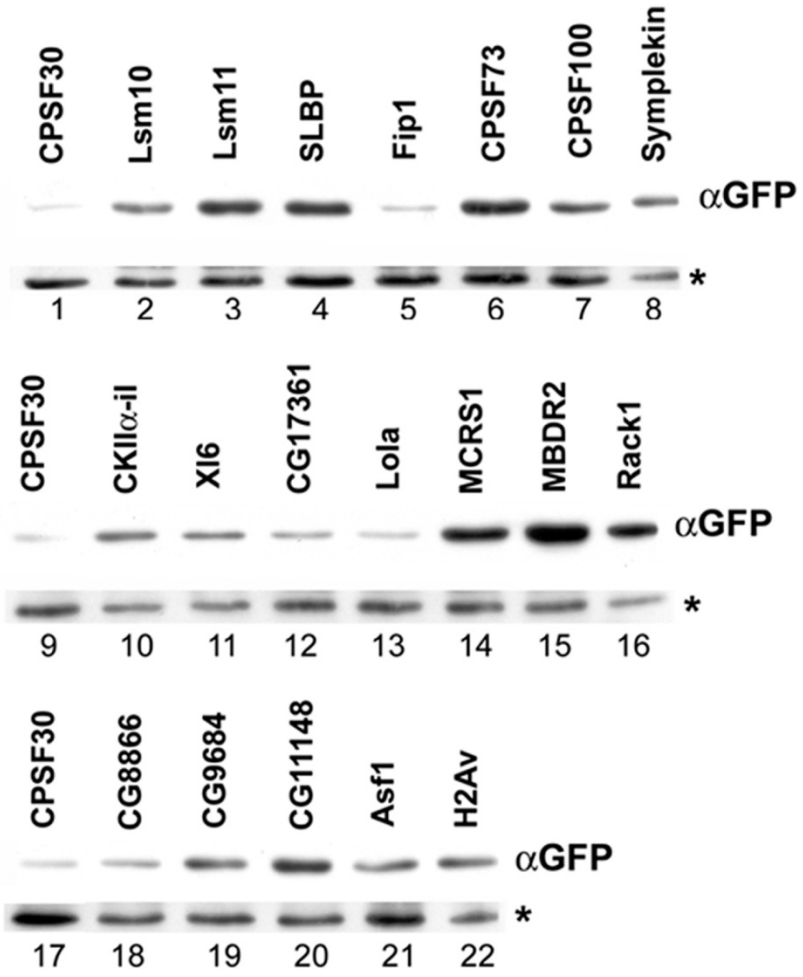
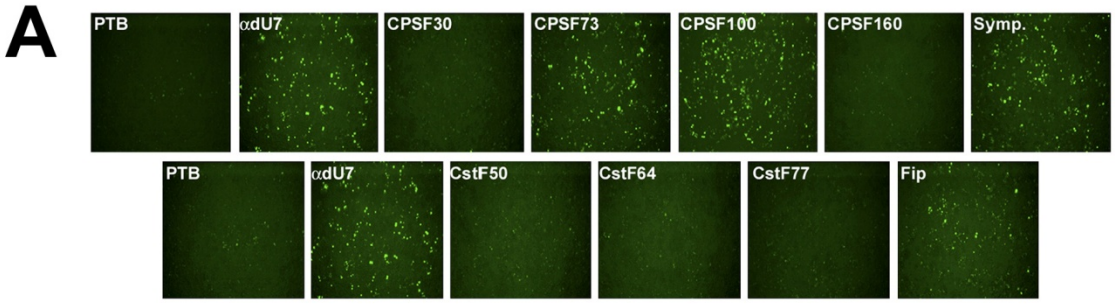
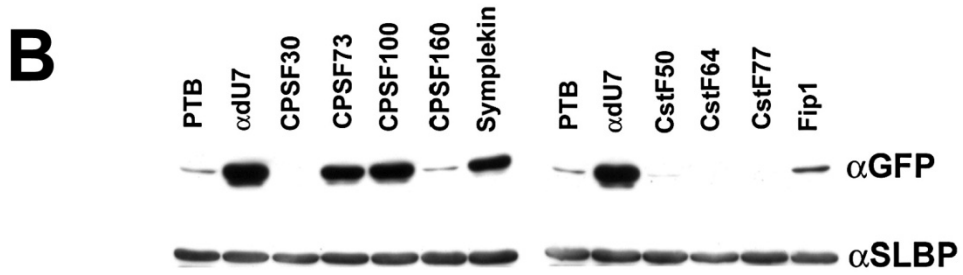
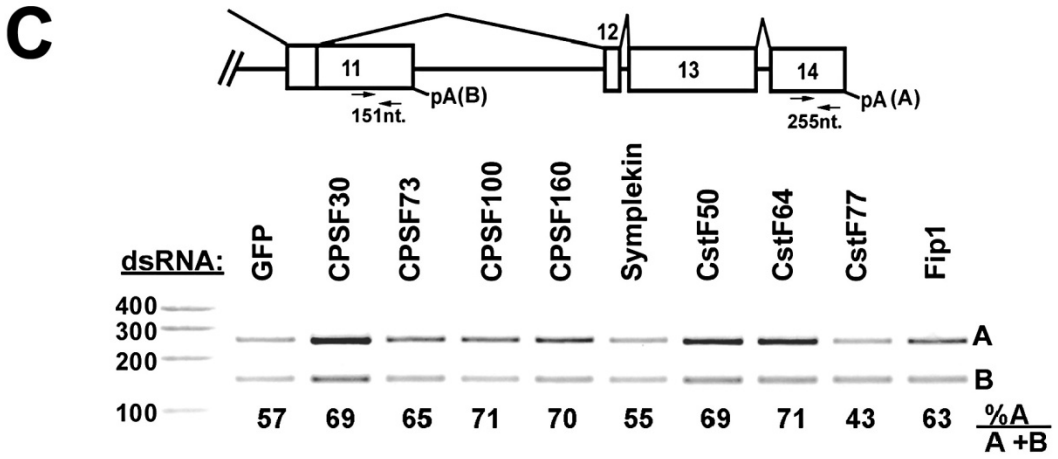

CPSF30
Lsm10
Lsm11
SLBP
Fip1
CPSF73
CPSF100
Symplekin
αGFP
*
1 2 3 4 5 6 7 8
CPSF30
CKIIα-il
Xl6
CG17361
Lola
MCRS1
MBDR2
Rack1
αGFP
*
9 10 11 12 13 14 15 16
CPSF30
CG8866
CG9684
CG11148
Asf1
H2Av
αGFP
*
17 18 19 20 21 22

Figure 3-4. A subset of cleavage and polyadenylation factors is required for histone pre-mRNA processing. (A) Fluorescence images of D.Mel-2 cells treated with dsRNA targeting the indicated genes and then transfected with reporter. (B) Western blot analysis of lysates from cells in panel A. (C) Top: Schematic of the 3' end of the *Drosophila domino* gene containing two alternative polyadenylation sites. Bottom: RT-PCR analysis of *domino* mRNA after treatment of D.Mel-2 cells with the indicated dsRNAs. Band A corresponds to use of the distal polyA site and band B to the proximal polyA site. PCR products were cloned to confirm identity and in all cases there were no products in the absence of RT. (D) Confocal images of D.Mel-2 cells transiently transfected with genes expressing the indicated myc-tagged proteins and stained with anti-Lsm11 (green) and anti-myc (red) antibodies. In merged images, yellow arrows indicate colocalization of a myc-tagged protein with Lsm11 and green arrows indicate HLB with Lsm11 only. Note that all cells contained HLB, not all of which could be visualized in the particular focal plane shown. Abbreviations: RT-PCR = reverse transcriptase polymerase chain reaction; HLB = histone locus body. (panels C and D courtesy of Eric Wagner)


A
PTB
αdU7
CPSF30
CPSF73
CPSF100
CPSF160
Symp.
PTB
αdU7
CstF50
CstF64
CstF77
Fip


B
PTB
αdU7
CPSF30
CPSF73
CPSF100
CPSF160
Symplekin
PTB
αdU7
CstF50
CstF64
CstF77
Fip1
αGFP
αSLBP


C
11
12
13
14
pA(B)
pA (A)
151nt.
255nt.
dsRNA:
GFP
CPSF30
CPSF73
CPSF100
CPSF160
Symplekin
CstF50
CstF64
CstF77
Fip1
400
300
200
100
A
B
57
69
65
71
70
55
69
71
43
63
%A
A +B

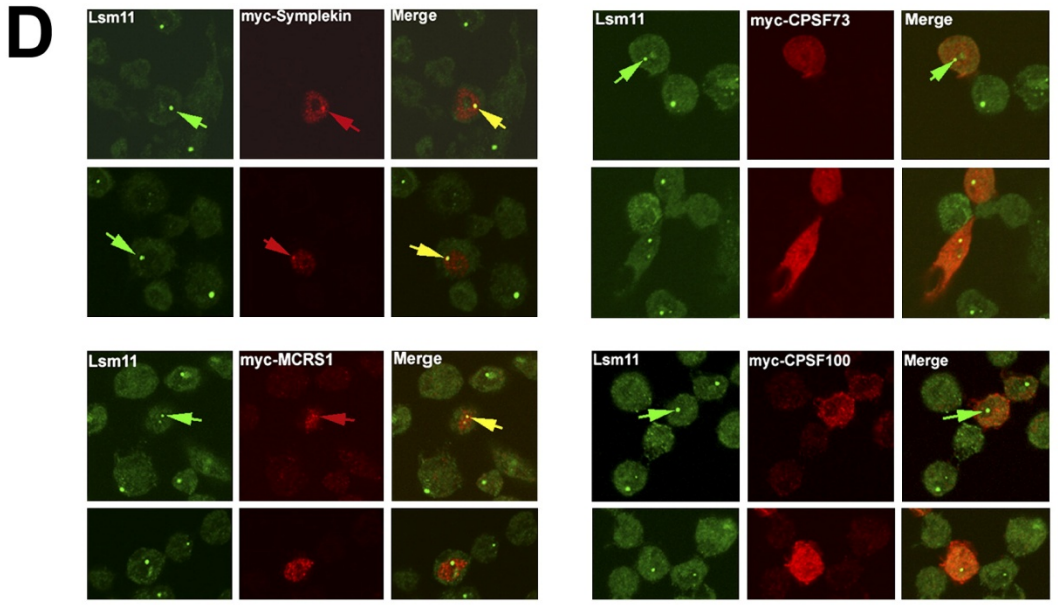
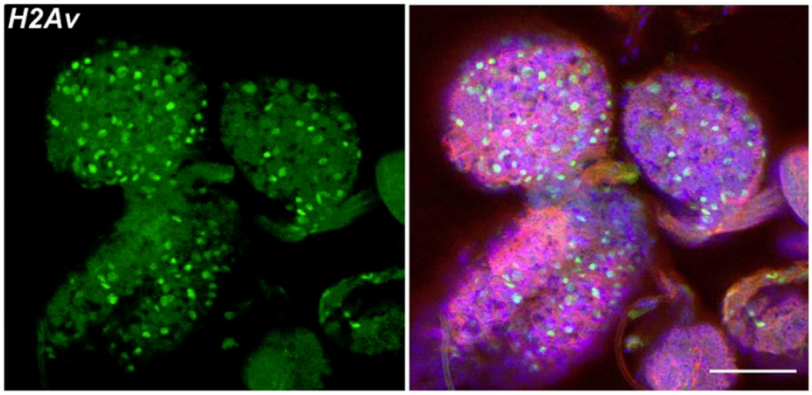

D
Lsm11
myc-Symplekin
Merge
Lsm11
myc-CPSF73
Merge
Lsm11
myc-MCRS1
Merge
Lsm11
myc-CPSF100
Merge

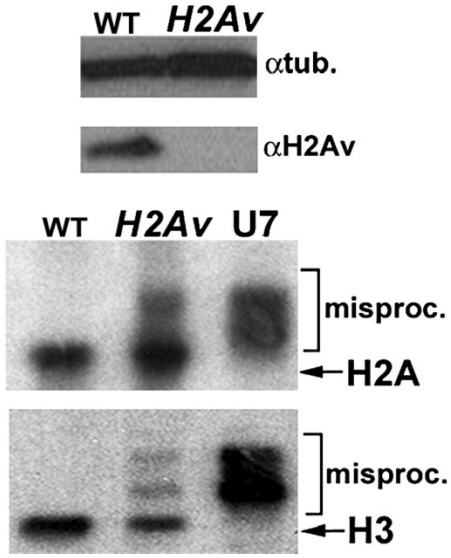
Figure 3-5. Knockdown of histone variants affect canonical histone pre-mRNA processing. (A) Confocal microscopy images of third instar larval brains from H2Av null mutant flies containing the reporter and stained with anti-GFP (green), anti-phosphotyrosine (for cell cortex labeling; red) and DAPI (blue). Scale bar = 50 µm. (B) Top: Western blot of H2Av from wild type and *H2Av* mutant 3rd instar larvae. Bottom: Northern blot of endogenous H2A mRNA (upper) and H3 mRNA (lower) from wild-type, *H2Av* null, and *U7* snRNA null mutant whole third instar larvae. (C) D.Mel-2 cells were treated with either PTB dsRNA (negative control), Lsm11 dsRNA (positive control), or with dsRNA targeting H2Av. Cells were transfected with a GFP-based processing reporter driven by either the Actin 5C promoter or the histone H3 promoter and visualized for GFP expression and by brightfield. (D) Confocal images of salivary gland nuclei isolated from either wild type or *H2Av* null third instar larvae stained for DNA (DAPI, blue), Lsm11 (green), and HP1 (pink). Scale bar = 20 µm. (E,F) Confocal images of a region of a third instar larval brain from wild type or *H2Av* mutant (E) and D.Mel-2 cells treated with control dsRNA (PTB) or with H2Av dsRNA (F) stained for Lsm11 (green) and Mpm2 (red). In the merged field, green arrows indicate HLB positive for Lsm11 and negative for Mpm-2 (cells not in S phase), yellow arrows indicate HLB containing both Lsm11 and Mpm-2 (cells in S phase), pink arrows indicate HLB with Mpm-2 and reduced levels of Lsm11, and red arrows indicate HLB in H2Av-depleted cells that only contain Mpm-2. Scale bar in panel E = 5 µm. Bottom panels in F are a higher magnification view of a different H2Av field. (G) D.Mel-2 cells treated with PTB dsRNA or H2Av dsRNA were analyzed by Western blot for H2Av and Lsm11 protein (top 3 panels). RNA prepared from the same cells served as a template for RT-PCR analysis for the indicated endogenous mRNAs (next 4 panels). Note that the CstF77 RT-PCR serves as the

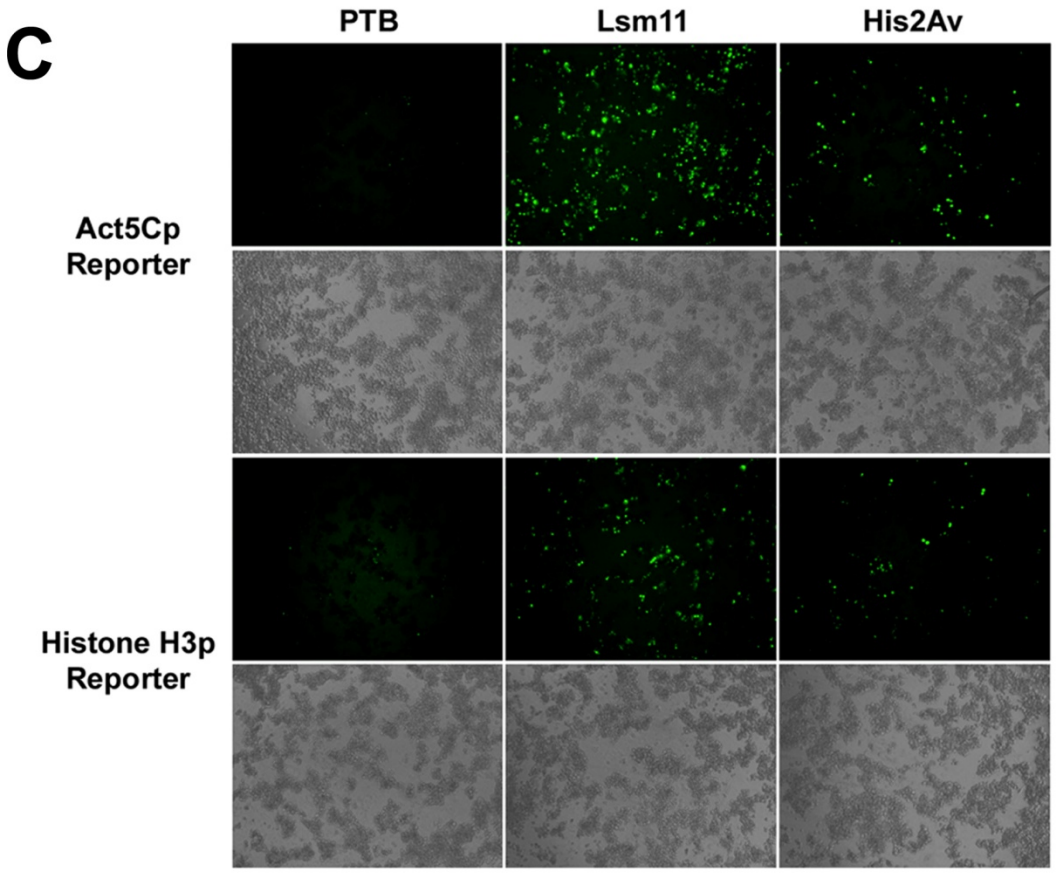
loading control for all of the RT-PCRs shown. The bottom panel is a Northern blot analysis of U7 snRNA in the same RNAs used in the RTPCR analysis. (H) In vitro processing of a labeled histone pre-mRNA incubated in nuclear extracts isolated from D.Mel-2 cells treated with the indicated dsRNA. S/L, H/S, and H/L indicate processing reactions in a 1:1 mixture of nuclear extract from SLBP-, Lsm11-, or H2Av-depleted cells (S, L, H, respectively). Abbreviations: H2Av = histone H2A variant; PTB = pyrimidine tract binding protein; Lsm = Sm-like; DAPI = 4',6-diamidino-2-phenylindole; HP1 = heterochromatin protein 1; MPM-2 = mitotic protein monoclonal #2; HLB = histone locus body. (panels A, D, and E courtesy of Harmoney Salzler, panels B, F, and H courtesy of Eric Wagner, panel G done in collaboration with Eric Wagner)

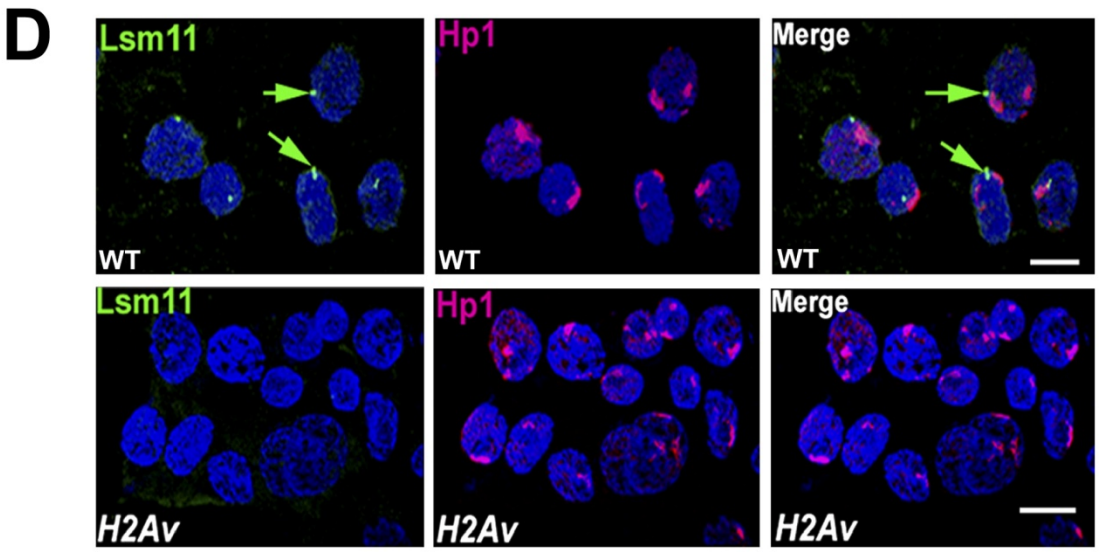
A

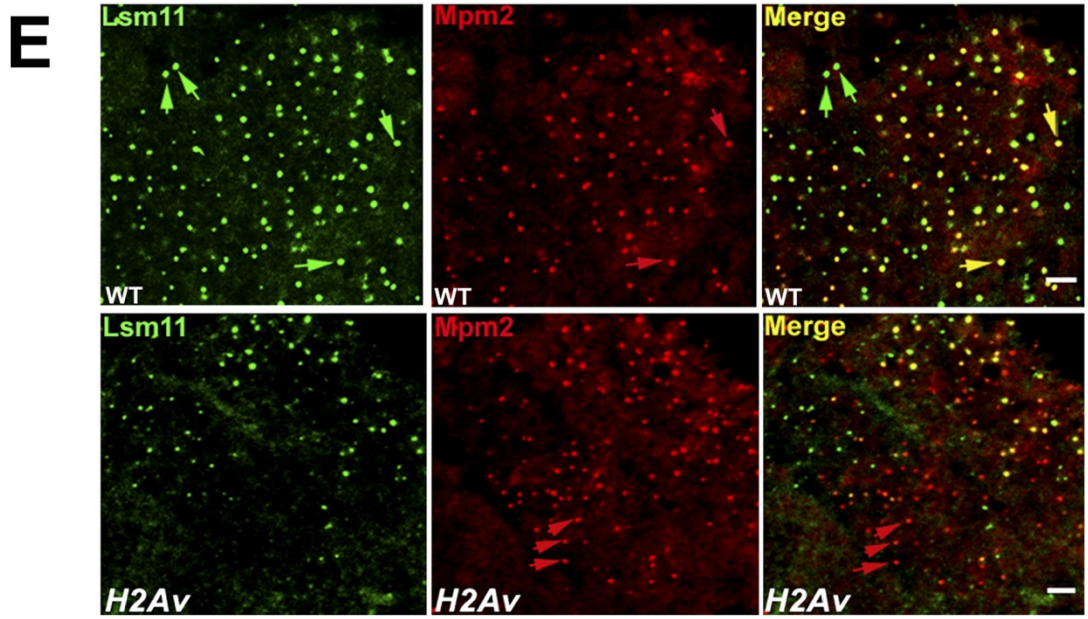


B

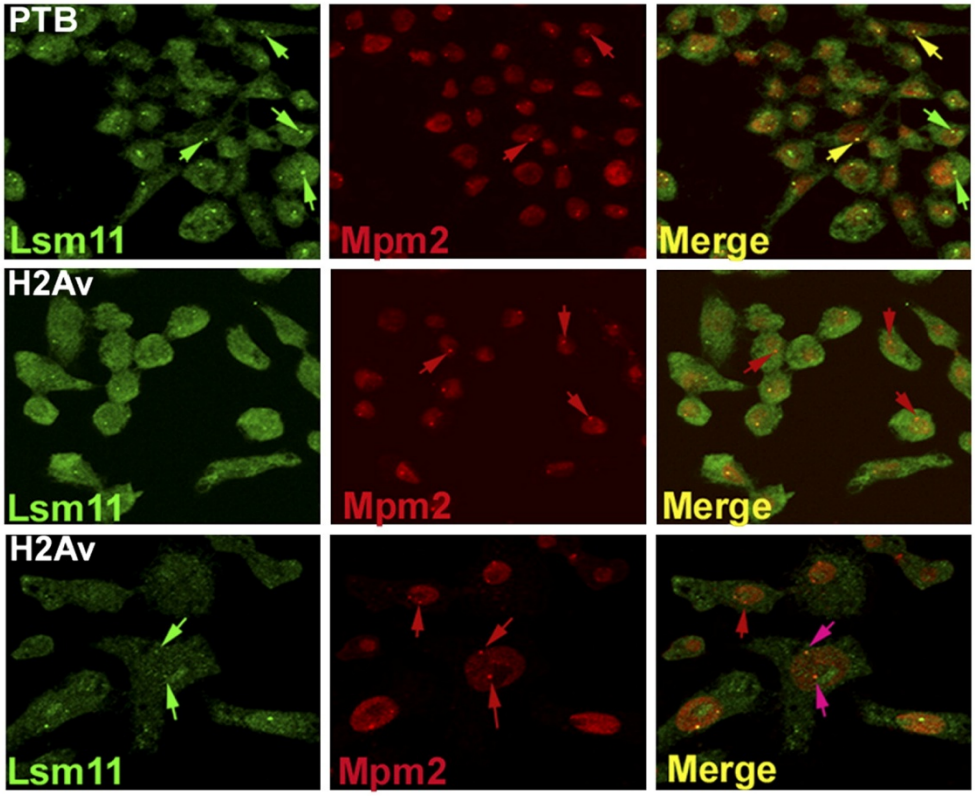



C
PTB
Lsm11
His2Av
Act5Cp
Reporter
Histone H3p
Reporter

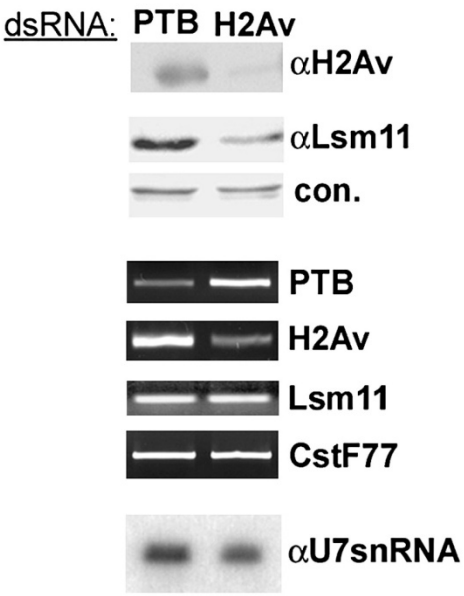

D
Lsm11
Hp1
Merge
WT
WT
WT
Lsm11
Hp1
Merge
H2Av
H2Av
H2Av

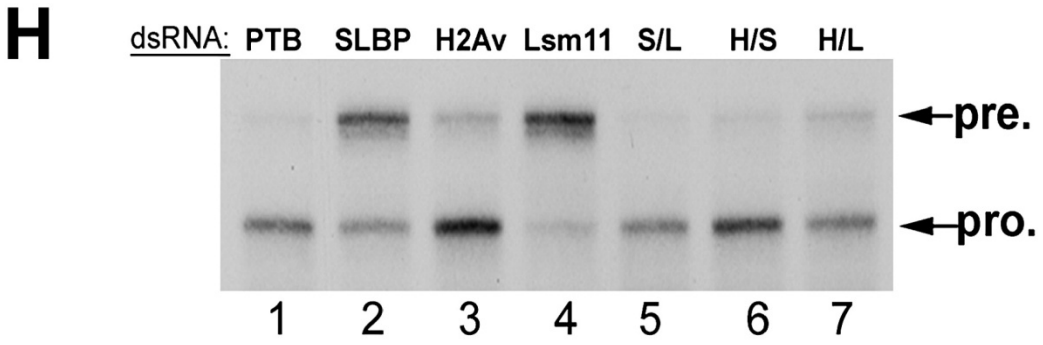

E
Lsm11
Mpm2
Merge
WT
WT
WT
Lsm11
Mpm2
Merge
H2Av
H2Av
H2Av

F


PTB
Lsm11
Mpm2
Merge
H2Av
Lsm11
Mpm2
Merge
H2Av
Lsm11
Mpm2
Merge

G


dsRNA:
PTB
H2Av
αH2Av
αLsm11
con.
PTB
H2Av
Lsm11
CstF77
αU7snRNA


H
dsRNA:
PTB
SLBP
H2Av
Lsm11
S/L
H/S
H/L
pre.
pro.
1
2
3
4
5
6
7

CHAPTER IV

CHARACTERIZATION OF THE INTERACTION BETWEEN FLASH AND LSM11 IN *DROSOPHILA*

Introduction

FLASH is a large (~2000 amino acid) protein originally identified as a protein involved in apoptosis, and later as a protein present in the mammalian histone locus bodies, where it was postulated to play a role in transcription of histone genes. Zbig Dominski recently identified FLASH as a histone pre-mRNA processing factor, and in this chapter I show that the critical biochemical function of a putative *Drosophila* FLASH orthologue is in histone pre-mRNA processing.

Fas-associated death domain-like IL-1-converting enzyme (FLICE) associated huge protein (FLASH), also known as caspase 8-associated protein 2 (CASP8AP2), is a multi-functional protein with roles in apoptosis, cell cycle regulation, transcription, and histone expression. This protein was first identified in a yeast-two-hybrid screen of a mouse T-cell cDNA library using two tandem-repeat death-effector domains (DED) of procaspase-8 as bait (Imai *et al.*, 1999). Further characterization revealed a DED-recruiting domain (DRD) at its C terminus between amino acids 1684-1785 that interacts with the DED domains of both caspase-8 (FLICE/MACH/Mch5) and FADD/MORT1, both involved in CD95/Fas-mediated apoptosis. Previous studies showed that caspase-8, FADD, and activated CD95 form a pro-apoptotic complex known as the death-inducing signaling complex (DISC). Based upon this work and the observation that FLASH only associates with CD95 when CD95 is in its activated form, FLASH was classified as a new component of DISC (Imai *et al.*, 1999). In

agreement with a cytoplasmic role for FLASH, a putative nuclear exclusion signal (NES) was documented in the mouse protein between amino acids 1124-1133 (Imai *et al.*, 1999). Interestingly, however, at least two putative nuclear localization signals (NLS), one from amino acids 1152-1169 and the other from 1813-1830, were also found, suggesting a more complex localization and functional profile for this protein (Imai *et al.*, 1999).

Although the early observations about FLASH all supported a primary role as a pro-apoptotic protein, this notion began to change rapidly as additional functions of FLASH began to emerge that pointed to roles distinct from its participation in promoting apoptosis. One of the first of these diverse roles to be described was participation by FLASH in the NFκB pathway though its interaction with TRAF2. This interaction, involving the region between amino acids 856-1191 in the human protein, is required for TNFα-induced NFκB activation (Choi *et al.*, 2001). Ultimately, this interaction results in the degradation of IκB, an inhibitor of NFκB, and accumulation of NFκB in the nucleus, which in turn allows transcription of NFκB responsive genes. NFκB activation is part of a protective pathway that leads to cell proliferation and survival, thereby offering an interesting counterpoint to the role of FLASH in pro-apoptotic pathways and adding further complexity to our understanding of this protein.

Despite these studies suggesting a purely cytoplasmic localization of FLASH, researchers began to describe functions of the protein that required nuclear localization. One of these involves translocation of FLASH into the nucleus and its participation in the regulation of glucocorticoid target genes through a suppressive effect on members of the p160 family of steroid hormone receptor coactivators (Kino *et al.*, 2003; Kino *et al.*, 2004; Obradovic *et al.*, 2004). Specifically, FLASH is able to interact with glucocorticoid

receptor-interacting protein 1 (GRIP1), and thyroid hormone receptor activator molecule 1 (TRAM1) through the nuclear-receptor binding domains of these proteins, thereby inhibiting the interaction between these proteins and the glucocorticoid receptor (GR) and preventing GR transcriptional activity (Kino *et al.*, 2004). This suppressive interaction involves the region of FLASH between amino acids 1709-1982, just downstream of the DRD (Kino *et al.*, 2004).

Shortly after the initial glucocorticoid receptor studies, other researchers discovered a link between FLASH and histone gene expression. In these studies, FLASH was shown to localize to Cajal bodies (Barcaroli *et al.*, 2006a,b; Kiriyama *et al.*, 2009; Yang *et al.*, 2009) and, even more specifically, in the same bodies in which the histone transcriptional regulator $p220^{NPAT}$ (nuclease protein, ataxia-telangiectasia locus) is found. Its localization to the histone locus was further indicated by chromatin immunoprecipitation experiments in which FLASH coprecipitated with histone gene chromatin (Barcaroli *et al.*, 2006a). Significantly, both the N-terminus and C-terminus of human FLASH are required for the formation of NPAT foci in the nucleus, suggesting that FLASH plays a critical role in organizing the HLB (Kiriyama *et al.*, 2009). These experiments demonstrated that FLASH is present on histone genes, leading Barcaroli and colleagues to postulate a role for FLASH in histone gene transcription (Barcaroli *et al.*, 2006b). The role of FLASH in histone gene transcription was further inferred from the observation that overexpression of FLASH results in increased transcription of a luciferase reporter construct driven by a histone H4 promoter (Barcaroli *et al.*, 2006b). In addition, FLASH was shown to be required, not only for NPAT body integrity, but for the integrity of the Cajal body, with loss of FLASH resulting in Cajal body

fragmentation, delocalization of NPAT from the histone locus, reduced levels of histone mRNA and protein, and arrest of cells in S-phase (Barcaroli *et al.*, 2006a,b).

Together, these studies point to a diversity of functions for FLASH and the importance of this protein's participation in several critical pathways in the cell. In addition to the functions described previously, FLASH has an additional role in RNA metabolism. Here, I discuss the newly-discovered role of the *Drosophila* orthologue of FLASH in histone pre-mRNA processing and describe characterization of key regions within the N-terminus of this protein that are required for its activity and localization to the *Drosophila* histone locus body.

Materials and Methods

Creation of Reporter Constructs

The His_{proc} reporter was created by subcloning the histone H3 promoter, partial open reading frame, downstream sequence, and an EGFP coding sequence into a promoterless pIZ/V5/His vector backbone (Invitrogen). Briefly, the histone H3 5'UTR and amino acids 1-67 were amplified by PCR using forward primer 5' GGCCGAATTCCGACAAAAAC CCGAGAGAGTAC 3' and reverse primer 5' GGCCGGTACCTTAGGCAGCTTG CGGATTAGAAGC 3' and subcloned into pEGFPN1 (Clontech, Palo Alto CA) using EcoRI and KpnI. The 3' end of the H3 gene starting immediately after the stop codon and continuing until 18 nt downstream of the HDE was amplified using forward primer 5' GGCCGGTACCACTTGCAGAT AAAGCGCTAGCG 3' and reverse primer 5' GGCCGGATCCTTGTTATAAATAG TCGGCAACA GAAAATTTTTTCTC 3' followed by ligation to the 5' product using KpnI and BamHI. The resulting construct contained the H3 promoter, amino acids 1-67 of histone H3 ORF, and H3 3' end containing a portion of the downstream intergenic region upstream of an EGFP ORF. Note that the H3 ORF is in frame with GFP and that we created a single mutation (U to A) within the HDE in order to disrupt a stop codon (red box in Fig. S1). The OpIE2 promoter was removed from pIZ-V5/His (Invitrogen, Carlsbad CA) vector by inverse PCR using forward primer 5' GGCCGCTAGCACAG CATCTGTTCGAATTTA 3' and reverse primer 5' GGCCGCTAGCAGACAT GATAAGATACATTGATGA 3' followed by digestion with NheI and religation. The reporter was then subcloned into the promoterless pIZ/V5/His vector using forward primer 5'

GGCCGAATTCCGACAAAAACCCGAGAGAG TAC 3' and reverse primer 5' GGCCTCTAGATTACTTGTACAGCTCGTCCAT GCC 3' followed by digestion with EcoRI and XbaI (NEB) and subsequent ligation to form the His_{proc} construct. Note that the polyA site used by the GFP ORF is the insect OpIE2 polyadenylation sequences 3' of the XbaI site in the pIZ/V5/His vector.

To create the Act_{proc} construct, the *Drosophila* Actin 5C promoter was amplified from genomic DNA by PCR using forward primer 5' GGCCAAGCTTTATGTATGTTTTGGCATACAATGAGTAGTTGG 3' and reverse primer 5' CCGGGAATTCGTGTCGGGAGGAGTATCCAC 3' and subcloned into the pIZ/promoterless/H3/GFP plasmid using HindIII and EcoRI. The resulting Act_{proc} reporter gene consisted of the Actin 5C promoter followed by a region encoding the first 67 amino acids of histone H3, a portion of the histone H3 3' UTR that included the SL and HDE, an in-frame GFP ORF, and a vector-encoded OpIE2 poly(A) signal.

The His_{trans} construct was created by amplifying the histone H3 promoter using forward primer 5' GGCCAAGCTTCGACAAAAACCCGAGAGAGTAC 3' and reverse primer 5' GGCCACCGGTCTCCGATTTGGGTTTCACTAAAGTTCACGTTC 3' and subcloning this piece into the pIZ/promoterless/H3/GFP plasmid using HindIII and AgeI. The resulting His_{trans} reporter consisted of the Histone H3 promoter followed by an EGFP ORF and a vector-encoded OpIE2 poly(A) signal.

The Act_{trans} construct was created by amplifying the Act 5C promoter with forward primer 5' GGCCAAGCTTTATGTATGTTTTGGCATACAATGAGTAGTTGG 3' and reverse primer 5' CCGGGAATTCGTGTCGGGAGGAGTATCCAC 3' and the EGFP ORF using forward primer 5' GGCCGAATTCATGGTGAGCAAGGGCGAGGAG 3' and reverse

primer 5' GGCCTCTAGATTACTTGTACAGCTCGTCCATGCC 3' and performing a three way ligation into the pIZ/promoterless vector using HindIII, EcoRI, and XbaI. The resulting Act_{trans} reporter consisted of the Actin 5C promoter followed by an EGFP ORF and a vector-encoded OpIE2 poly(A) signal.

Generation of dsRNAs and RNA interference

DNA templates for double stranded RNAs (dsRNAs) were generated by PCR using *Drosophila* Gene Collection library clones (Open Biosystems) or genomic DNA, if appropriate, and primers containing a T7 promoter (Table 4-1). *In vitro* transcription was carried out in a total volume of 300 µl using T7 RNA polymerase (Fermentas) or the TranscriptAid T7 high yield transcription kit (Fermentas), as recommended by the manufacturer. Following transcription, the RNA was treated with 5 units of RNase-free DNase Q (Promega), purified on G-50 micro columns (GE Healthcare), boiled for 5 minutes and cooled to room temperature to form dsRNA.

To induce RNAi-mediated knockdown of a factor of interest, dsRNAs targeting the factor were introduced into *Drosophila* D.Mel-2 cultured cells by adding the dsRNAs directly to the growth medium. For 384-well experiments, 5 ul diluted dsRNA containing 250 ng dsRNA in water was added to 10 µl SF-900II SFM (Gibco) containing $8x10^3$ cells in a black-walled, clear bottom tissue culture plate (Corning) in a final volume of 15 µl. For 96-well plates, $8x10^4$ cells were plated in 40 µl SF-900II SFM and supplemented with 800 ng dsRNA. For 24-well plates, $5x10^5$ cells were plated in 250 µl SF-900II SFM to which 5 µg dsRNA was added. For 6-well plate experiments, $3x10^6$ cells were plated in 1 ml SF-900II SFM and 30 µg dsRNA was added. For all experiments, the cells were allowed to

knockdown at 27 °C in a humidified growth chamber for 3-6 days, depending upon the experiment and dsRNA target. If RNAi was to proceed longer than 3 days, an amount of dsRNA equal to the original application was added every 2 days to ensure continuous knockdown for all plating formats except 384-well format, for which dsRNA was added only on Day 0.

Cell Transfections and Reporter Assays

For experiments involving the His_{proc}, and His_{trans} reporters, RNAi-depleted cells were transfected using the Amaxa Nucleofector electroporation system (Lonza) and the Nucleofector V kit (Lonza) according to the manufacturer's protocol for the *Drosophila* S2 cell line. Briefly, $5x10^6$ RNAi-depleted D.Mel-2 cells were isolated by centrifugation at 1000xg for 3 min., followed by aspiration of the media from the cell pellet and addition of 100 µl room-temperature supplemented Nucleofector solution. To this was added 2 µg reporter plasmid DNA and the cells were resuspended gently and transferred to a Nucleofector electrporation cuvette. The cuvette was capped and placed inside the Nucleofector device and the cells were subjected to electroporation using the G-030 Nucleofector program. After electroporation, 500 µl of room temperature growth medium (SF-900II Serum Free Medium supplemented with antibiotic/antimycotic solution) was added to the cuvette and the cells were transferred with a plastic pipette into 6-well tissue culture plates containing 400 µl room temperature growth medium. The cells were then allowed to grow at 27 °C in a humidified growth chamber until viewing.

The Act_{proc} construct was integrated stably into *Drosophila* D.Mel-2 cells using the Amaxa transfection protocol above and selection of the cells from Day 3 through Day 10

with 500 µg/ml Zeocin selection reagent (Invitrogen). Note that most of the experiments using this reporter utilized the reporter stable line instead of the transiently transfected reporter. RNAi experiments were performed in this cell line the same way as was done for experiments in which the reporter was transiently transfected.

RT-PCR Analysis

Relative reporter and endogenous mRNA levels were measured using reverse transcriptase PCR (RT-PCR). D.Mel-2 cells were harvested after treatment with dsRNAs and total RNA was isolated using TRIzol reagent (Invitrogen). An aliquot of the isolated RNA (2 µg) was incubated with RNase-free DNase Q (Promega) followed by reverse transcription using MMLV-RT (Invitrogen) and random hexamers, as suggested by the manufacturer. PCR amplification was carried out with 0.5 µl of cDNA using the following forward (F) and reverse (R) primers written in the 5'-3' orientation: SLBP (F-CAGTAAACAAACGAAAGATAATTGCACAAC3', R-GACCTCCTGGGCCCAGCTCTG); FLASH (F-CGAAGCCAATCCGCATCG, R-CATCAGCTGGGCAATCTTCAG); misprocessed reporter (F within H3 ORF-GAGCACCGAGCTTCTAATCC, R within GFP ORF-GGCGGACTTGAAGAAGTCGTGC); Histone H1 total (F-CGTGGCGGTTCATCACTTC, R-GCTTCGCTGCAGTCACTTTC); Histone H2a total (F-GGCCATGTCTGGACGTGGAAAAGGT, R-GGCCTTAGGCCTTCTTCTCGGTCTT); Histone H2b total (F-CCAAGAAGGCTGGCAAGG, R-GCTGGTGTACTTGGTGACAG); Histone H1 misprocessed (F-GCTAAGAAGGCTGTGGCTACC, R-GTTGCTGCGAACGTCAGC); Histone H2A misprocessed (F-

GTGACAACAAGAAGACTAGAATTATTC, R-CTAATTACAACAAATTGCCAAGC); Histone H2A misprocessed (F-CTGGCAAGGCTCAGAAGAAC, R-GTAAATTCATATTCGATGATTGGTGGTTG); α1-Tubulin 84B (F-GCCCTACAACTCCATCCTGA, R-GGTCACCAGAGGGAAGTGAA). PCR products were run on a 2 % agarose gel and detected by ethidium bromide staining.

Creation of GST-fusion and *in vitro* Transcription/Translation Constructs

Plasmid constructs encoding wild type and mutant dFLASH or dLsm11 ORFs N-terminally fused to gluthione-s-tranferase were created by cloning the ORFs into the pET42a vector using EcoRI and XhoI (Fermentas). *In vitro* transcription/translation constructs were created by cloning the sequence of interest into the pOT2 vector using EcoRI and XhoI (Fermentas). The resulting construct contained a wild type or mutant dFLASH or dLsm11 ORF driven by a T7 promoter. This construct was used in coupled *in vitro* transcription/translation reactions resulting in the synthesis of radiolabeled peptide corresponding to the dFLASH or dLsm11 region of interest. A list of primers used for PCR amplification of inserts is shown in Tables 4-2 and 4-3.

GST-fusion Protein Purification

Chemically-competent E. coli BL21(DE3) cells were transformed by standard heat shock with pET42a plasmid constructs encoding glutathione-S-transferase (GST) fused in frame to wild type or mutant dFLASH or dLsm11 coding sequences and plated onto selective LB agar plates containing 30 ug/ml kanamycin. A single colony was picked from the plate and used to inoculate 300 ml of LB medium. The cultures were grown at 37 °C with shaking

until their absorbance at 595 nm reached 0.6, at which time IPTG was added to a final concentration of 0.5 mM. The cells were allowed to induce in the presence of IPTG for 3 h at 37 °C with shaking before harvesting by centrifugation. Cell pellets were retained and frozen at -80 °C until needed for purification.

Bacterial cell pellets were resuspended in 20 ml lysis buffer (50 mM Tris, pH 8.0, 5 mM EDTA, 50 mM NaCl, 10% glycerol, 0.5% NP-40). After resuspension, PMSF was added to a final concentration of 1 mM and the cells were incubated 15 min with occasional mixing. The cell suspension was then sonicated on ice three times for 20 seconds, with at least 2 min. between sonications. The sonicated lysate was transferred to a 50 ml Oak Ridge tube and centrifuged for 10 min. at 10,000 RPM at 4 °C in a Sorvall RC-5B centrifuge. The supernatant was transferred to a fresh 50 ml conical tube on ice and 300 µl of glutathione-agarose beads (Sigma) pre-equilibrated with lysis buffer were added. The beads were incubated with the lysate for 4 h at 4 °C with gentle rocking. After incubation, the beads were recovered by centrifugation at 300xg for 5 min. at 4 °C and removal of the supernatant. The beads were then washed three times with 15 ml cold PBS at 4 °C with gentle rocking for 10 min. per wash. GST-fusion proteins were eluted from the beads by adding 500 µl elution buffer (50 mM Tris, pH 8.8, 10 mM glutathione) and rotating at 4 °C for 30 min. Following this incubation, the beads were centrifuged and the supernatant, containing the free GST-fusion proteins, was reserved and stored at -20 °C.

In Vitro Translation

In vitro translation reactions were performed using either a coupled transcription and translation (TNT) kit (Promega) to express a plasmid construct containing a T7 promoter

sequence driving a FLASH or Lsm11 coding sequence or using nuclease-treated rabbit reticulocyte lysate (Promega) to translate *in vitro* transcribed and capped Lsm11 mRNA. Each reaction was set up according to the manufacturer's protocol. Briefly, for coupled transcription and translation reactions, 1 µg of a T7 promoter-containing plasmid encoding my protein of interest was added to a reaction mixture containing 25 µl TNT rabbit reticulocyte lysate, 2 µl TNT reaction buffer, 1 µl 1mM amino acid mixture minus methionine, 1 µl 40u/µl RiboLock ribonuclease inhibitor (Fermentas), 2 µl [35S]methionine (>1000Ci/ml), 1 µl TNT RNA polymerase (Promega) or 1 µl T7 RNA polymerase (Fermentas), and autoclaved deionized water to a total reaction volume of 50 µl. The reactions were incubated at 30 °C for 90 min, after which _____ mM Ribonuclease A (Sigma) was added and the samples were incubated an additional 5 min. at 30 °C. The reactions were then stored at -20 °C until needed.

For uncoupled *in vitro* translation reactions, capped messenger RNA to be used as template in the IVT reaction was first created using the mMESSAGE mMACHINE kit (Ambion) according to the manufacturer's protocol. pOT2 plasmid constructs containing a T7 promoter and the coding sequence of interest were linearized using (PUT THE NAME OF THE ENZYME HERE) and 1 µg of the linear plasmid was added to a reaction containing 2 µl 10X reaction buffer (salts, buffer, DTT, and proprietary ingredients), 10 µl2X NTP/CAP mixture (neutralized buffer containing 15 mM each ATP, CTP, UTP, 3 mM GTP, 12 mM cap analog), 2 µl enzyme mix (buffered 50% glycerol containing T7 RNA polymerase, RNase inhibitor, and proprietary components), and nuclease-free water to a total 20 µl reaction volume. Transcription reactions were incubated at 37 °C for 2 h, after which 1 µl TURBO DNase (Ambion) was added and the samples were incubated at 37 °C for an

additional 15 min. Following incubation, the RNA was purified suing phenol:chloroform:isoamyl alcohol extraction followed by isopropanol precipitation. Briefly, 115 µl nuclease-free water was added to the transcription reaction along with 15 µl ammonium acetate stop solution (5 M ammonium acetate, 100 mM EDTA). To this, 150 µl phenol:chloroform:isoamyl alcohol (Invitrogen) was added, the sample was shaken vigorously for 15 sec., and the sample was incubated at room temperature for 3 min. After incubation, the sample was centrifuged at 16000xg in a microcentrifuge at 4 °C for 15 min. The aqueous phase was recovered and placed into a fresh microfuge tube. The RNA was precipitated by adding 150 µl isopropanol and incubating at -20 °C for 15 min., followed by centrifugation at 16000xg for 15 min. The supernatant was removed and the RNA was resuspended in nuclease-free water.

The *in vitro* synthesized capped mRNA was then used as template for *in vitro* translation reactions using nuclease-treated rabbit reticulocyte lysate (Promega) according to the manufacturer's protocol. Briefly, 2 µg capped mRNA were added to a reaction mixture containing 35 µl nuclease-treated rabbit reticulocyte lysate, 1 µl 1 mM amino acid mixture minus methionine, 1 µl 40u/µl RiboLock RNase inhibitor (Fermentas), 2 µl [35S]methionine (>1000Ci/ml), and autoclaved deionized water to a total reaction volume of 50 µl. The reactions were incubated at 30 °C for 90 min. The reactions were then stored at -20 °C until needed.

GST Pulldown Experiments

GST pulldown reactions were set up by combining 6 µl *in vitro* translated protein (IVT) with 10 µg GST-fusion protein in a total of 100 µl of 1X binding buffer (15 mM Tris,

pH 8.0, 20 mM EDTA, 75 mM KCl, 0.375 mM DTT, 10% glycerol, and 0.1% NP-40). Samples were then incubated on ice for 30 min. with occasional mixing, followed by a 30 min. incubation at room temperature for 30 min. with occasional mixing. Samples were then clarified by centrifuging at 13000xg for 15 min. at 4 °C, after which the supernatant was transferred to a fresh microcentrifuge tube. To the supernatant was added 370 µl of 1X binding buffer and 30 µl of a 1:1 slurry of glutathione-agarose beads (Sigma) pre-equilibrated with 1X binding buffer and the supernatant and beads were rotated together for 1 h at 4 °C. After incubation, the beads were collected by centrifuging at 300xg for 1 min. and removing the supernatant. The beads were then washed four times with 1X binding buffer for 10 minutes per wash. After the final wash, the supernatant was removed by careful aspiration and 25 µl SDS loading buffer (100mM Tris-Cl, pH 6.8, 200 mM DTT, 4 % SDS, 20 % glycerol, and 0.2% bromophenol blue) was added to the beads. The samples were then boiled for 3 minutes and centrifuged at 300xg at room temperature to pellet the beads. The entire non-bead volume was then loaded onto a 12.5% or 15% polyacrylamide gel. Following electrophoresis, the gels were stained with Coomassie stain and destained using a destain solution consisting of 10% acetic acid and 20% methanol in water, dried on a gel drier, and exposed to film or a phosphor screen.

Creation of *Drosophila* Expression Constructs

Drosophila expression constructs were created by cloning the dFLASH promoter into the promoterless pIZ/V5/His vector, followed by subcloning a full or partial dFLASH ORF into this construct. The endogenous dFLASH promoter and 5'UTR were amplified by PCR off of genomic DNA from wild type adult flies using primers targeting the entire genomic

region between the upstream gene, *latheo* (dORC3), and the start codon of dFLASH. This region consists of a putative promoter of 233 nt followed by 336 nucleotides containing the dFLASH 5'UTR up to the start codon of the ORF. The region encoding the 5'UTR is annotated to contain the entire first exon of dFLASH, from nucleotides 1-236, an intron spanning the next 70 nucleotides of the genomic sequence, and the first 30 nucleotides of the second exon, which continues into the dFLASH ORF. This promoter/5'UTR sequence was cloned into the promoterless pIZ/V5/His gene using HindIII and EcoRI restriction enzymes (Fermentas), resulting in the pIZ/FLASHp/V5/His construct. Portions of the dFLASH ORF were subcloned into this construct using EcoRI (Fermentas) and SacII (NEB) restriction enzymes. The result was a construct containing the dFLASH promoter driving expression of the full length or truncated dFLASH ORF, followed by an in-frame V5 epitope and polyhistidine tag and a strong, vector-encoded polyadenylation signal. A list of primers used for PCR amplification of inserts is shown in Table 4-4.

Drosophila Cell Culture and Transfection of Expression Constructs

Drosophila D.Mel-2 cells (Invitrogen) were cultured in 10 cm tissue culture plates (Greiner) in SF-900II Serum Free Medium (Gibco) supplemented with 1X antibiotic/antimycotic solution (Sigma) containing penicillin, streptomycin, and amphotericin B in a humidified growth chamber at 27 °C.

For dFLASH localization experiments, pIZ/V5/His constructs containing the FLASH coding region of interest preceded by the natural FLASH promoter and 5'UTR were transfected into D.Mel-2 cells using the Amaxa Nucleofector electroporation system (Lonza) and the Nucleofector V kit (Lonza) according to the manufacturer's protocol for the

Drosophila S2 cell line. Briefly, $5x10^6$ D.Mel-2 cells were isolated by centrifugation at 1000xg for 3 min., followed by aspiration of the media from the cell pellet and addition of 100 µl room-temperature supplemented Nucleofector solution. To this was added 2 µg plasmid DNA containing the FLASH coding region of interest and the cells were resuspended gently and transferred to a Nucleofector electrporation cuvette. The cuvette was capped and placed inside the Nucleofector device and the cells were subjected to electroporation using the G-030 Nucleofector program. After electroporation, 500 µl of room temperature growth medium (SF-900II Serum Free Medium supplemented with antibiotic/antimycotic solution) was added to the cuvette and the cells were transferred with a plastic pipette into 6-well tissue culture plates containing 400 µl room temperature growth medium. The cells were allowed to grow at 27 °C in a humidified growth chamber for 48-72 h. before fixation.

Confocal Microscopy

Drosophila D.Mel-2 cells transfected with pIZ/V5/His plasmid constructs containing a FLASH promoter and 5'UTR followed by a FLASH coding sequence were harvested on Day 2 or 3 post-transfection and replated onto laminin-treated glass coverslips (BD Biosciences) in 24-well tissue culture plates and allowed to adhere to the coverslips for 1 h at 27 °C in a humidified growth chamber. After the cells had adhered to the coverslips, the media was removed and the cells were washed briefly with PBS and fixed with 4% electron microscopy grade paraformaldehyde for 10 min. at room temperature. The cells were washed briefly with PBS and permeabilized with 0.1% Triton X-100 (Sigma) in PBS for 30 min. The cells were then washed briefly three times with 0.1% Triton X-100 in PBS.

Following the washes, the cells were blocked with 5% normal goat serum (Sigma) in PBS for 30 min. Next, the coverslips were transferred onto parafilm, cell side up, into a humidified chamber and primary antibody (anti-V5 diluted at 1:500 to detect recombinant FLASH, anti-Mxc diluted at 1:500 to mark HLBs) diluted in 5% normal goat serum in PBS was applied to the entire surface of the coverslip, followed by incubation at room temperature for 1 h. After incubation, the coverslips were returned to the 24-well plate and washed with PBS three times for 10 min. per wash. Next, the coverslips were transferred back to the parafilm and covered with a solution of 5% normal goat serum in PBS containing appropriate secondary antibodies, followed by incubation at room temperature for 1 h. The coverslips were then transferred back to the 24-well plate and washed with PBS three times for 10 min. per wash. Next, the cells were stained with DAPI diluted to 0.5 µg/ml in PBS for 5 min., followed by three brief washes with PBS. The coverslips then mounted onto microscope slides, allowed to dry, and stored at -20 °C. Antibody-stained cells were viewed using a confocal microscope at 400X magnification.

Functional Reporter Assays

Drosophila D.Mel-2 cells stably expressing the Act_{proc} reporter were transfected with pIZ/FLASHp/$FLASH_{ORF}$/V5/His constructs encoding FLASH ORFs lacking the C-terminal-most 111 amino acids using the Amaxa Nucleofector electroporation protocol. The cells were allowed to express the FLASH constructs for 2 days before the addition of a dsRNA targeting the region of FLASH encoding the C-terminal 111 amino acids of the FLASH open reading frame. Cells were then viewed 2-3 days later using an inverted fluorescence light microscope to detect GFP signal.

Results

Recent work by our group in both *Drosophila* and mammalian systems has further defined the role of FLASH in histone expression. Human FLASH was identified as a binding partner of the N-terminal region of Lsm11 in a yeast-two-hybrid screen, and the *Drosophila* ortholog was identified bioinformatically, and also shown to interact with Lsm11 (Yang *et al.*, 2009). A fragment containing less than 200 amino acids of human FLASH contains all the elements required for histone pre-mRNA processing *in vitro*.

The Role of dFLASH in Histone Transcription and pre-mRNA Processing

Early studies of the role of FLASH in histone gene expression suggested that FLASH is involved in histone gene transcription in mammals (Barcaroli *et al.*, 2006a,b). However, because dFLASH was only recently identified (Yang *et al.*, 2009), no information about any possible role in transcription role of dFLASH in *Drosophila* was available. To help answer this question, I used the reporters described in Chapter II to distinguish effects on histone pre-mRNA processing and transcription. RNAi-mediated knockdown of dFLASH had no effect on fluorescence off of the His_{trans} reporter when compared to PTB, Lsm11, or SLBP-depleted cells (Fig. 4-1A), suggesting that dFLASH is not involved in the transcription of histone genes in *Drosophila*.

Since FLASH was found to interact with Lsm11, a known component of the U7 snRNP, this strongly suggested a role for FLASH in the histone pre-mRNA processing reaction. Despite this, dFLASH did not score strongly in the initial histone pre-mRNA processing screen (Wagner *et al.*, 2007; Chapter III of this dissertation), although it was

placed in the "weakest category" of positive hits and reproducibly activated the reporter to a small extent (Fig. 4-1B). Since there is not quality control on all 22,000 siRNAs used in the initial screen, I performed experiments using dsRNA targeting dFLASH using the original His_{proc} reporter and found that RNAi-mediated depletion of dFLASH results in significant fluorescence off the His_{proc} reporter (Fig.4-1C), suggesting technical problems with the dsRNA targeting this factor in the initial screen. Consistent with the interaction data, these results support the notion that dFLASH is not required for transcription off the histone promoter, but that it is required for the histone pre-mRNA cleavage reaction.

To investigate the processing role of dFLASH further, I used the Act_{proc} reporter that I stably integrated into *Drosophila* cultured cells (Chapter II). This stable cell line was used in experiments conducted over a period of 3 days, with fluorescence images taken after two and three days (Fig. 4-1D). As expected, RNA interference-mediated knockdown of PTB, a splicing factor with no role in histone pre-mRNA processing, resulted in only background levels of GFP fluorescence off the reporter. When SLBP, a factor required for histone pre-mRNA processing, was depleted, the cells showed robust fluorescence by the second day, which was intensified by the third day, likely due to increased depletion of SLBP. When dsRNAs targeting dFLASH were added to the reporter stable cell line, a similar result was observed. By the second day, appreciable amounts of fluorescence were visible from these cells and by the third day, this fluorescence was even more robust. Note that two independent dsRNAs, targeting two completely separate regions of dFLASH, were used in this experiment, indicating that the fluorescence that arises is specific to knockdown of dFLASH and not due to secondary effects. These results suggested that dFLASH is required for processing histone pre-mRNAs and played no role in transcription of histone genes.

These experiments provided functional evidence that dFLASH has an active role in processing.

To detect the presence of misprocessed histone mRNAs in the reporter cells, total RNA was extracted and reverse transcriptase PCR (RT-PCR) was employed (Fig. 4-1E) using primers specific for misprocessed histone mRNAs (Fig. 4-1E). Also, total histone mRNA levels were assessed using primers located within the histone ORF, and knockdown was assessed using primers to detect the levels of SLBP and FLASH (Fig. 4-1E). There was significant knockdown of both SLBP and FLASH in these experiments, but no effect of FLASH siRNAs on SLBP mRNAs. As was seen visually using the reporter stable line (Fig. 4D), while PTB dsRNA-treated cells showed only background expression off the reporter, cells depleted of SLBP and FLASH showed high amounts of misprocessing (Fig. 4-1E). Similar results were found with the endogenous histone mRNAs, which showed little to no misprocessing in PTB dsRNA-treated cells, while cells depleted of SLBP or FLASH showed significant amounts of misprocessing by RT-PCR. Note that histone H1 was affected less by knockdown of dFLASH than by knockdown of dSLBP. The significance of this is unknown.

Similar results were found when histone H3 mRNA from FLASH knockdown cells was analyzed by Northern blotting (Fig. 4-1F). Substantial amounts of polyadenylated histone H3 mRNA were present in total cell RNA, similar to that observed when SLBP was knocked down. Importantly, despite suggestions from earlier studies in mammalian cells that FLASH is involved in transcription of histone genes, no effect is seen on the overall histone mRNA levels for histones H1, H2A, and H2B. This is in agreement with the results using the histone H3 promoter-driven transcriptional reporter (Fig. 4-1C).

The *in vivo* processing data from RNAi experiments described above is corroborated by the results of *in vitro* histone pre-mRNA processing assays using nuclear extract from *Drosophila* Kc cells (Yang *et al.*, 2009). When antibody targeting the N-terminus of dFLASH is added to nuclear extract that is competent to support processing of histone mRNA, processing is completely blocked (Fig. 4-1G, lanes 4 and 6), while addition of an antibody targeting the N-terminus of human FLASH has no effect (Fig. 4-1G, lane 3). When the N-terminal fragment of dFLASH used to create the antibody is added to the reaction along with the antibody, the processing defect observed for the dFLASH antibody-treated sample is eliminated (Fig. 4-1G, lane 8). Together, these results, along with the reporter experiments described above, reveal that dFLASH is required for processing histone pre-mRNA both *in vivo* and *in vitro*.

Sequence Conservation of FLASH and Lsm11 from Mammals to Flies

At 1982 amino acids (~200 kDa), mammalian FLASH is considerably larger than its *Drosophila* counterpart, which is only 844 amino acids (~96 kDa) in length. In addition, the mammalian protein contains C-terminal domains that are not readily identifiable in dFLASH. The two proteins share some conserved amino acid residues within their N termini, with the region between amino acids 45-177 in dFLASH containing 25 % sequence identity and 48 % sequence similarity with amino acids 30-169 of human FLASH (Fig. 4-2A). This region of conservation is contained within the longer N-terminal fragment (amino acids 1-177 in flies) that was found to interact with the N-terminus of U7 snRNP component Lsm11 via yeast-two-hybrid assay and later by GST-pulldown experiments (Yang *et al.*, 2009). There is no significant homology outside of this region.

There is a high degree of conservation from mammals to flies within the N-terminus of Lsm11 (Fig.4-2B). Most intriguing is a stretch of conserved residues located between amino acids 20-37 in the *Drosophila* protein (24-40 in human Lsm11). Within the conserved region, 12 out of 17 amino acids are perfectly conserved. No factors that interact with this region have been described. Other regions of Lsm11 have been characterized, including the Sm binding domains (Pillai *et al.*, 2003) and the ZFP100-interacting domain in mammals (Azzouz *et al.*, 2005), but the identification of FLASH as a binding partner gave the first suggestion of a possible function for these highly conserved N-terminal residues. Specifically, the binding interaction with FLASH absolutely requires amino acids 1-40 of Lsm11, as deletion of these residues abolishes binding (Yang *et al.*, 2009). However, it remained unclear precisely which residues were involved.

Biochemical Interactions between dFLASH and dLsm11

Regions of dFLASH required for binding. Recent work by our group, including yeast-two-hybrid assays and GST pulldowns, has shown that the N-terminus of dFLASH physically interacts with the N-terminus of dLsm11. Specifically, it was shown that a conserved region corresponding to the first 177 amino acids of dFLASH interacts with the first 153 amino acids of dLsm11 and that the first 40 amino acids of dLsm11, containing several conserved residues, are essential for binding to occur (Yang *et al.*, 2009). In order to investigate the details of this binding interaction further, I created deletion and point mutants for both proteins (Figs. 4-3A and 4-4A), cloned them into a vector encoding an N-terminal glutathione-S-transferase (GST) cassette, and assayed binding interactions using GST-pulldown assays. For these assays, ^{35}S-labeled dLsm11 or dFLASH, as appropriate, was

synthesized by *in vitro* translation and incubated with the recombinant GST-fusion proteins, and the bound proteins were detected by autoradiography (Fig. 4-3B-E and 4-4B-E).

In order to narrow down the possible binding region within dFLASH, I created a number of deletion mutants in which residues were removed from the N-terminal and C-terminal ends of a polypeptide consisting of the first 178 amino acids of dFLASH (Fig. 4-3A) and used these mutants in GST pulldown experiments with ^{35}S-labeled dLsm11 and GST-FLASH fusion proteins. Deleting amino acids 1-66 or 1-77 from the N-terminus of FLASH protein had no effect on binding (Fig. 4-3B, lanes 3 and 4), suggesting the binding region lies downstream of position 77. Similarly, removal of amino acids 155-178 had no effect on binding (Fig. 4-3B, lane 5), indicating that this region is not required for the interaction with dLsm11 and narrowing the binding region to between amino acids 78-154.

I next systematically deleted regions from the C terminal end of the N-terminal 178 amino acids using ^{35}S-labeled FLASH and GST-Lsm11 and found that deleting residues 144-178 decreased binding (Fig. 4-3C, lane 12) and deletion of residues 138-178 and 132-178 abolished binding (Fig. 4-3C, lanes 9 and 6, respectively), suggesting critical binding residues reside C-terminal to amino acid 137. I tested point mutants in and around this region and found that two conserved residues, E136 and E141, led to reduced binding when mutated (Fig. 4-3D, lanes 4 and 8, respectively). Mutation of residues RK139,140 (Fig. 4-3, lane 6), I144 (Fig. 4-3, lane 10), and L147 (Fig. 4-3, lane 12) had no significant effect on binding compared to wild type (Fig. 4-3, lane 2).

In order to define better the N-terminal boundary of the binding region, I created deletion mutants from which the first 104, 121, 141, and 166 amino acids were removed from the N-terminus. I found that deletion of amino acids 1-104 had no effect on binding (Fig. 4-

3E, lane 3). However, removal of residues 1-122 decreased binding almost completely (Fig. 4-3E, lane 4) and removal of residues 1-141 and 1-166 completely abolished the ability of dFLASH to bind dLsm11 (Fig. 4-3E, lanes 5 and 6, respectively). Together with earlier data, this indicates that crucial binding residues reside in a region between amino acids 105-154. Consistent with this, mutation of residues NL125,126 and LL129,130 to alanine abolished binding (Fig. 4-3E, lanes 7 and 8, respectively), as did mutating leucines 118 and 119. My results are consistent with a set of parallel experiments exploring the interaction between human FLASH and human Lsm11 (X. Yang and Z. Dominski, unpublished results). In these experiments, mutation of human FLASH at residues LI118,119 (corresponding to LL129,130 in flies) abolishes binding to Lsm11 (Fig. 4-3F, lane 5), while mutation of KD129,130 (KE140,141 in flies) reduces binding (Fig. 4-3F, lane 6). Further, mutation of hFLASH residue L136 (L147 in flies) had no effect on binding (Fig. 4-3F, lane 4), in agreement with my findings for dFLASH (Fig. 4-3D, lane 12), nor did mutation of residue N101 (Fig. 4-3F, lane 3).

Region of dLsm11 that binds dFLASH. In order to define the residues of dLsm11 that bind dFLASH, I used a similar approach as above by creating deletion and point mutants (Fig. 4-4A) of dLsm11 and testing them in GST-pulldown assays. It was shown previously that deleting the first 40 amino acids eliminates proper binding between the two proteins (Yang *et al.*, 2009). However, even the presence of the first 54 amino acids fused to GST is not enough to facilitate proper binding (Fig. 4-4B, lane 6), which indicates that residues 1-40, while necessary, are not sufficient for binding. When I used a GST fusion protein containing amino acids 1-78, binding was restored to near control levels (Fig. 4-4B, lane 5). This suggests the presence of important residues, not only within the first 40 amino acids, but also

further downstream, and broadens the possible total region of interaction to the first 78 amino acids of dLsm11.

Because the first 40 amino acids of dLsm11 were shown previously to be required for binding dFLASH, I focused my efforts for determining essential binding residues in this region. Mutation of residues FN27,28 to alanine eliminates the binding interaction between dLsm11 and the N-terminal 178 amino acids of dFLASH (Fig. 4-4C, lane 7), while mutation of residues SD24,25 (Fig. 4-4C, lane 6), PL29,30 (Fig. 4-4C, lane 8), or PN36,37 (Fig. 4-4C, lane 9) had no effect on binding. Note that, as revealed above, inclusion of only the first 40 amino acids of dLsm11 was not sufficient to allow binding to occur (compare Fig. 4-4C, lanes 6 and 11, and lanes 9 and 12). This result was repeated using full length dFLASH protein (Fig. 4-4D), revealing again that residues FN27,28 are critical for the binding interaction between dFLASH and dLsm11. Using the reverse approach than I used in panels B-D, I *in vitro* synthesized and ^{35}S-labeled wild type and mutant dLsm11 and incubated these with wild type GST-FLASH protein (AA1-178). The results of this experiment confirm the earlier findings that FN27,28 are critical residues for binding (Fig. 4-4E, lane 9) and also reveal that mutation of residues LDV20,21,22 or residues LY33,34 to alanine abolishes binding (Fig. 4-4E, lanes 15 and 18, respectively). Together, these results indicate that critical residues exist between amino acids 20-34 and three distinct patches of residues in this 14 amino acid region have been identified as essential for binding (Fig. 4-4, panel B).

These results agree with data from the human proteins (X. Yang and Z. Dominski, unpublished results), indicating that mutation of residues LDV24,25,26 (LDV20,21,22 in flies) results in reduced binding (Fig. 4-4F top, lane 5; Fig. 4-4F bottom, lane 6) and mutating residues FD31,32 (FN27,28 in flies) causes complete loss of binding (Fig.4-4F top, lane 7),

while mutating residues PL33,34 (PL29,30 in flies) has no effect on binding (Fig. 4-4F top, lane 9). In contrast to the fly data, mutation of residues LY37,38 (LY33,34 in flies) has no effect on binding (Fig. 4-4F bottom, lane 9).

Localization of dFLASH

FLASH is known to localize to the HLB in both mammals and flies, based upon colocalization experiments with NPAT in mammals (Barcaroli *et al.*, 2006) and MPM-2 in flies (Yang *et al.*, 2009; Fig. 4-5A). Immunofluorescence experiments were used to help characterize the localization of dFLASH and, specifically, whether dLsm11 is required for its localization. I created V5 epitope-tagged dFLASH deletion mutants (Fig. 4-5B), expressed them in *Drosophila* cultured cells, and used immunofluorescence against the V5 epitope to determine localization of these proteins. Localization to the HLB was determined by costaining for Mxc, a known HLB component. Deletion of the first 165 residues, including the dLsm11 binding region, did not affect dFLASH localization (Fig. 4-5C). Further, deletion of the last 111 amino acids from the C-terminus of dFLASH has no effect on its localization (Fig. 4-5C). Taken together, these results point to the requirement for an uncharacterized region of dFLASH between amino acids 165 and 733 for proper localization to the HLB. Further deletions will allow us to narrow down this region further.

Functional Characterization of dFLASH

The first 177 amino acids of dFLASH are necessary and sufficient to support processing of histone pre-mRNA in nuclear extracts (Yang *et al.*, 2009). This region binds dLsm11, and presumably at least one other factor required for histone pre-mRNA processing.

In order to help determine which regions of the N-terminus of dFLASH are required for proper processing of histone pre-mRNA and not for binding dLsm11, I developed a functional assay involving the use of the constitutive histone pre-mRNA processing reporter. The idea behind this assays is that depletion of dFLASH from *Drosophila* cultured cells results in misprocessing of the reporter, but if recombinant dFLASH protein containing the region required for processing is introduced into cells that are subsequently RNAi-depleted of endogenous dFLASH, this should rescue the misprocessing phenotype, preventing the cells from turning green. Ideally in this case, dsRNAs targeting UTR sequences of dFLASH could be used, effectively knocking down the endogenous message while leaving the recombinant dFLASH mRNA, which does not contain this sequence, able to accumulate in the cell. I tested several dsRNAs targeting 5' and 3' UTRs of dFLASH, yet none of the dsRNAs used knocked down dFLASH well enough to trigger misprocessing of the reporter (Fig. 4-6A). One problem was that the 3' UTR of FLASH was quite short and there is one annotated form that only has a 20 nt 3' UTR.

Because targeting the UTRs was ineffective, I designed a dsRNA targeting the very C-terminus of the dFLASH ORF, in a region encoding amino acids 734-844, which is not required for localization of the protein to the HLB or for its interaction with dLsm11. Importantly, this dsRNA knocks down dFLASH sufficiently to elicit misprocessing of the reporter, as evidenced by increased green fluorescence in treated cells versus the negative control (Fig. 4-6B). For the preliminary rescue experiment, a dFLASH construct was created that did not encode the C-terminal 111 amino acids of the endogenous protein. In this way, endogenous dFLASH can be knocked down while the deletion mutant is expressed, allowing for a determination of rescue to be made based upon the activity of the reporter.

The results of these RNAi/reporter experiments reveal that transfection of the construct encoding dFLASH amino acids 1-733 driven by the natural dFLASH promoter reduces the amount of fluorescence from the reporter, although it is not sufficient to rescue the misprocessing phenotype completely (Fig. 4-6C). Before attempting the rescue experiment with N-terminal deletion mutants, I transfected a number of the dFLASH constructs with deletions from the N and C-terminus into the reporter stable line with no dsRNA treatment. Transfecting full-length FLASH or FLASH 1-733 had no effect on the reporter, but surprisingly, transfection of a construct with a deletion of amino acids from 1-77 from FLASH (FLASH 78-733) resulted in significant fluorescence off the reporter (Fig. 4-6D), indicating that this protein inhibits histone pre-mRNA processing *in vivo*. FLASH 78-133 acts as a dominant negative inside the cell. This protein still binds Lsm11 normally. This result suggests that the region from 1-77 is important for histone pre-mRNA processing. Presumably the mutant protein may bind Lsm11 but is not able to bind another essential factor, effectively rendering the U7 snRNP inactive.

In agreement with this interpretation are results from *in vitro* processing assays using human FLASH. A recombinant protein containing GST fused to the N-terminus of hFLASH (amino acids 1-139) is active in processing histone pre-mRNA in mammlian nuclear extract (Yang *et al.*, 2009). When the first 61 amino acids (corresponding to dFLASH amino acids 1-77) are deleted from the active recombinant hFLASH and this fragment (amino acids 62-139) is added to processing-competent nuclear extract from HeLa cells, the processing reaction is inhibited (X. Yang and Z. Dominski, unpublished results), demonstrating that the protein is a dominant negative in a biochemical system.

Discussion

Mammalian FLASH is a protein of many and varied functions, playing key roles in pro-life and pro-death processes, both inside the nucleus and out in the cytoplasm. Initially characterized as a factor involved in apoptosis, FLASH was later implicated in histone gene expression. In this chapter I report experiments defining a role for FLASH in histone pre-mRNA processing in *Drosophila*, which support the biochemical experiments in mammals that show it is required for histone pre-mRNA processing in nuclear extracts (Yang *et al.*, 2009). Many questions about how FLASH participates in histone pre-mRNA processing, as well as other possible functions of FLASH, remain to be answered. Here, I have attempted to answer a few of these, providing information about the portions of dFLASH required for efficient histone pre-mRNA 3' end formation *in vivo*, for the binding of FLASH to dLsm11, and for localizing dFLASH to the histone locus body. These studies precisely define a site in dLsm11 that is required to bind dFLASH, as well as the site on dFLASH required to bind dLsm11.

FLASH and Histone mRNA Expression

The finding that dFLASH interacts with dLsm11, a known histone pre-mRNA processing factor (Yang *et al.*, 2009), suggested that dFLASH may have a role in histone pre-mRNA processing. However, if this were the case, this factor should have been among the validated hits from the genome-wide screen for factors involved in histone pre-mRNA processing (Wagner *et al.*, 2007; Chapter III of this dissertation). One possibility was that if FLASH was required for transcription as well as processing, the reporter would not have

been expressed and hence FLASH would not have been detected. In fact, dFLASH, then known only as CG4616, was among over 300 factors from the screen identified as preliminary positive hits. However, this factor scored as a weak hit, a 1 on an initial scoring scale of 1 to 3, with 3 being a strong positive like SLBP or Lsm11, 2 being a moderate hit, and 1 being classified as weak, putting it between 100 and 300 in the list of the top 300 hits. Since we now know from my more recent experiments that RNAi depletion of FLASH using the same dsRNA target as in the original screen strongly activates the screening reporter, it seems likely that this factor did not score very well in the screen for technical reasons, most likely related to poor quality or insufficient concentration of dsRNA in that well.

Besides its relatively low score in the initial screen, bioinformatics did not identify a clear mammalian homolog, since there is only a small amount of homology between dFLASH and hFLASH. As a result, although over 100 of the top hits from the screen underwent further validation, based on intensity of scoring and potential interest based on bioinformatics, FLASH was not included in this group.

The results in this chapter suggest that dFLASH plays a role predominantly, if not exclusively in processing histone pre-mRNA *in vivo*, and is not involved in histone gene transcription. There was no difference in the ability of RNAi against FLASH to activate the processing reporter driven by the histone promoter or the actin promoter. In addition, RNAi against FLASH did not affect expression of GFP driven by the histone promoter. The RNAi experiments utilizing the Act_{proc} reporter indicated a functional role of dFLASH in processing in the context of live cells. The GFP signal for FLASH RNAi-depleted cells was similar to the signal from cells depleted of SLBP, putting it in the top tier of genes in activation of the reporter. Subsequent analysis using RT-PCR confirmed the results of the

reporter experiment, that is, that FLASH depletion results in readthrough of the reporter mRNA as well as readthrough of the endogenous histone mRNA. Indeed, FLASH-depleted cells exhibited misprocessed histone message for each histone class. Northern blotting demonstrated that a majority of the endogenous histone H3 mRNA in *Drosophila* cells was polyadenylated when dFLASH was depleted. These results indicated that FLASH likely plays a direct role in formation of the 3' end of the mature histone mRNA. For the core histones, the amount of readthrough for FLASH-depleted cells was similar to that of SLBP-depleted cells. However, for histone H1, the linker histone, there was a reduced level of misprocessed message relative to the SLBP control. The significance of this is not known, though it could be due to the nature of the histone H1 gene, which differs somewhat from the core histone genes and may be regulated in a slightly different way (Isogai *et al.*, 2007).

The RNAi and reporter experiments indicated that dFLASH has some role in the histone pre-mRNA processing reaction. However, though the data are highly suggestive, they do not unequivocally reveal whether this role is direct or indirect. Evidence of the direct biochemical involvement of dFLASH in the cleavage reaction requires isolation of the processing reaction outside of the context of the living cell. A role for FLASH in this reaction has been clearly demonstrated in mammalian cells. *In vitro* processing reactions performed using *Drosophila* Kc cell nuclear extract reveal that antibody targeting the N-terminal region of dFLASH can inhibit the cleavage reaction, suggesting that this region is required for the processing reaction to occur (Yang *et al.*, 2009). Addition of the N-terminal fragment of dFLASH reversed this inhibition (Yang *et al.*, 2009). Taken together, these experiments reveal that dFLASH not only affects processing of histone message in the cell, but that it is directly involved in the biochemical reaction resulting in cleavage.

Another important finding from the RNAi and reporter experiments is that depletion of FLASH does not have any appreciable effect on overall levels of histone message. Based upon RT-PCR data for histones H1, H2A, and H2B, Northern blots on total cell RNA, as well as reporter experiments using histone promoter-driven reporters, there was no effect of FLASH depletion on histone gene transcription or overall histone mRNA levels (Fig.4-1A,C,E,F). This is in contrast to previous reports from mammalian systems suggesting that FLASH has a role in transcription of the histone genes. This interpretation was largely based on the loss of histone mRNA when FLASH was knocked down. Since there is not extensive production of polyadenylated histone mRNA in mammalian cells when processing is inhibited, the reduction in histone mRNA levels in mammalian cells is almost certainly due to the failure to process the pre-mRNA, and not to a reduction in transcription. It is also possible that FLASH function varies between flies and mammals.

Direct Biochemical Interactions between dFLASH and dLsm11

The first suggestion that FLASH might be involved in the processing reaction was made through yeast-two-hybrid experiments aimed at detecting novel factors that interact with the N-terminus of Lsm11 (Yang *et al.*, 2009). The shortest length of FLASH that interacted with Lsm11 was the N-terminal 138 amino acids of the human protein.

Ultimately, the yeast-two-hybrid experiments, in conjunction with GST-pulldown assays, pointed to an interaction between an N-terminal fragment of human FLASH (amino acids 1-178 in flies) and the N-terminal half of human Lsm11 (amino acids 1-153 in flies). Additionally, it was found that deletion of the first forty amino acids of Lsm11, which contain several conserved residues, were essential for the binding interaction. Beyond this,

however, not much more was known about which residues of each protein are required for the interaction. In order to answer this question, I created several deletion and point mutants for both dFLASH and dLsm11.

I found that deleting the region of dFLASH from amino acid 1-105 or from 155-178 had no effect on dLsm11 binding, while deletion of amino acids 1-122 or 144-178 showed significantly reduced binding and deletion of residues 1-141 or 138-178 completely abolished the ability to bind dLsm11. These results defined a region between amino acids 105-154 that contains residues essential for binding.

Within this binding region, I created a number of point mutants in an attempt to define which residues are required for the binding interaction. Ultimately, I found mutating the asparagine and leucine residues at 125-126 (NL125,126) or the two leucine residues at 129 and 130 (LL129,130) abolished binding *in vitro*, while mutation of the glutamate residues at 136 and 141 (E136 and E141, respectively) reduced binding, but did not abolish it. None of the other residues tested had any appreciable effect on binding. These results agree with data from human FLASH generated by Xiao-cui Yang at the same time, which showed that mutation of the leucine and isoleucine residues (LI118,119; corresponding to LL129,130 in *Drosophila*) abolishes binding to Lsm11, while mutation of the lysine and aspartate residues at sites 129 and 130 (KD129,130; KE140,141 in *Drosophila*) reduces binding (X. Yang and Z. Dominski, unpublished results). Together, these results point to the requirement for several key residues within the N-terminus of dFLASH that are absolutely required for binding Lsm11 (Fig. 4-7A).

I also mapped the precise region of Lsm11 required for binding FLASH. I demonstrated that mutations of dLsm11 in the region from 20-35 revealed three distinct

sections of that region required for binding, LDV 20-22, FN 27-28 and LY 32-33. In particular the FN27,28 mutant did not bind Lsm11, while mutation of adjacent amino acids (SD24,25) had no effect (Fig. 4-4C,D,E).

Mutations in the FLASH binding site do not rescue an Lsm11 mutant.

A transgene encoding Lsm11 tagged with V5 at the N-terminus and containing either the SD24,25 mutation or the FN27,28 mutation were created (A. Godfrey and R. Duronio) to allow characterization of the binding interaction in the context of the whole organism. These mutants and wild-type Lsm11 were introduced into a site on chromosome 3 using site-specific recombination. The transgenic flies were crossed with heterozygous flies carrying an Lsm11 null mutation (Godfrey *et al.*, 2009). The wild-type and the SD24,25 dFLASH mutant rescues lethality of the Lsm11 mutant fly, while the FN27,28 mutation does not (P. Gasdaska and R. Duronio, unpublished results). This result supports the biochemical data that the FN27,28 residues are critical for the interaction between dFLASH and dLsm11, and that disruption of this interaction is lethal. It is interesting to note that the Lsm11 null mutant flies have a lethal phenotype, while the U7 snRNA null mutants do not, despite the fact that nearly all of the histone message in U7 snRNA null flies being misprocessed (Godfrey *et al.*, 2007; Godfrey *et al.*, 2009). This suggests that Lsm11 may have an additional role in the cell beyond histone pre-mRNA processing (Godfrey *et al.*, 2009). The implications of this for dFLASH are presently unknown, but may suggest that whatever the cause of lethality, it likely requires the interaction between dFLASH and dLsm11. Whether that involves histone pre-mRNA processing or some other cellular process remains to be seen.

Surprisingly, although removal of the first 40 amino acids from Lsm11 abolished binding to FLASH, these 40 amino acids alone did not bind FLASH. However, inclusion of an additional 38 amino acids restores binding. This suggests that although critical binding residues reside within the first 40 amino acids of dLsm11, there are other required residues. Whether these residues directly contact FLASH or whether they help fold the first 40 amino acids correctly is not known.

The residues required for binding are not continuous with each other but are interrupted by conserved residues that are not required for binding. Indeed, mutation of the serine and aspartate residues at sites 24-25, the proline and leucine residues at sites 29-30, or the proline and asparagine residues at sites 36-37 had no effect on the ability of dLsm11 to bind dFLASH. Together, these results suggest that this binding portion of dLsm11 may adopt a structure in which one face interacts with dFLASH, while the other does not. Whether the non-dFLASH binding face interacts with another protein or has some other function has not been determined.

Localization of dFLASH to the HLB Does Not Require dLsm11 Binding

Because dFLASH binds dLsm11 and both proteins localize to the HLB, it is possible that dFLASH could be recruited to the HLB by the U7 snRNP. Alternatively, it is possible that FLASH is recruited first by another factor and that FLASH may then be responsible for recruitment of the U7 snRNP by binding Lsm11. The results of my experiments reveal that deleting the Lsm11 binding site from FLASH had no effect on the ability of dFLASH to localize to the HLB. Thus, dFLASH localization is not dependent upon its interaction with

dLsm11 and there must be a region of FLASH between amino acids 166 and 733 that is required for recruitment to the HLB.

This indicates that a factor other than dLsm11 is required for dFLASH localization to the HLB. In mammalian cells, FLASH has been reported to interact with at least two other proteins related to histone expression, NPAT and Ars2 (Barcaroli *et al.*, 2006a,b; Kiriyama *et al.*, 2009). The *Mxc* gene product in *Drosophila* is almost certainly the NPAT homologue, since it is a direct target of cyclin E/cdk2 and reacts directly with the MPM-2 antibody (A. White and Duronio, unpublished results). Indeed, when Mxc is depleted by RNAi, dFLASH is delocalized from the HLB (Anne White, unpublished results). Depletion of Mxc affects both the histone processing and transcription reporters (Chapter 2). In the case of Ars2, there has been no evidence for or against binding of these two proteins in flies. It is possible that Mxc is a core factor for formation of the HLB, and that it recruits FLASH. There may also be another unknown factor(s) that participates in HLB formation.

Testing the Functionality of dFLASH *in vivo*

I have established that the portion of the N-terminus of dFLASH (aa 105-154) responsible for binding dLsm11 is not required for localizing dFLASH to the HLB. There are conserved residues in the region upstream of the dLsm11 binding region that may interact with other processing factors. In support of this possibility, deletion of this region (which leaves the dLsm11 binding site intact), inhibits histone pre-mRNA processing *in vivo* in the presence of the wild-type FLASH, consistent with it binding a critical processing factor. A diagram depicting this and other characterized regions of dFLASH and dLsm11 is shown in Figure 4-7.

To identify other regions of FLASH essential for processing, I have taken advantage of the fact that I can deplete endogenous FLASH using RNA I against the C-terminus of the protein, which is not essential for function, and then introduce different FLASH proteins to see if they restore processing (as evidenced by diminished GFP). Thus, I will be able to carry out an *in vivo* "structure-function" analysis of FLASH in cultured cells in the future. A similar approach can also be taken in flies, since there is a mutant of FLASH which produces large amounts of polyadenylated histone mRNA (T. K. Rajendra and Z. Dominski, unpublished). FLASH likely plays important roles in addition to its role in histone pre-mRNA processing, and these will be best elucidated in the intact organism, rather than in cultured cells.

Table 4-1. Oligonucleotides used to create DNA template for transcription of dsRNAs. Shown is a list of oligonucleotides that were used to prime amplification of DNA to make template for transcription. The final amplicon contained T7 promoter sequences on each end, allowing for transcription off of both strands. "Target" indicates the gene being amplified, "Annotation ID" is the CG number for that gene, "Forward" and "Reverse" indicate primers that prime in these orientations with respect to the gene being amplified. Note that the T7 promoter sequence is abbreviated as "T7" for brevity. The T7 promoter sequence is GGTAATACGACTCACTATAGGG.

Target	Forward	Reverse
PTB	T7-GCCCATAGCGACTACAGC	T7-TGGAATGAATTGTTCTTTGTGAA
SLBP	T7-TCCAGTTCCTTGAATAGCAG	T7-AGTCCGCTCGTCCTTTG
FLASH-1	T7-GCAACCCCTGGACAAATTCAAGG	T7-ATCGGGCTTGGCGTATAAT
FLASH-2 (INTERNAL)	T7-CGAAAGTAAGCGTCCGAAAG	T7-ATTCCTGTGATGATCTCGCC
FLASH 5' UTR	T7-ACAAATACACGGCAGGGAAA	T7-CGCTTACGTTTCTTCTTGGC
FLASH 3' UTR-1	T7-CCGCCGATGATAAGGAATAA	T7-ATGGGAACGTTTGAGGTCAG
FLASH 3' UTR-2	T7-AATGATTGGGTGAATAGGTG	T7-TCTTTTCCAGAAATCCAAAA
FLASH C-TERM.	T7-GATCCACTAAGATCCCAAACG	T7-AATTGGTTTGTCCGCAGCAGT

Table 4-2. Oligonucleotides used to create GST-FLASH constructs. Shown is a list of oligonucleotides used to create deletion and point mutants of FLASH fused to glutathione-S-transferase (GST) for use in GST pulldown assays. "dFLASH Amplicon" indicates the region of dFLASH that is amplified (for deletion mutants) or the mutation introduced into a fragment of dFLASH containing amino acids 1-178 (for point mutants). For point mutants, 5' indicates the primer pair used to amplify the 5' end of the mutated dFLASH and 3' indicates the primer pair used to amplify the 3' end of the mutated dFLASH. After this first round of PCR, these fragments were used as template for overlap PCR reactions containing forward primer 5'-GGCCGAATTCATGGAAACGCCTGCATATG-3' and reverse primer 5'-GGCCCTCGAGTCATTTCGGACGCTTACTTTC-3' to make a dFLASH fragment from amino acids 1-178 containing the desired point mutations. "Forward" and "Reverse" indicate primers that prime dFLASH in these orientations with respect to the 5' and 3' ends of the gene.

dFLASH Amplicon	Forward	Reverse
1-131	GGCCGAATTCATGGAAACGCCTG CATATG	GGCCCTCGAGTCAGTCCAGTAGATTCTGTAGATT AACCTC
1-137	GGCCGAATTCATGGAAACGCCTG CATATG	GGCCCTCGAGTACCTCCGCCTTAGCTGTGTC
1-143	GGCCGAATTCATGGAAACGCCTG CATATG	GGCCCTCGAGCAGTGTCTCTTTTCTCTTTACCTC
1-154	GGCCGAATTCATGGAAACGCCTG CATATG	GGCCCTCGAGTCATACATCGTCCTTTTCGTTTCG
1-178	GGCCGAATTCATGGAAACGCCTG CATATG	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
67-178	GGCCGAATTCAGATCCCTTGAACT GGACATCTAC	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
78-178	CGCGGAATTCGACGACTTTCAGA AGGCCGA	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
105-400	GGCCGAATTCATGATTGAAGCGC TAAAGGTTGAGAATAAG	GGCCCTCGAGTCACTTCGTCTCTACTTTTGATTCC TTTGA
122-400	GGCCGAATTCATGGAGGTTAATC TACAGAATCTACTG	GGCCCTCGAGTCACTTCGTCTCTACTTTTGATTCC TTTGA
141-400	CGCGGAATTCATGGAGACACTGA TTGCCGAGC	GGCCCTCGAGTCACTTCGTCTCTACTTTTGATTCC TTTGA
166-400	CGCGGAATTCATGGTACCAGGTG CAAGGGAAC	GGCCCTCGAGTCACTTCGTCTCTACTTTTGATTCC TTTGA
NL125, 126AA 5'	GGCCGAATTCATGGAAACGCCTG CATATG	GTCCAGTAGATTCTGAGCAGCAACCTCCATGGTT TT
NL125, 126AA 3'	AAAACCATGGAGGTTGCTGCTCA GAATCTACTGGAC	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
LL129, 130AA 5'	GGCCGAATTCATGGAAACGCCTG CATATG	CGCCTTAGCTGTGTCAGCAGCATTCTGTAGATTA AC
LL129, 130AA 3'	GTTAATCTACAGAATGCTGCTGAC ACAGCTAAGGCG	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
E136A 5'	GGCCGAATTCATGGAAACGCCTG CATATG	CTCTTTTCTCTTTACAGCCGCCTTAGCTGTGTC
E136A 3'	GACACAGCTAAGGCGGCTGTAAA GAGAAAAGAG	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
RK139, 140AA 5'	GGCCGAATTCATGGAAACGCCTG CATATG	GGCAATCAGTGTCTCAGCAGCCTTTACCTCCGCC TT
RK139, 140AA 3'	AAGGCGGAGGTAAAGGCTGCTGA GACACTGATTGCC	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
E141A 5'	GGCCGAATTCATGGAAACGCCTG CATATG	CTCGGCAATCAGTGTAGCTTTTCTCTTTACCTC
E141A 3'	GAGGTAAAGAGAAAAGCTACACT GATTGCCGAG	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
I144A 5'	GGCCGAATTCATGGAAACGCCTG CATATG	GTTTCGCAGCTCGGCAGCCAGTGTCTCTTTTCT
I144A 3'	AGAAAAGAGACACTGGCTGCCGA GCTGCGAAAC	GGCCCTCGAGTCATTTCGGACGCTTACTTTC
L147A 5'	GGCCGAATTCATGGAAACGCCTG CATATG	ATCGTCCTTTTCGTTTCGAGCCTCGGCAATCAGT GT
L147A 3'	ACACTGATTGCCGAGGCTCGAAA CGAAAAGGACGAT	GGCCCTCGAGTCATTTCGGACGCTTACTTTC

Table 4-3. Oligonucleotides used to create GST-Lsm11 constructs. Shown is a list of oligonucleotides used to create deletion and point mutants of dLsm11 fused to glutathione-S-transferase (GST) for use in GST pulldown assays. "dLsm11 Amplicon" indicates the region of dLsm11 that is amplified (for deletion mutants) or the mutation introduced into full length dLsm11 (for point mutants). For point mutants, 5' indicates the primer pair used to amplify the 5' end of the mutated dFLASH and 3' indicates the primer pair used to amplify the 3' end of the mutated dFLASH. After this first round of PCR, these fragments were used as template for overlap PCR reactions containing forward primer 5'-GGCCGAATTCATGGAATCGAGGGACCGGA-3' and reverse primer 5'-GGCCCTCGAGCTATTTTGTTGGTATTAAACTAACTAGAACCAC-3' to make full length dLsm11 containing the desired point mutations. "Forward" and "Reverse" indicate primers that prime dLsm11 in these orientations with respect to the 5' and 3' ends of the gene.

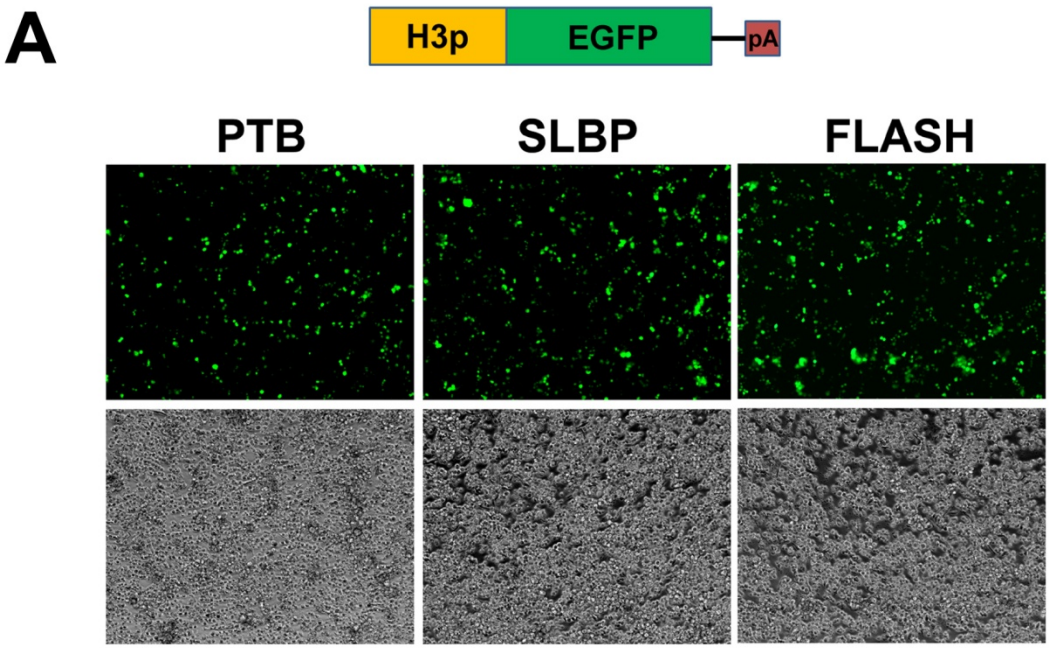
dLsm11 Amplicon	Forward	Reverse
1-40	GGCCGAATTCATGGAATCGAGGGACCGGA	GGCCCTCGAGTCACACCCTAAAATTGGGTTCGTATAACGCT
1-54	GGCCGAATTCATGGAATCGAGGGACCGGA	GGCCCTCGAGTCAGGCAGCCAGATTCTGGTAGATGTC
1-78	GGCCGAATTCATGGAATCGAGGGACCGGA	GGCCCTCGAGTCATCCTTCTGCGGATCCCGGCTTTTG
1-100	GGCCGAATTCATGGAATCGAGGGACCGGA	GGCCCTCGAGTCATCGCTCAGGAGGTAGAATATCTAC
1-153	GGCCGAATTCATGGAATCGAGGGACCGGA	GGCCCTCGAGTCACCGAACTCTTCTCCTTTCGTTGGA
FULL LENGTH	GGCCGAATTCATGGAATCGAGGGACCGGA	GGCCCTCGAGCTATTTTGTTGGTATTAAACTAACTAGAACCAC
LDV20,21,22AAA 5'	GGTTCTTTGTTCTATCAGACTTC	AAACCTATCACTGCCAGCAGCAGCTTCGGAGATTTCCGA
LDV20,21,22AAA 3'	TCGGAAATCTCCGAAGCTGCTGCTGGCAGTGATAGGTTT	GGCCCTCGAGCTATTTTGTTGGTATTAAACTAACTAGAACCAC
SD24, 25AA 5'	GGTTCTTTGTTCTATCAGACTTC	CAGAGGATTAAACCTAGCAGCGCCCACATCCAATTC
SD24, 25AA 3'	GAATTGGATGTGGGCGCTGCTAGGTTTAATCCTCTG	GGCCCTCGAGCTATTTTGTTGGTATTAAACTAACTAGAACCAC
FN27, 28AA 5'	GGTTCTTTGTTCTATCAGACTTC	TAACGCTCGCAGAGGAGCAGCCCTATCACTGCCCAC
FN27, 28AA 3'	GTGGGCAGTGATAGGGCTGCTCCTCTGCGAGCGTTA	GGCCCTCGAGCTATTTTGTTGGTATTAAACTAACTAGAACCAC
PL29, 30AA 5'	GGTTCTTTGTTCTATCAGACTTC	TTCGTATAACGCACGAGCAGCATTAAACCTATCACT
PL29, 30AA 3'	AGTGATAGGTTTAATGCTGCTCGTGCGTTATACGAA	GGCCCTCGAGCTATTTTGTTGGTATTAAACTAACTAGAACCAC
LY33, 34AA 5'	GGTTCTTTGTTCTATCAGACTTC	CCTAAAATTGGGTTCAGCAGCCGCTCGCAGAGGATTAAA
LY33, 34AA 3'	TTTAATCCTCTGCGAGCGGCTGCTGAACCCAATTTTAGG	GGCCCTCGAGCTATTTTGTTGGTATTAAACTAACTAGAACCAC
PN36, 37AA 5'	GGTTCTTTGTTCTATCAGACTTC	ATCGGTCACCCTAAAAGCAGCTTCGTATAACGCTCG
PN36, 37AA 3'	CGAGCGTTATACGAAGCTGCTTTTAGGGTGACCGAT	GGCCCTCGAGCTATTTTGTTGGTATTAAACTAACTAGAACCAC

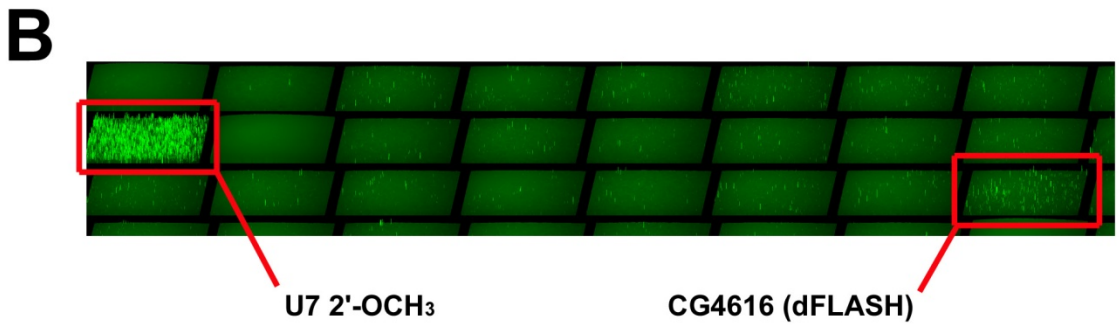
Table 4-4. Oligonucleotides used to create pIZ-FLASH constructs. Shown are oligonucleotides used to create *Drosophila* expression constructs using the pIZ/V5/His vector (Invitrogen). "dFLASH Amplicon" indicates the region of dFLASH that is amplified. "dFLASHp + 5' UTR" is the dFLASH promoter and 5' UTR. "Forward" and "Reverse" indicate primers that prime dFLASH in these orientations with respect to the 5' and 3' ends of the gene.

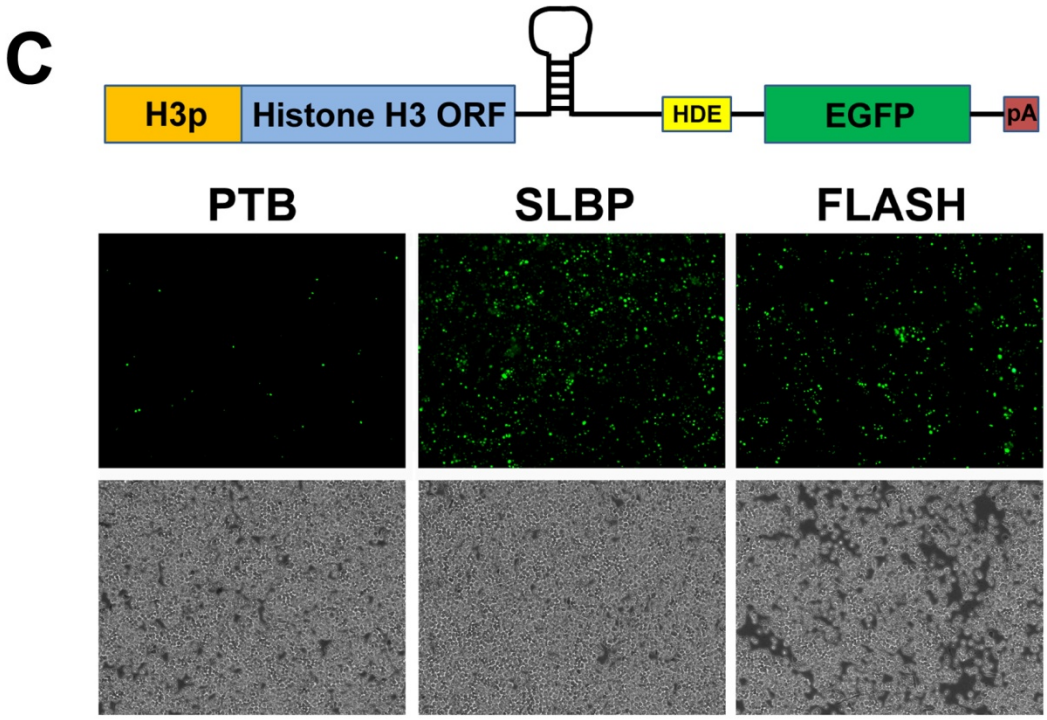
dFLASH Amplicon	Forward	Reverse
dFLASH p + 5' UTR	GGCCAAGCTTCGAGTATTATTGTTTGTATTGTGATGT	GGCCGAATTCACCAATCGCTTACGTTTCTT
1-844	GGCCGAATTCCAAAATGGAAACGCCTGCATATG	GAACCCGCGGAATTGGTTTGTCAGCAGCA
105-844	GGCCGAATTCCAAAATGATTGAAGCGCTAAAGGTTGAGAATAAG	GAACCCGCGGAATTGGTTTGTCAGCAGCA
166-844	GGCCGAATTCCAAAATGGTACCAGGTGCAAGGGAAC	GAACCCGCGGAATTGGTTTGTCAGCAGCA
1-733	GGCCGAATTCCAAAATGGAAACGCCTGCATATG	GAACCCGCGGAGCCTGTTTTGGAGTCTGGGTGAG
78-733	GGCCGAATTCCAAAATGGACGACTTTCAGAAGGCCGA	GAACCCGCGGAGCCTGTTTTGGAGTCTGGGTGAG

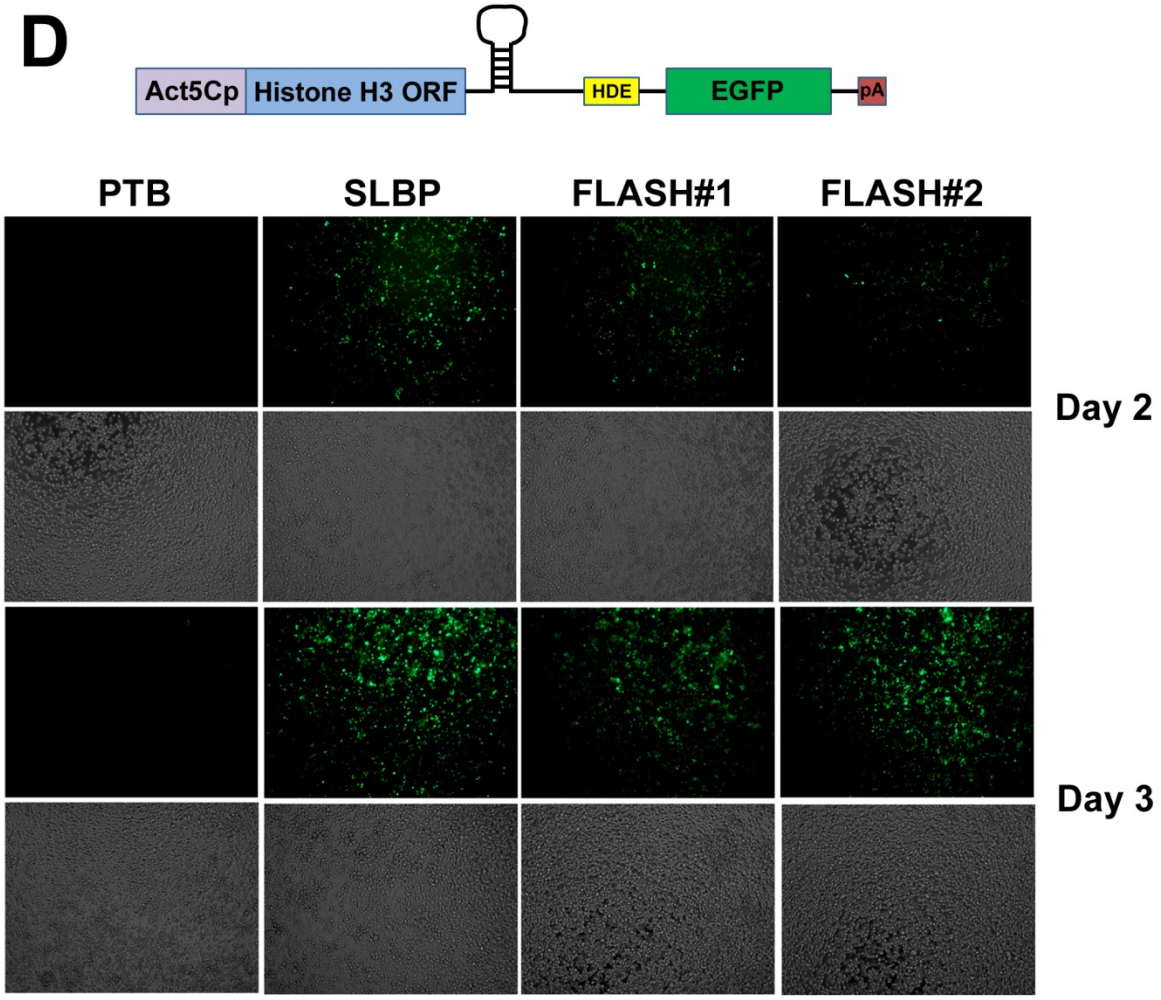
Figure 4-1. *Drosophila* FLASH is required for processing histone pre-mRNA but not for transcribing it. (A) *Drosophila* D.Mel-2 cells were treated with dsRNAs targeting PTB, SLBP, or FLASH, transfected with the His_{trans} reporter (model shown at top), and viewed on Day 5 by fluorescence microscopy to detect GFP expression. (B) Image of wells from the genome-wide screen for factors that participate in histone pre-mRNA processing. The postive control well containing a 2'-OCH_3 oligonucleotide targeting dU7 snRNA and the experimental well containing dsRNA targeting CG4616 (dFLASH) are indicated by red boxes. (C) D.Mel-2 cells were treated with dsRNAs targeting PTB, SLBP, or FLASH, transfected with His_{proc} reporter (model shown at top), and viewed on Day 5 by fluorescence microscopy to detect GFP. (D) D.Mel-2 cells stably expressing the Act_{proc} reporter (model shown at top), were treated with dsRNAs targeting PTB, SLBP, or one of two independent regions of FLASH, then viewed by fluorescence microscopy on Days 2 and 3 to detect GFP. (E) Total RNA from the Day 3 cells shown in panel D was isolated and amplified by PCR with (+) or without (-) prior reverse transcription. RT-PCR targets were SLBP and FLASH (to show knockdown), misprocessed reporter, total histone H1, H2A, and H2B (to show total mRNA levels), misprocessed histone H1, H2A, and H2B (to show improperly cleaved mRNA), and tubulin (loading control). A model of the histone-specific primer locations is shown at the bottom. Note that p1 and p2 are forward and reverse primers, respectively, for detecting total histone message, while p1' and p2' are forward and reverse primers, respectively, for detecting misprocessed histone message. The normal cleavage site is depicted by a red slash just after the stem loop. (F) Northern blot analysis of total RNA taken from cells RNAi-depleted of (G) *In vitro* 3' end processing of *Drosophila* H3 (dH3)

histone pre-mRNA in a nuclear extract from Kc cells. Processing was carried out under control conditions (lanes 2 and 5) or in the presence of the indicated antibodies (lanes 3, 4, and 6) and/or 1 µg of the N-terminal dFLASH protein fused to GFP (lanes 7 and 8). Lane 1 contains the input dH3 substrate. Abbreviations: PTB = pyrimidine tract binding protein; SLBP = stem loop binding protein; GFP = green fluorescent protein; Act_{proc} = Actin 5C promoter-driven histone pre-mRNA processing reporter. (panels F and G courtesy of Xiao-cui Yang and Zbigniew Dominski; Yang *et al.*, 2009)

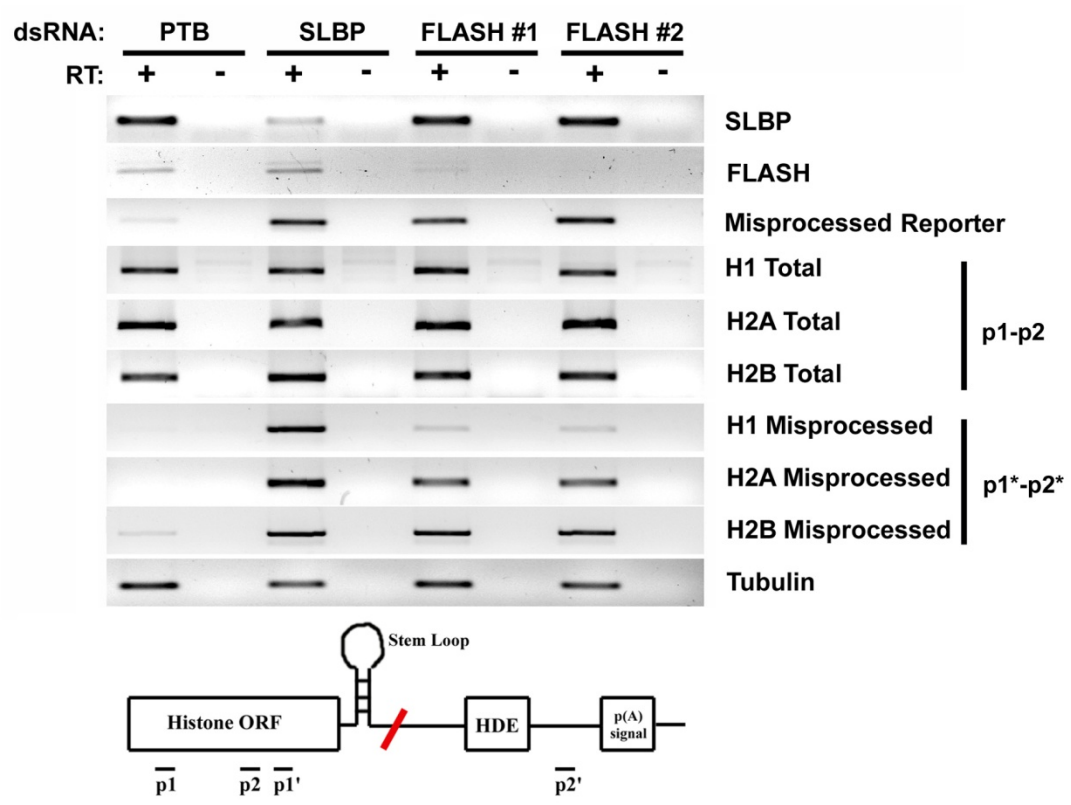

A
H3p
EGFP
pA
PTB
SLBP
FLASH


B
U7 2'-OCH3
CG4616 (dFLASH)


C
H3p
Histone H3 ORF
HDE
EGFP
pA
PTB
SLBP
FLASH


D
Act5Cp
Histone H3 ORF
HDE
EGFP
pA
PTB
SLBP
FLASH#1
FLASH#2
Day 2
Day 3

E


dsRNA:
PTB
SLBP
FLASH #1
FLASH #2
RT:
+ - + - + - + -
SLBP
FLASH
Misprocessed Reporter
H1 Total
H2A Total
H2B Total
p1-p2
H1 Misprocessed
H2A Misprocessed
H2B Misprocessed
p1*-p2*
Tubulin
Stem Loop
Histone ORF
HDE
p(A) signal
p1
p2
p1'
p2'

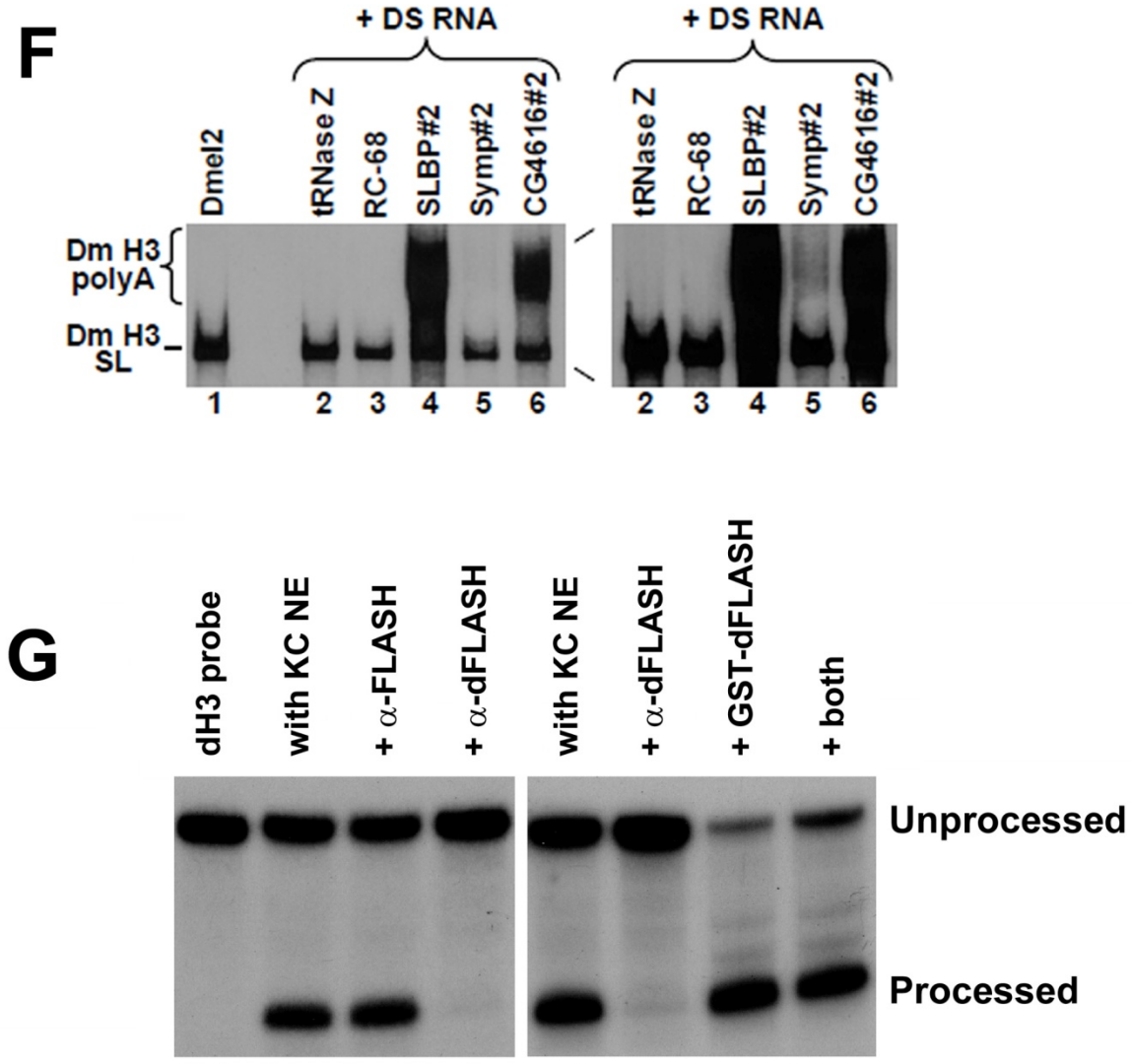

F
+ DS RNA
+ DS RNA
Dmel2
tRNase Z
RC-68
SLBP#2
Symp#2
CG4616#2
tRNase Z
RC-68
SLBP#2
Symp#2
CG4616#2
Dm H3 polyA
Dm H3 SL
1
2
3
4
5
6
2
3
4
5
6
G
dH3 probe
with KC NE
+ α-FLASH
+ α-dFLASH
with KC NE
+ α-dFLASH
+ GST-dFLASH
+ both
Unprocessed
Processed

Figure 4-2. *Drosophila* FLASH shares a limited sequence similarity with human FLASH. (A) A comparison of the amino acid sequences of human FLASH (Hs) and *Drosophila* FLASH (Dm). (B) A comparison of the amino acid sequences of human Lsm11 (Hs) and *Drosophila* Lsm11 (Dm). The first 40 amino acids containing a region conserved between the two species are denoted by a double-headed arrow and the highly conserved region between *Drosophila* amino acids 20 and 37 is highlighted. (Figure adapted from Yang *et al.*, 2009)

A

FLASH

```
Hs   30  LDIYAGLDSAVSDSASKSCVPSRNC-LDLYEEILTEEGTAKEATYNDLQVEYGKCQLQMK  88
         LDIY  L       + K     R+  LD+Y++ L +   A++    +LQ    K +  +
Dm   45  LDIYHDLYDIEPAGSKKPDQMDRSLELDIYDD-LDDFQKAEDRNTKELQAWEAKYEKALA 103

Hs   89  ELMKKFKEIQTQNFSLINENQSLKKNISALIKTARVEINRKDEEISNLHQRLSEFPHFRN 148
         E+    + ++ +N +L  + ++++ N+  L+ TA+ E+ RK+  I+ L     +   FR
Dm  104  EI----EALKVENKALGKKIKTMEVNLQNLLDTAKAEVKRKETLIAELRNEKDDV-CFR- 157

Hs  149  NHKTARTFDTVKTKDLKSRSP 169
           K AR FD    ++ +S+ P
Dm  158  -RKRARAFDVPGAREPESKRP 177
```

B

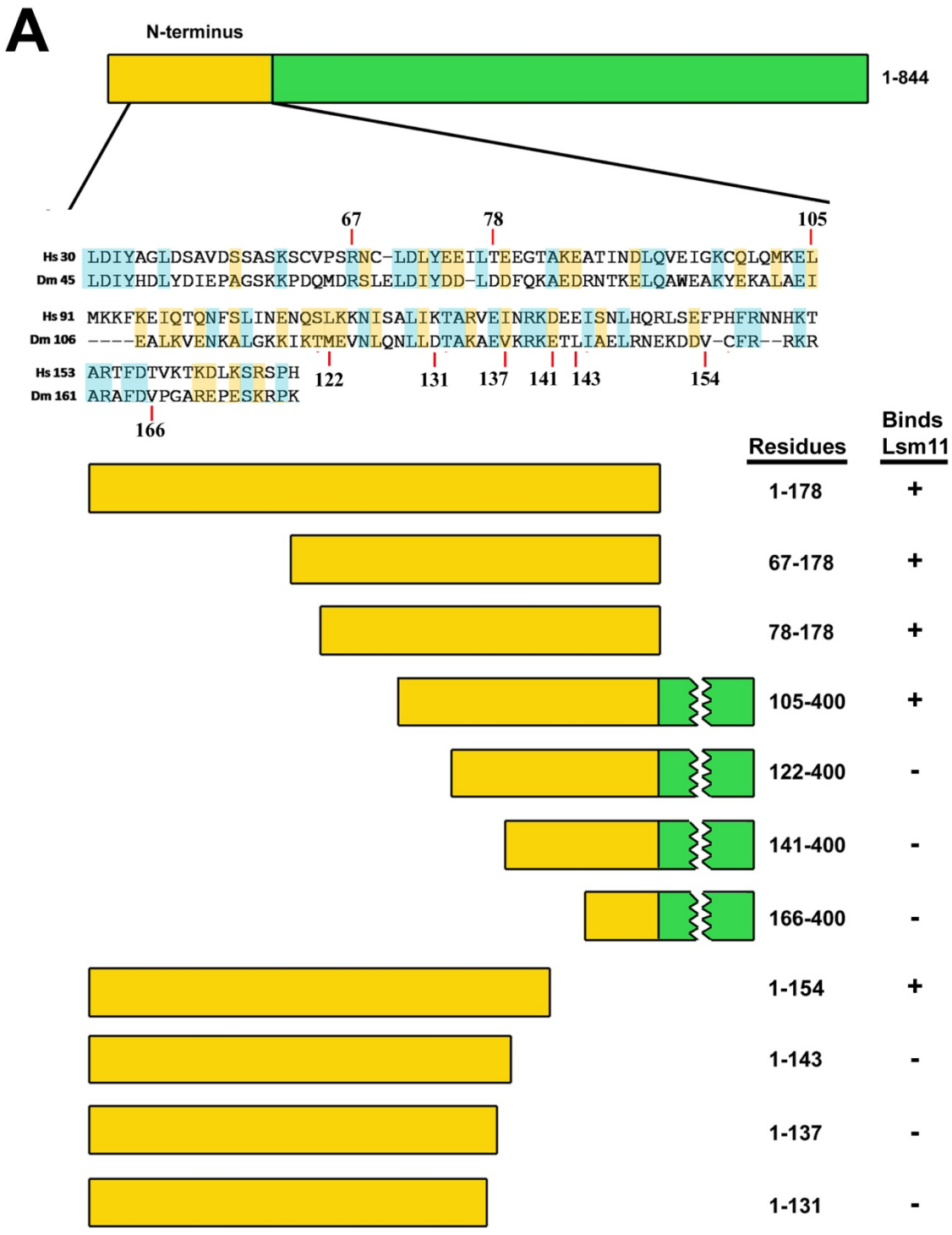
Lsm11

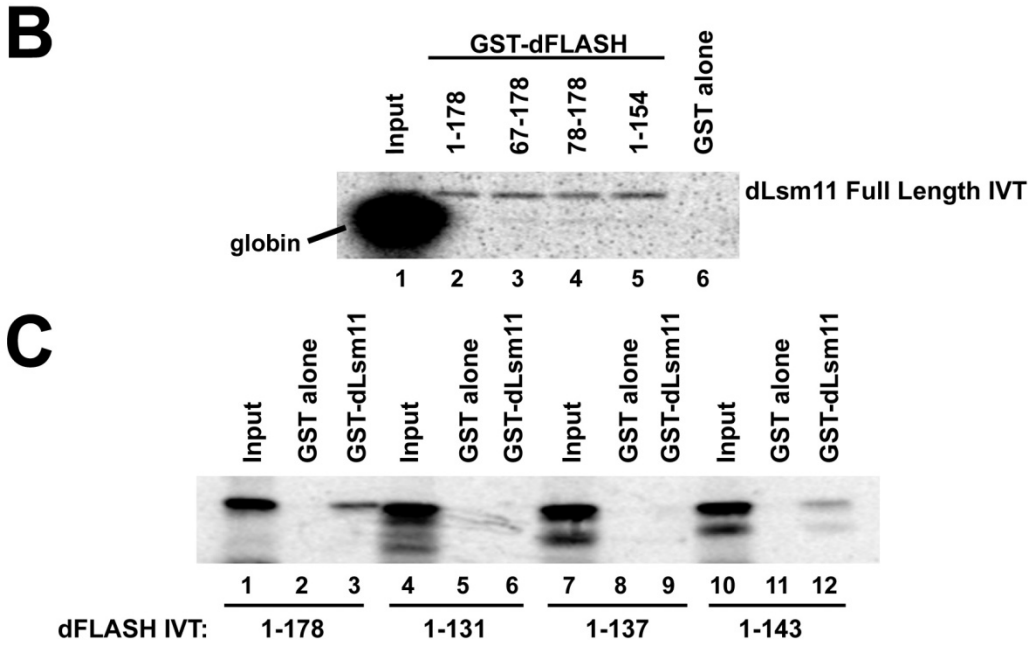
40-amino acid region with conserved residues

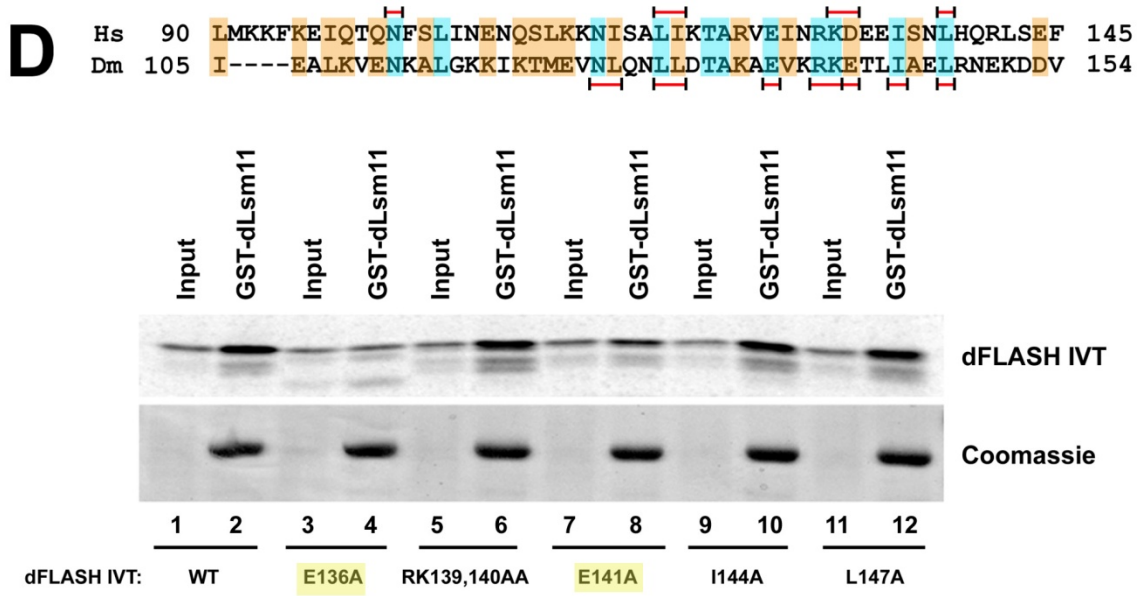
```
Hs    1  MEERERGARSAGAGSPARPPSPRLDVSSDSFDPLLALYAPRLPPIPYPNAPCFNNVAEYE  60
         ME R+R   +    S         LDV SD F+PL ALY P            + N+A +E
Dm    1  MESRDRKTSNEDDSSEI----SELDVGSDRFNPLRALYEPNFRVTDAVPKVIYQNLAAFE  56

Hs   61  SFLRTGVRGGGRGRGRAR-GAAAGSGVPAAPGPSGRTRRRPDAPAPD---PERIQRLRRL 116
         S L+    G  +   R + G+A G       G  G +++   + A D   PER
Dm   57  SALKKF--GIWQLNKRQKPGSAEG-------GDQGTSKKASTSEAVDILPPER------- 100

Hs  117  MVAKEEGDGAAGAGRRGPGRSRKAPRNVLTRMPLHEGSPLGELHRCI--------REGVK 168
            + E         R     +K  RN+ T M    G PL  L +CI         E ++
Dm  101  ---RFEPHQMPTTSSR-----KKHHRNIFTYMEAAVG-PLELLTKCISPASHNESNERIR 151
```

Figure 4-3. Biochemical characterization of the region of dFLASH that binds dLsm11. (A) Model of the dFLASH N-terminus showing sequence conservation between flies and mammals, with blue boxes indicating perfectly conserved residues and orange boxes indicating residues with similar characteristics. Numbers represent *Drosophila* residues used in deletion mutants created for GST pulldown assays. A diagram of the dFLASH deletion mutants is shown below. Orange bars correspond to N-terminal residues (amino acids 1-178) and green bars to downstream residues. The ability of each mutant to bind dLsm11, determined by experiments depicted in panels B-E of this figure, is shown, with + indicating binding and – indicating failure to bind. (B) Full length dLsm11 was *in vitro* synthesized and labeled with ^{35}S, followed by incubation with GST-FLASH fusion proteins containing the full N-terminus (AA1-178) or containing the N-terminus with either N- (AA67-178, AA78-178) or C-terminal (AA1-154) residues deleted. GST alone was used as a negative control to show any non-specific binding. The ^{35}S-labeled dLsm11 bound to GST-FLASH or GST was bound by glutathione agarose beads, separated on a 12 % SDS-PAGE gel and detected by autoradiography. Input = 10 % of radiolabeled protein used in the assay (C) dFLASH proteins containing either the full N-terminus (1-178) or the N-terminus with C-terminal deletions (AA1-131, AA1-137, AA1-143) were *in vitro* synthesized and labeled with ^{35}S, followed by incubation with GST-Lsm11 fusion proteins containing the first 153 amino acids of dLsm11. GST alone was used as a negative control. The ^{35}S-labeled dFLASH bound to GST-Lsm11 or GST was bound by glutathione agarose beads, separated by electrophoresis, and detected by autoradiography. Input = 5 % of radiolabeled protein used in the assay. (D) Diagram of FLASH sequence conservation between human and flies with the location of point mutations examined in panels D-F underlined (top). dFLASH proteins containing

amino acids 1-178 with the wild type (WT) amino acid sequence or containing point mutations were *in vitro* synthesized and ^{35}S-labeled and incubated with GST-dLsm11 fusion proteins containing the first 153 amino acids of dLsm11. The radiolabeled dFLASH proteins bound to GST-dLsm11 were bound by glutathione agarose beads, separated by electrophoresis, and detected by autoradiography (middle). Note that the GST-dLsm11 was combined with the other reagents and added to each radiolabeled dFLASH protein as a master mix. A Coomassie stain of the SDS-PAGE gel is shown to verify equal loading (bottom). (E) As in Panel B, *in vitro* synthesized and ^{35}S-labeled dLsm11 (AA1-153) was combined with GST-dFLASH deletion and point mutants, bound by glutathione agarose beads, separated by electrophoresis, and detected by autoradiograhy (top). A Coomassie stain of the SDS-PAGE gel is shown as a loading control (bottom). (F) The N-terminus of human Lsm11 up to the first Sm binding site was *in vitro* synthesized and labeled with ^{35}S and combined with GST-hFLASH fusion proteins containing wild type amino acid sequence (WT) or point mutations. Abbreviations: GST = glutathione-S-transferase; SDS-PAGE = sodium dodecyl sulfate polyacrylamide gel electrophoresis. (panel F courtesy of Xiao-cui Yang and Zbigniew Dominski)


A
N-terminus
1-844
67
78
105
Hs 30 LDIYAGLDSAVDSSASKSCVPSRNC-LDLYEEILTEEGTAKEATINDLQVEIGKCQLQMKEL
Dm 45 LDIYHDLYDIEPAGSKKPDQMDRSLELDIYDD-LDDFQKAEDRNTKELQAWEAKYEKALAEI
Hs 91 MKKFKEIQTQNFSLINENQSLKKNISALIKTARVEINRKDEEISNLHQRLSEFPHFRNNHKT
Dm 106 ----EALKVENKALGKKIKTMEVNLQNLLDTAKAEVKRKETLIAELRNEKDDV-CFR--RKR
Hs 153 ARTFDTVKTKDLKSRSPH
Dm 161 ARAFDVPGAREPESKRPK
122
131
137
141
143
154
166
Residues
Binds Lsm11
1-178 +
67-178 +
78-178 +
105-400 +
122-400 -
141-400 -
166-400 -
1-154 +
1-143 -
1-137 -
1-131 -


B
GST-dFLASH
Input
1-178
67-178
78-178
1-154
GST alone
dLsm11 Full Length IVT
globin
1 2 3 4 5 6
C
Input
GST alone
GST-dLsm11
1 2 3 4 5 6 7 8 9 10 11 12
dFLASH IVT:
1-178
1-131
1-137
1-143


D
Hs 90 LMKKFKEIQTQNFSLINENQSLKKNISALIKTARVEINRKDEEISNLHQRLSEF 145
Dm 105 I----EALKVENKALGKKIKTMEVNLQNLLDTAKAEVKRKETLIAELRNEKDDV 154
Input
GST-dLsm11
dFLASH IVT
Coomassie
1 2 3 4 5 6 7 8 9 10 11 12
dFLASH IVT:
WT
E136A
RK139,140AA
E141A
I144A
L147A

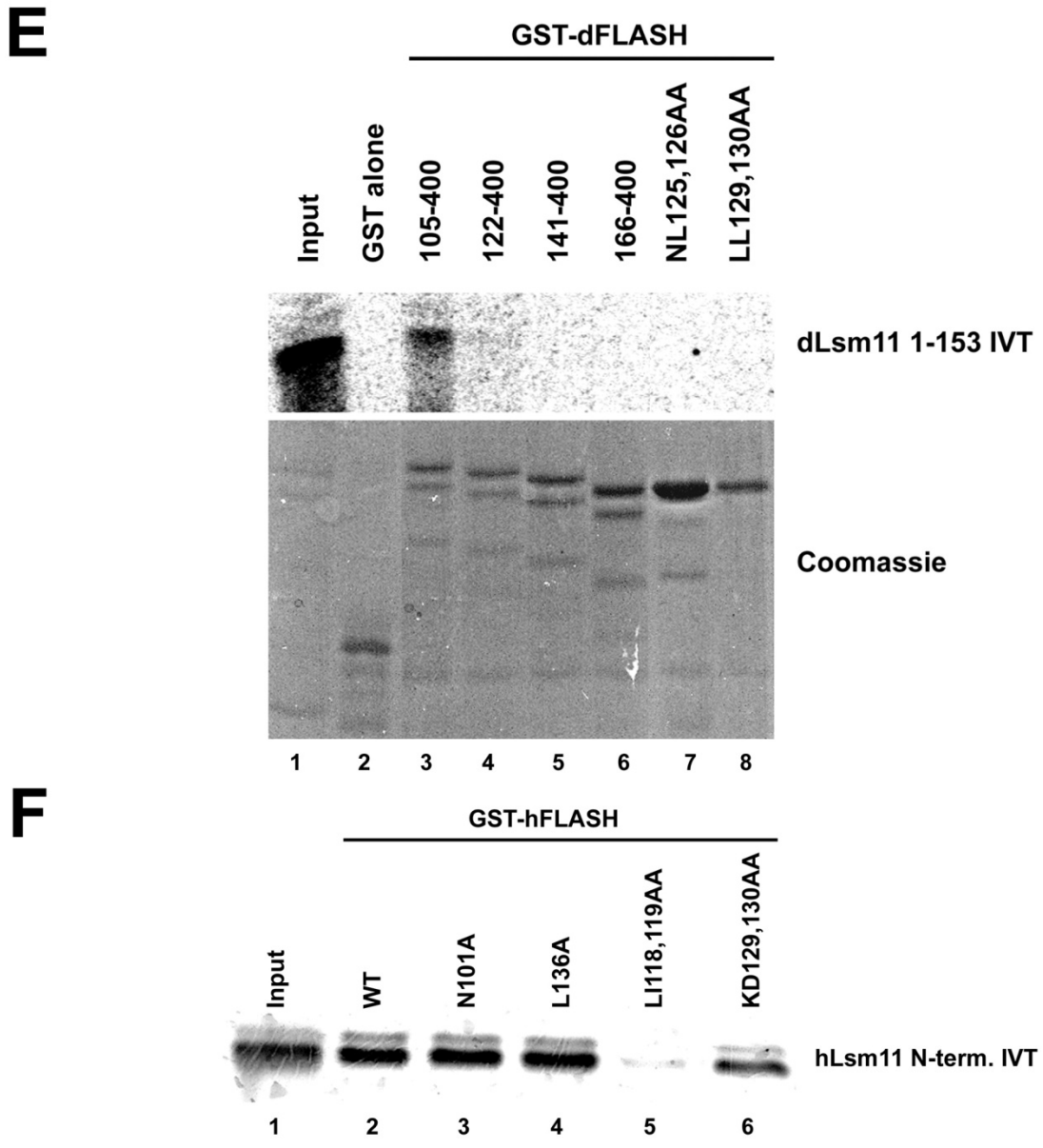
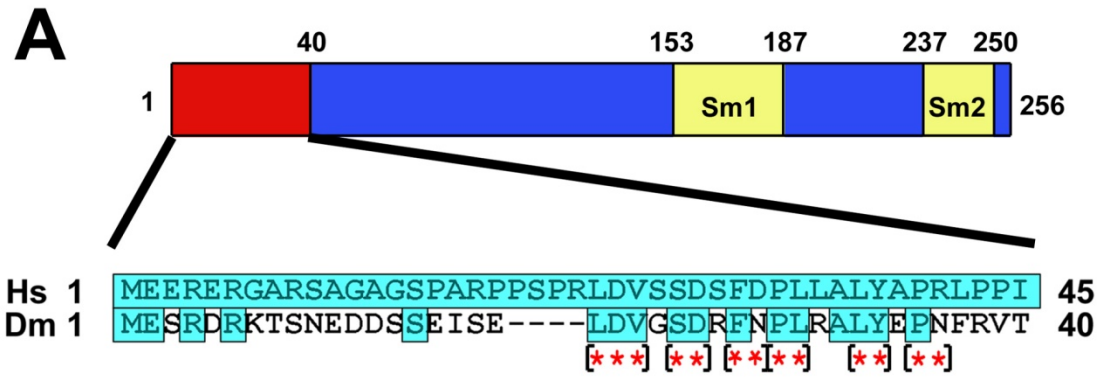
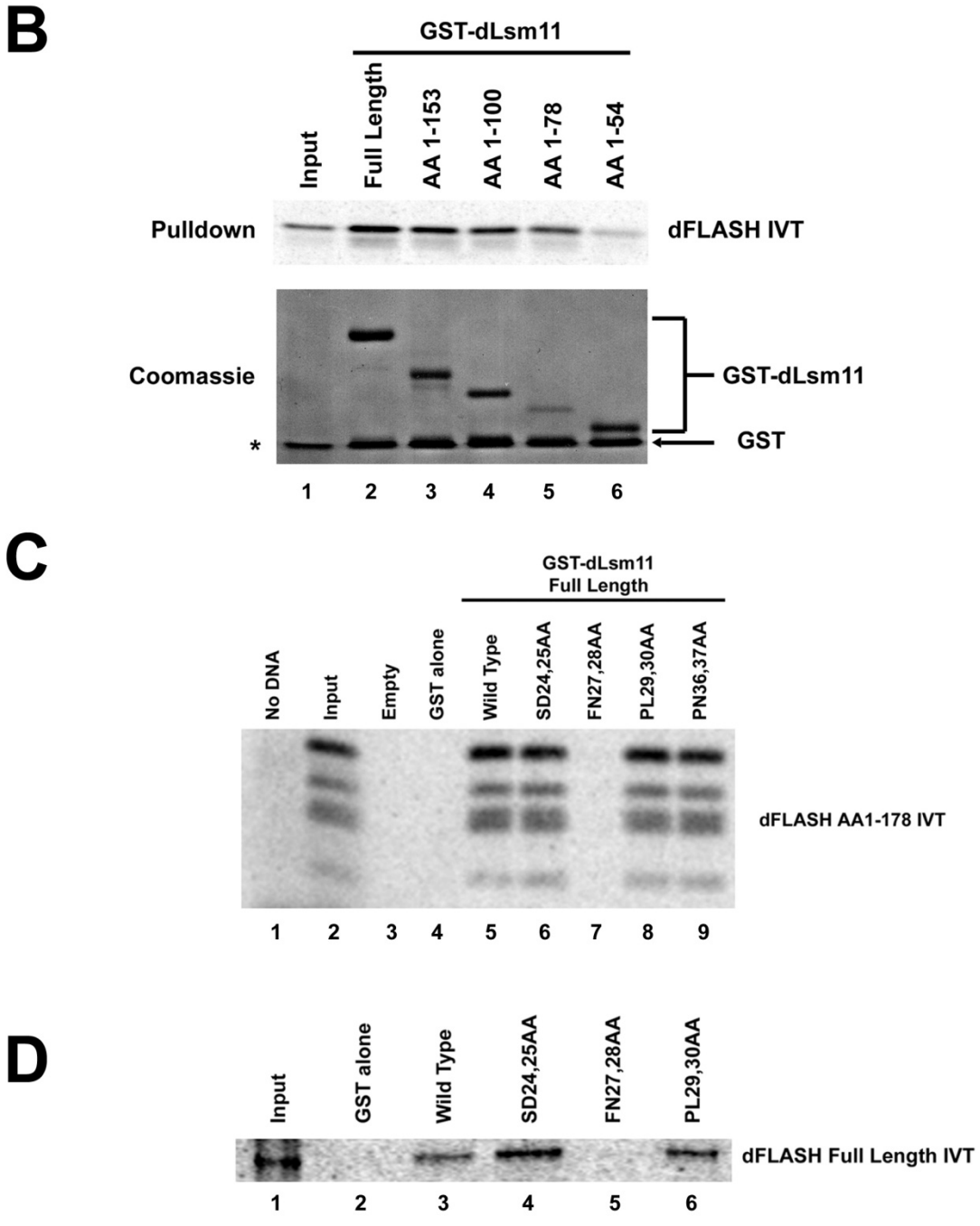
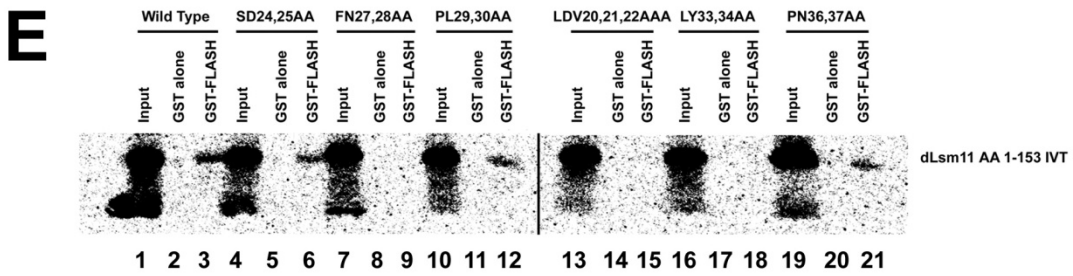

E
GST-dFLASH
Input
GST alone
105-400
122-400
141-400
166-400
NL125,126AA
LL129,130AA
dLsm11 1-153 IVT
Coomassie
1
2
3
4
5
6
7
8
F
GST-hFLASH
Input
WT
N101A
L136A
LI118,119AA
KD129,130AA
hLsm11 N-term. IVT
1
2
3
4
5
6

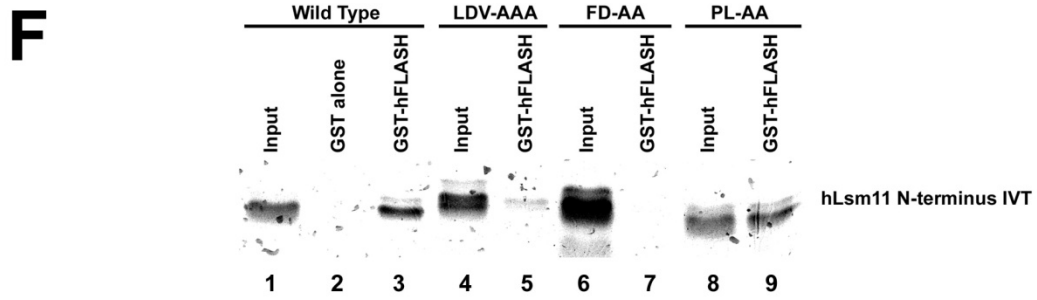
Figure 4-4. Biochemical characterization of the region of dLsm11 that binds dFLASH. (A) Schematic of dLsm11 with a depiction of the conserved residues within the first 40 amino acids of the protein. Asterisks indicate residues that were altered to make point mutants for GST pulldown assays. (B) dFLASH protein consisting of the first 178 amino acids of the N-terminus was *in vitro* synthesized and labeled with ^{35}S and incubated with GST-dLsm11 fusion proteins, either full length or C-terminal deletions. The labeled dFLASH bound to GST-dLsm11 was bound by glutathione agarose beads, separated by electrophoresis, and detected by autoradiography. A Coomassie stain of the gel is shown below to indicate loading. (C) dFLASH protein consisting of the first 178 amino acids of the N-terminus was *in vitro* synthesized and labeled with ^{35}S and incubated with GST-dLsm11 fusion proteins, either full length or corresponding to the first 40 amino acids of the N-terminus, that contain either wild type sequence or point mutations. GST alone is included as a control for non-specific binding. The labeled dFLASH bound to GST-dLsm11 was bound by glutathione agarose beads, separated by electrophoresis, and detected by autoradiography. (D) Full length dFLASH protein was *in vitro* synthesized and labeled with ^{35}S and incubated with GST-dLsm11 fusion proteins that contain either wild type sequence or point mutations. GST alone is included to control for non-specific binding. The labeled dFLASH bound to GST-dLsm11 was bound by glutathione agarose beads, separated by electrophoresis, and detected by autoradiography. (E) dLsm11 protein corresponding to the first 153 amino acids of the N-terminus containing either wild type sequence or point mutations was *in vitro* synthesized and labeled with ^{35}S and incubated with GST-FLASH fusion protein consisting of the first 178 amino acids of its N-terminus. GST alone is included as a control for non-specific binding. Radiolabeled dLms11 bound to GST-FLASH was bound by glutathione

agarose beads, separated by electrophoresis, and detected by autoradiography. (F) Human Lsm11 protein corresponding to the N-terminus up to the first Sm binding site was *in vitro* synthesized and labeled with ^{35}S and incubated with the N-terminal portion of hFLASH fused to GST. GST alone is included as a control for non-specific binding. Radiolabeled hLsm11 bound to GST-hFLASH or GST was bound by glutathione agarose beads, separated by electrophoresis, and detected by autoradiography. GST = glutathione-S-transferase (Panel 4F courtesy of Xiao-cui Yang and Zbigniew Dominski)


A
40
153
187
237
250
1
Sm1
Sm2
256
Hs 1 MEERERGARSAGAGSPARPPSPRLDVSSDSFDPLLALYAPRLPPI 45
Dm 1 MESRDRKTSNEDDSSEISE----LDVGSDRFNPLRALYEPNFRVT 40


B
GST-dLsm11
Input
Full Length
AA 1-153
AA 1-100
AA 1-78
AA 1-54
Pulldown
dFLASH IVT
Coomassie
GST-dLsm11
GST
*
1 2 3 4 5 6
C
GST-dLsm11
Full Length
No DNA
Input
Empty
GST alone
Wild Type
SD24,25AA
FN27,28AA
PL29,30AA
PN36,37AA
dFLASH AA1-178 IVT
1 2 3 4 5 6 7 8 9
D
Input
GST alone
Wild Type
SD24,25AA
FN27,28AA
PL29,30AA
dFLASH Full Length IVT
1 2 3 4 5 6


E
Wild Type
SD24,25AA
FN27,28AA
PL29,30AA
LDV20,21,22AAA
LY33,34AA
PN36,37AA
Input
GST alone
GST-FLASH
dLsm11 AA 1-153 IVT
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21


F
Wild Type
LDV-AAA
FD-AA
PL-AA
Input
GST alone
GST-hFLASH
hLsm11 N-terminus IVT
1 2 3 4 5 6 7 8 9

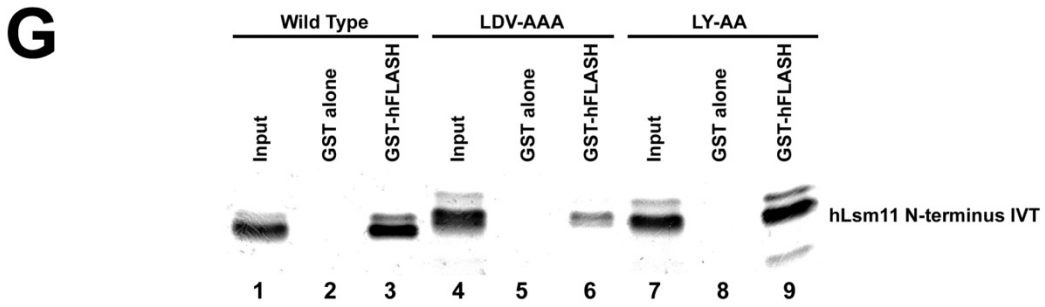
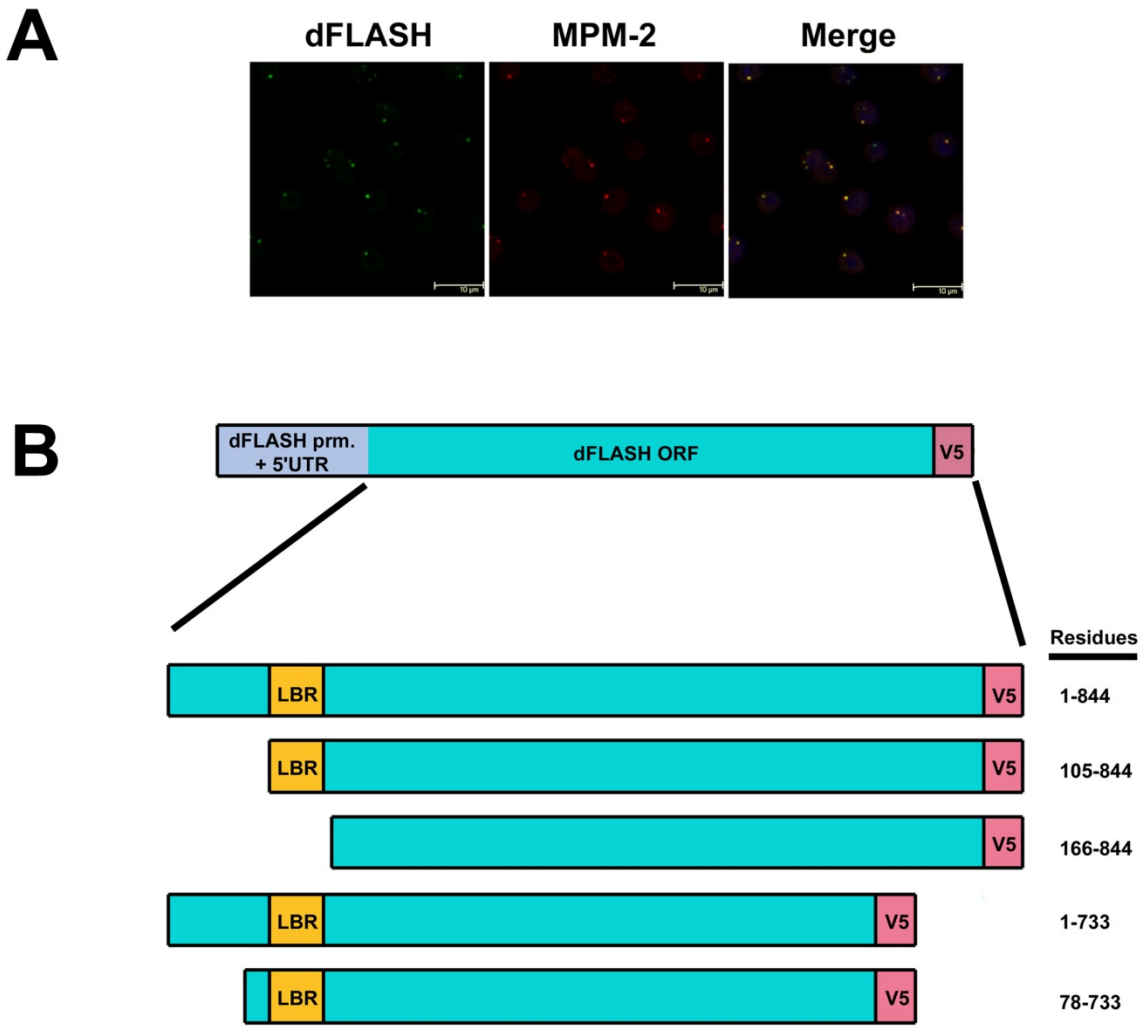

G
Wild Type
LDV-AAA
LY-AA
Input
GST alone
GST-hFLASH
hLsm11 N-terminus IVT
1 2 3 4 5 6 7 8 9

Figure 4-5. dFLASH does not require dLsm11 for recruitment to the HLB. (A) *Drosophila* D.Mel-2 cells were stained with antibodies targeting dFLASH and the MPM-2 reactive epitope, which marks the HLB in *Drosophila*. (B) Diagram of dFLASH V5-tagged expression constructs created for localization experiments. These constructs contain the natural dFLASH promoter and 5' UTR followed by full length dFLASH (AA 1-844) or dFLASH deleted at its N- or C-terminus. Note: None of the deletions introduced resulted in delocalization of recombinant dFLASH from the HLB, as shown in panel C. (C) *Drosophila* D.Mel-2 cells were transfected with the indicated expression constructs, allowed to express V5-tagged dFLASH for two days, then fixed and stained with antibodies targeting the V5 tag on the recombinant protein (red), Mxc as an HLB marker (green), and DAPI (blue) to stain the nuclear material. Untransfected cells (Mock) are included to show the specificity of staining of the V5 antibody. Abbreviations: MPM-2 = mitotic protein monoclonal #2; HLB = histone locus body; V5 = epitope tag consisting of the amino acids GLPIPNPLLGLDST; UTR = untranslated region.


A
dFLASH
MPM-2
Merge
B
dFLASH prm.
+ 5'UTR
dFLASH ORF
V5
Residues
LBR
V5
1-844
LBR
V5
105-844
V5
166-844
LBR
V5
1-733
LBR
V5
78-733

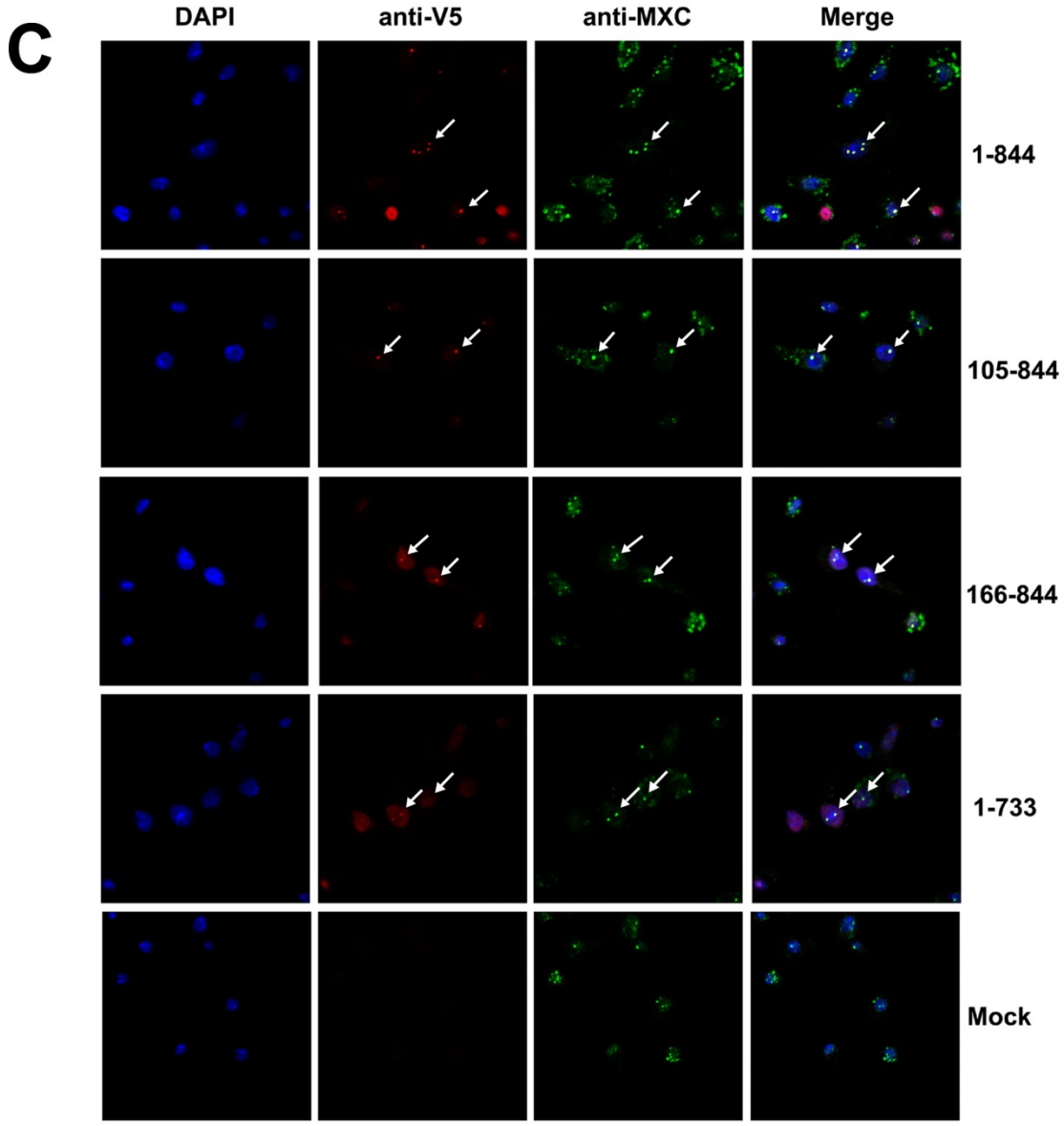
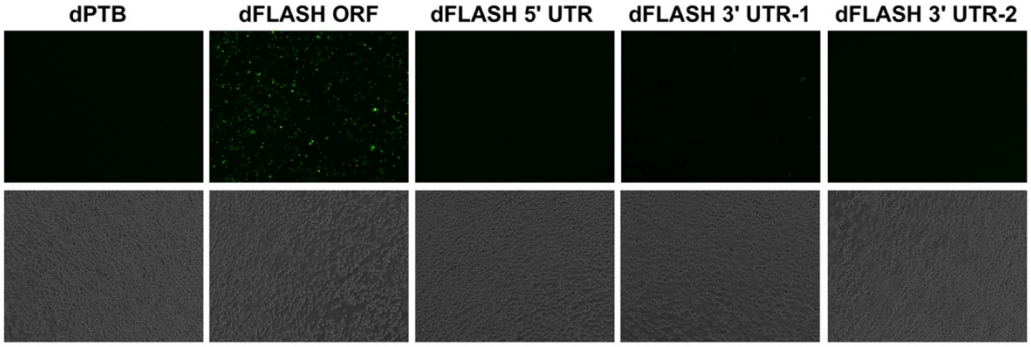

C
DAPI
anti-V5
anti-MXC
Merge
1-844
105-844
166-844
1-733
Mock

Figure 4-6. The N-terminus of dFLASH contains a region upstream of the dLsm11 binding site that is required for histone pre-mRNA processing. (A) Sequence of the 3' end of the dFLASH ORF and 3' UTR as defined in online databases (top). The stop codon is indicated by bold red lettering and polyadenylation signals by bold blue lettering. Sites of polyadenylation indicated by cDNA sequences in databases are depicted by two red vertical bars. Amplicon corresponding to dFLASH 3' UTR-1 is shown between green brackets and amplicon corresponding to dFLASH 3' UTR-2 is shown between purple brackets. *Drosophila* D.Mel-2 cells stably expressing the Act_{proc} reporter were treated with dsRNAs targeting dPTB, the dFLASH ORF, the dFLASH 5' UTR, or one of two regions within the dFLASH 3' UTR and incubated for 5 days, after which the cells were viewed using a fluorescence microscope to detect GFP signal (bottom). (B) *Drosophila* D.Mel-2 cells stably expressing the Act_{proc} reporter were treated with dsRNAs targeting either a region of the dFLASH ORF upstream of amino acid 733 (dFLASH Internal) or the C-terminal region of dFLASH downstream of amino acid 733 (dFLASH C-term.) and incubated for three days, after which the cells were viewed under a fluorescence microscope to detect GFP signal. (C) D.Mel-2 cells containing the constitutive processing reporter were transfected with an expression construct encoding dFLASH amino acids 1-733 (Δ734-844) and driven by the dFLASH promoter and 5'UTR. Untransfected cells are used as a control. Cells were treated 36 hours later with dsRNA targeting the region of dFLASH downstream of amino acid 733 (AA733-844) and incubated for 3 days. The cells were the viewed under a fluorescence microscope to detect GFP signal. (D) D.Mel-2 cells containing the constitutive processing reporter were transfected with an expression construct encoding dFLASH full length or deletion constructs. Untransfected cells are used as a control. Cells were viewed after five

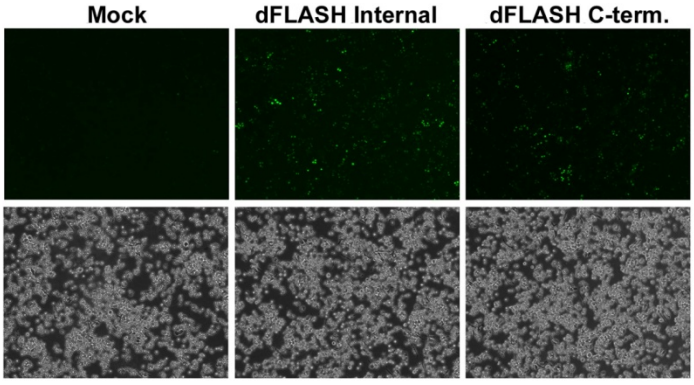
days using a fluorescence microscope to detect GFP signal. Note that there is no RNAi in this experiment and that the same GFP expression pattern shown here is observable just one day after transfection. Abbreviations: PTB = pyrimidine tract binding protein; ORF = open reading frame; UTR = untranslated region; GFP = green fluorescent protein; Act_{proc} = Actin 5C promoter-driven histone pre-mRNA processing reporter.

A

CCAATTTGAAATAAAATTGTGTAATCGATTTGA||AGTGTCCACTTCGAATTTATTTATATATATTACATTGCTGCACCCAATGATTGG
GTGAATAGGTGTTAATTCACGCGATCAGGACTTTCGATCATCTTGAACCATGAGTAGGCAACGATCCAAAAGTAAATGGCTACACCT
GTGGGTTTTGAATTATTATTAATAATAAGGTACTTAAAATAAGCAAACTATTCTTAATTCTGGATAGTGTGAAAATATATCAAATATTTT
GCAAGCTTTACCAAATCCGCCGATGATAAGGAATAAGGTCAAAATCACAGGGCCAATTGCGACGCAAACATAAATGAAAAACATA
AGATCATAAACAATGCGAATTGCGCCTGTTATTAGATAGCCAATGACCAGTTTTTTGGATTTCTGGAAAAGACGAGAGTCTTAATTA
ATAGATTGCTTCTAACTAGAAATGAATACTTACCACCCAAACAGAGCCAATAACTGAGACACAAAACACTAGGTGGGCAAAGTGGA
TGACAAGTGTGATGATCCAAACGTAGTAGGTGGAAGTGTAGCTGACCTCAAACGTTCCCATATCATGATCAATCCACCAGTCCTCAT
AATAGGCTAGATCTGTGCGACACTTTATATTCAAATCGATTGAGATCATATTAAAACACATACACTTAGTTTCCA||AAACAACTGCC



B



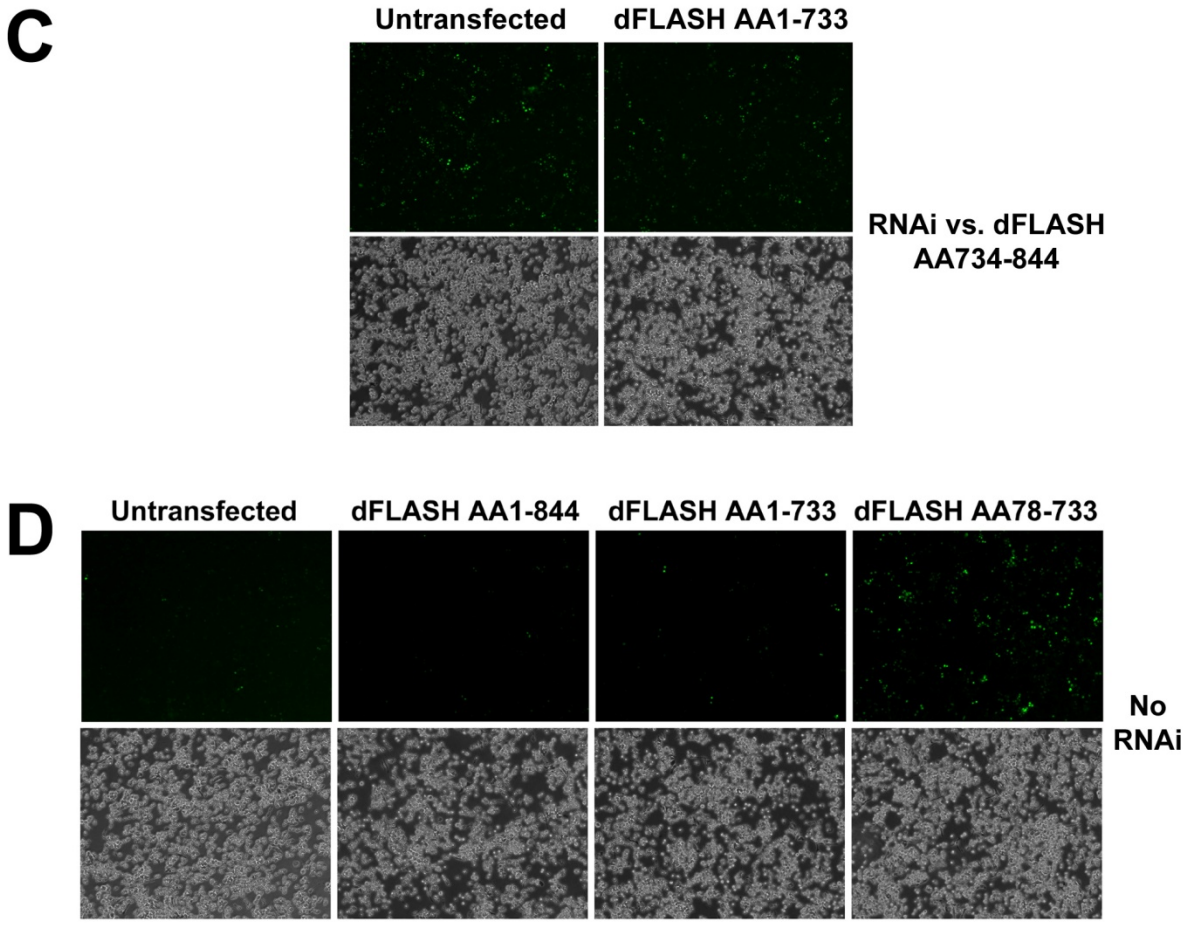
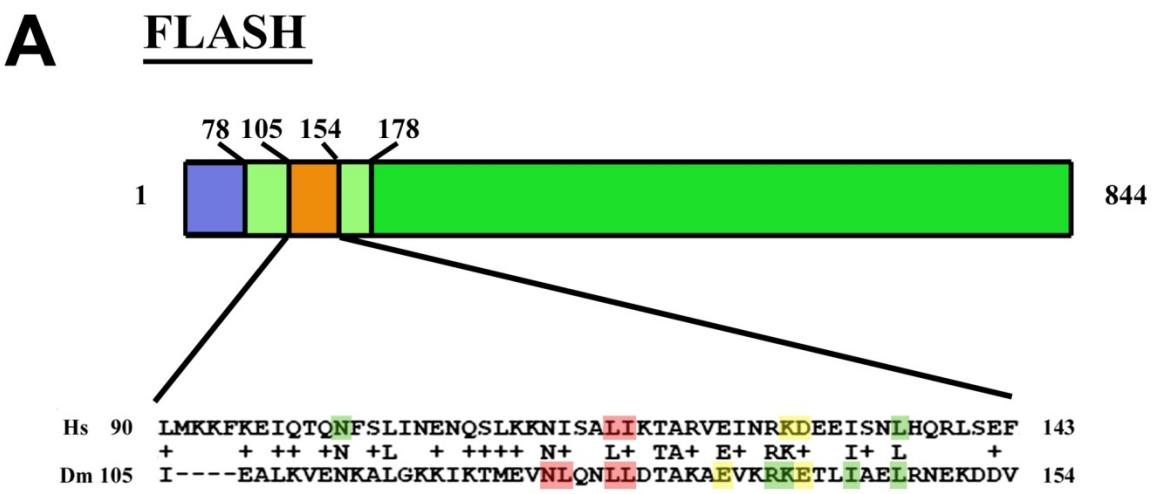
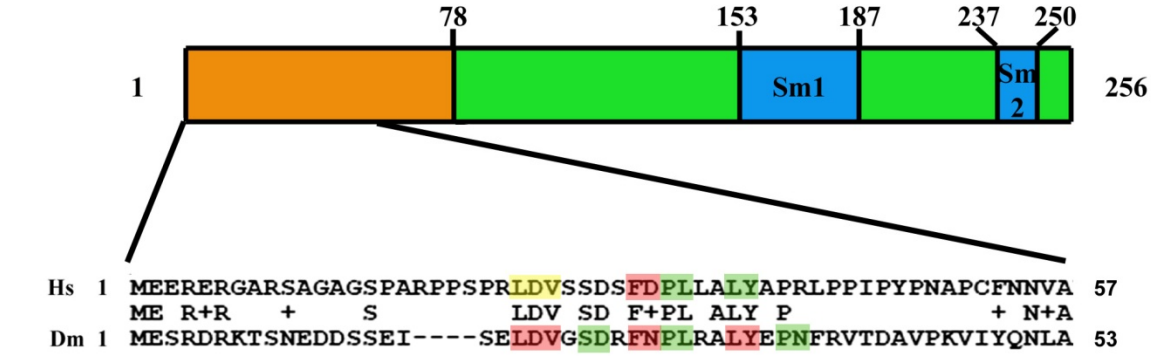

C
Untransfected
dFLASH AA1-733
RNAi vs. dFLASH AA734-844
D
Untransfected
dFLASH AA1-844
dFLASH AA1-733
dFLASH AA78-733
No RNAi

Figure 4-7. Summary of newly identified regions of dFLASH and dLsm11. (A) Diagram of dFLASH with the dLsm11 binding region sequence zoomed in. The purple box is a region of dFLASH that is not required for binding dLsm11, but is required for histone pre-mRNA processing. The orange box is the region of dFLASH containing residues required for binding dLsm11. The residues that are required for this interaction are underlined in red and the residues that are not required are underlined in blue. (B) Diagram of dLsm11 with the dFLASH binding region zoomed in. The orange box is the region of dLsm11 containing residues required for binding dFLASH. The residues that are required for this interaction are underlined in red and the residues that are not required are underlined in blue. The blue boxes correspond to Sm-binding domains 1 and 2.





= No Binding

= Reduced Binding

= No Effect on Binding

CHAPTER V

SUMMARY AND CONCLUSIONS

INTRODUCTION

Histone expression is highly regulated and tightly linked to DNA replication and the cell cycle. These proteins comprise the primary protein component of chromatin and are responsible for packaging the cell's DNA. Histone proteins H2A, H2B, H3, and H4 form a core octamer around which approximately 146 base pairs of DNA are wound, thus forming the primary unit of chromatin, the nucleosome. Histone H1 acts as a linker protein outside the core octamer and participates in the formation of higher order chromatin structures. Nucleosomal chromatin is condensed into higher order chromatin fibers and ultimately into its most condensed form, the chromosome.

In addition to a structural role, histone proteins are also important for gene regulation. Post-translational modifications of the protruding histone tails allow the cell to define regions of active chromatin that are available for transcription and, conversely, to demarcate inactive, or heterochromatic chromatin that is transcriptionally inactive. This allows the cell to exact greater control of gene expression than would be possible from the DNA sequence alone. Because of its role in chromatin structure and gene expression, proper coordination of histone levels is crucial for cell viability. For this reason, a number of regulatory targets exist that allow modulation of histone expression.

Most of the control of histone expression exists at the level of the histone mRNA. Regulatory targets include histone gene transcription, pre-mRNA processing, and mRNA stability. The greatest contributors to increased histone expression during S phase are upregulation of transcription and increased pre-mRNA processing efficiency. Most of my work has focused on describing these early events in histone mRNA metabolism, with special emphasis on pre-mRNA processing and the factors that participate in it.

THE USE OF REPORTERS TO DEFINE EARLY ASPECTS OF HISTONE mRNA METABOLISM

In Chapter II, I discussed the design and validation of a histone pre-mRNA processing reporter by Dr. Eric J. Wagner and me, as well as the creation of a number of other reporters I have used to learn more about histone pre-mRNA processing, histone transcription, and snRNA processing. Three different reporters allow one to distinguish effects on transcription, pre-mRNA processing, and a role in both transcription and processing. One of the most useful tools I have created during my research has been the constitutive histone pre-mRNA processing reporter, especially once integrated into a stable line of *Drosophila* cultured cells. This construct gave the first functional indication that CG4616, later known as dFLASH, is an essential histone pre-mRNA processing factor (Yang *et al.*, 2009). In addition to this, I have used this reporter extensively to screen many other factors for a possible role in the cleavage reaction, trying to determine whether there are factors or pathways that impact both histone gene transcription and histone pre-mRNA processing. Among these are the origin recognition complex (ORC) subunits and several other factors that have a role in DNA replication. Knockdown of these factors reduce

expression of a reporter driven by the histone promoter, but had a minimal impact on the processing reporter. Although none of these proteins appear to have any significant role in processing histone mRNA, these experiments have given us a better understanding of how DNA replication and histone expression are linked, likely primarily at the level of transcription and/or mRNA degradation, by ruling out the possibility of significant modulation through the histone pre-mRNA cleavage reaction.

In addition to these experiments, the constitutive histone pre-mRNA processing reporter can be readily used as a rapid and robust secondary screening tool for factors identified in a primary genome-wide screen. One such primary screen was performed by Anne White in Dr. Bob Duronio's lab. The screen is based upon the observation that the MPM-2 antibody recognizes a cell cycle regulated phosphoepitope that localizes to the HLB. However, it is unknown what protein is being recognized there. It was hypothesized that this factor may be the *Drosophila* homolog of NPAT, a crucial factor for histone expression in mammalian cells that relies on phosphorylation for its activity and localization to the HLB. Thus, Anne developed an assay in which she used RNAi to deplete various factors, and then stained these cells with the MPM-2 antibody to visualize the presence or absence of this phosphoepitope at the HLB. She did this on a genome-wide scale and came up with a list of factors that, when depleted, result in loss of the MPM-2 reactive epitope from the HLB. This list was pared down to around 100 factors of special interest. These 100 factors were subsequently depleted by RNAi in my constitutive reporter stable line and analyzed for production of GFP arising as a result of a histone processing defect. We found a number of interesting factors that scored positively in this secondary screen, among these, dFLASH, MBD-R2, MCRS1, and Mxc. The first three of these had been implicated previously for

involvement in the histone pre-mRNA processing reaction (Wagner *et al.*, 2007; Yang *et al.*, 2009). The possible involvement of Mxc, however, had never been addressed. Of note, depletion of this factor by RNAi in the reporter stable line never elicits fluorescence at the level seen for depletion of a core processing factor, such as SLBP or FLASH. However, the effect on the reporter is reproducible and always above background. This may suggest a more indirect role for this protein in the processing reaction, with depletion of Mxc resulting in some sort of structural perturbation of the HLB, for example. Indeed, results from Anne indicate that depletion of Mxc results not only in elimination of the MPM-2 reactive epitope from the HLB, but also delocalization of a number of other factors, including dFLASH. This could explain the misprocessing phenotype observed in the secondary screen.

A select set of these factors were tested with all three reporters to define at what steps they might function. Results from my histone transcription reporter suggest that Mxc may also have a role in transcribing histone mRNA, which is supported by data from others that depletion of this factor results in an overall reduction in histone mRNA levels (Z. Dominski, unpublished results). Because of its localization to the HLB, the effect of its depletion on the existence of the MPM-2 reactive epitope, its putative role in transcription, the presence of a LisH domain, and the recent finding the MPM-2 antibody directly binds to phosphorylated Mxc (White and Duronio, unpublished), Mxc has been suggested to be the *Drosophila* homolog of mammalian NPAT. Like Mxc, NPAT contains a LisH domain and has a role in histone gene transcription and HLB integrity. Future experiments will determine whether these two proteins are indeed orthologous.

THE USE OF REPORTERS TO CHARACTERIZE THE U7 snRNA PROCESSING REACTION

I also introduced the creation of U7 snRNA reporters to probe snRNA 3' end formation and these results are shown in Chapter II. My contribution to this project came on the heels of Eric's genome-wide screen for U7 pre-snRNA processing factors. In this screen, he used a U7 pre-snRNA processing reporter construct akin to the histone pre-mRNA processing reporter used in the earlier screen.

In addition to Eric's screen for processing factors, I created a number of new reporters in an effort to better define the cis elements that are required for properly processing U7 pre-snRNA. Through these experiments, I discovered that the U7 promoter sequence is required for maximally efficient pre-snRNA processing, but that a histone promoter can be substituted and still give significant amounts of correctly processed U7 snRNA, while the actin promoter is completely inactive in supporting U7 snRNA processing. This is in agreement with previous studies on U1 pre-snRNA processing in sea urchins (Wendelburg and Marzluff, 1992), but in contrast to results from vertebrates (Neuman de Vegvar *et al.*, 1986; Hernandez and Weiner, 1986; Pilch and Marzluff, 1991), indicating that requirements for the snRNA processing reaction differ somewhat between invertebrates and vertebrates. The fact that the histone promoter allows some U7 snRNA processing may be related to the fact that the histone genes are the other set of genes transcribed by RNA polymerase II that do not encode polyadenylated RNAs.

I also tested whether U7 pre-snRNA processing has a length dependence, as has been seen for mammalian snRNAs. Indeed, my results indicate that insertions in excess of 300 nucleotides are sufficient to elicit some level of readthrough. Again, these data are in

agreement with published results revealing that the degree of snRNA misprocessing increases as the transcript length increases (Ramamurthy *et al.*, 1996). Together with the promoter data, these experiments reveal important details about the requirements for efficient pre-snRNA processing, and Eric has continued to modify the reporter to define elements in the snRNA sequence required for efficient cleavage. I am a co-author on the paper describing these studies.

A GENOME-WIDE SCREEN FOR FACTORS REQUIRED FOR HISTONE mRNA 3' END FORMATION

One of the most important lines of research to come out of the initial reporter experiments was the genome-wide screen for factors that participate in histone pre-mRNA processing that Eric and I did (Wagner *et al.*, 2007), which is discussed in Chapter III. This was the first genome-wide screen that targeted RNA metabolism. This project led to the identification of 24 protein factors that appear to have some role in processing histone pre-mRNA. Some of these factors, such as SLBP, Lsm10, Lsm11, CPSF73, and Symplekin had been described previously, (although two of these, CPSF-73 and Symplekin) were identified as we were preparing to do the screen. Most, however, had not been previously implicated and offer us many new research projects for the future.

IDENTIFICATION AND CHARACTERIZTAION OF FLASH AS A HISTONE PRE-mRNA PROCESSING FACTOR

One of the factors that was initially identified in the genome-wide screen, but which was not subsequently validated since it scored weakly, was an uncharacterized factor known

at the time as CG4616. This factor was a weak hit in the screen, did not have an obvious human homologue, and was not among the top 100 hits that were explored further. The importance of this factor to the cleavage reaction was not suggested again until a yeast-two-hybrid screen for factors that interact with the N-terminus of Lsm11 was performed by Xiao-cui Yang and Dr. Zbigniew Dominski. In their work, they found that human FLASH interacts with human Lsm11 and subsequently identified CG4616 as the closest putative homolog of the human FLASH protein based upon a region of partial sequence conservation in the N-terminus of each protein. The first functional data implicating that this factor plays an essential role in processing histone pre-mRNA came from my experiments using the constitutively expressed histone pre-mRNA processing reporter. Using RNA interference to deplete CG4616 from *Drosophila* cultured cells stably expressing the processing reporter construct gave robust fluorescence off the reporter at levels comparable to that elicited by depletion of core processing factors, such as SLBP or Lsm11. The strength of the fluorescence signal indicated that CG4616 most likely had a direct role in the processing reaction. Significantly, the results from the reporter accurately reflect the response of the endogenous histone message upon depletion of CG4616. Using RT-PCR, I determined that endogenous histone mRNAs were being misprocessed upon depletion of CG4616. In addition to the misprocessing phenotype, I also addressed whether this factor has any role in transcription of histone genes, as had been suggested for human FLASH. Contrary to these suggestions, CG4616 did not have any effect on expression off of a histone transcription reporter I created, nor was there an effect on overall endogenous histone levels. This indicated that this factor is not involved in transcribing histone genes. Further work by Xiao and Zbig revealed that this protein is, however, required for processing histone message *in*

vitro, again suggesting that it directly participates in the cleavage reaction. Based upon the accumulated evidence, CG4616 was dubbed dFLASH.

Upon identification of dFLASH as an essential histone pre-mRNA processing factor that interacted with Lsm11, I defined the region of dFLASH that binds dLsm11 and define at least one obligate binding region between amino acids 125-130, as well as potential accessory regions slightly upstream and slightly downstream of these residues. In addition I identified the specific amino acids in dLsm11 required to bind dFLASH, a region of interaction that includes several conserved residues between amino acids 20-34 present in human and *Drosophila*. Significantly, the interaction between dFLASH and dLsm11 does not require all of the conserved residues in this region, but rather, three distinct patches that are separated from each other by intervening conserved residues that are not involved in binding dFLASH. This suggests the contribution of some level of secondary structure in the binding interaction in which binding occurs on one face of the structure, while the other is reserved for some other function. That the residues in this region not involved in binding dLsm11 are well conserved through evolution suggests their importance, though their precise role is unknown. In addition, I used deletion constructs to determine that, while the region from amino acids 1-40 is necessary for binding to dFLASH, it is not sufficient for binding to occur. Based upon these results, it now seems that there are residues between amino acids 40-78 that are required for binding, though their identity is presently unknown.

In addition to refining the definition of the binding regions of dFLASH and dLsm11, I also addressed the requirement of dLsm11 binding for dFLASH recruitment to the Histone Locus Body (HLB). Using V5-tagged dFLASH mutants, I found that deletion of the N-terminus of dFLASH up to amino acid 166, which includes the region responsible for

binding dLsm11, did not affect the ability of dFLASH to localize to the HLB. These results indicate that dFLASH is recruited to the HLB independently of dLsm11 through the interaction of a region downstream of amino acid 166 with an unknown factor. It is possible that the region between amino acid 166-177, which contains at least a few relatively conserved residues, may be involved in this interaction. Equally plausible is that recruitment requires a region well downstream of this, though upstream of amino acid 733, since deletion of amino acids 734-844 has no effect on localization. Future experiments will include the creation of N-terminal and C-terminal deletion mutants to better define where this recruitment domain may lie.

One of the most surprising results to come out of this study was the finding that deleting the first 77 amino acids from the N-terminus of dFLASH results in significant fluorescence off the constitutive histone pre-mRNA processing reporter even in the presence of wild-type FLASH, identifying this protein as an inhibitor of processing and a dominant negative. Importantly, deletion of these residues results in a dFLASH protein that can still localize to the HLB and still bind dLsm11. These results suggest a model in which this protein localizes to the HLB, where it interacts with dLsm11, but is unable to stimulate processing of the histone pre-mRNA because it cannot bind to another key factor(s) which is as yet unknown. Whether these residues bind some other factor or whether they are a target for some modification that activates the dFLASH/dU7 snRNP particle is unclear. It is possible that this region is required for recruitment of the cleavage factor or for converting the naturally exonucleolytic CPSF73 into an endonuclease once it has arrived at the cleavage site. In the case of the former, cleavage will not occur because the nuclease is not available to make the cut. If the latter is the case, deleting this region of dFLASH would result in

assembly of the cleavage machinery, but in an inactive form. In this event, cleavage does not occur, thus allowing for the use of downstream polyadenylation signals. At this juncture, the precise mechanism of cleavage factor recruitment and activation remains unknown, as does the seemingly crucial role of FLASH in these processes.

FINAL REMARKS AND FUTURE DIRECTIONS

At the beginning of my research, the general mechanism of histone pre-mRNA processing was known and a few of the key components were described in relative detail. Included among the known core processing factors were the stem loop binding protein (SLBP) and the U7 snRNP components, Lsm10, Lsm11, and, in mammals, ZFP100. However, little else was known, including the identity of the nuclease that cleaves the histone pre-mRNA or any other factors that are required for processing.

Our current knowledge about the processing reaction, now several years later, has been dramatically transformed by several key observations by us and others. Among the highlights are the identification of CPSF73 as the histone pre-mRNA endonuclease (Dominski *et al.*, 2005), the definition of components of the core cleavage factor, including CPSF73, CPSF100, and Symplekin (Kolev and Steitz, 2005; Wagner *et al.*, 2007; Sullivan *et al.*, 2009), the identification of several novel processing factors using a genome-wide RNAi screen (Wagner *et al.*, 2007), and the identification and characterization of FLASH as an essential component of the processing machinery in mammals and flies (Yang *et al.*, 2009). I have been privileged to be a part of much of this work and am excited about the prospects for the future that have been made possible by these studies.

Among the new directions for study made possible by this research is the characterization of the remaining proteins identified in the genome-wide screen for histone pre-mRNA processing factors. The most promising of these include the factors that scored strongest in the screen, including MCRS1, MBD-R2, Rack1, and CG17361. The first three of these have been characterized as participating in other cellular functions; although, like FLASH, this does not preclude them from also particpating in processing. The last protein, CG17361, is an uncharacterized protein that has been suggested bioinformatically as a zinc binding protein that binds nucleic acids. Further work on these factors will determine their precise role in the processing reaction.

Another aspect of this work that is continuing is refinement of the regions of FLASH and Lsm11 that are responsible for binding the other, as well as a determination of the region of dFLASH that is responsible for its recruitment to the HLB. I have developed a system for deleting regions of dFLASH and expressing them efficiently in *Drosophila* cultured cells that will allow me to determine the answer to this question. In addition to deleting the N- and C-terminal regions, I will also introduce dFLASH protein having point mutations that I have shown abolish dLsm11 binding *in vitro*. Although these residues are not required for dFLASH recruitment, it remains possible that the dU7 snRNP is recruited to the HLB by dFLASH and these experiments will allow me to determine if this is the case.

Finally, the observation that deletion of the first 77 amino acids from dFLASH results in a dominant negative effect on histone pre-mRNA processing when this construct is transfected into the processing reporter stable line has given us a great tool for discovering inhibitors of processing. Natural inhibitors can be detected by overexpressing candidate proteins in the reporter stable line and looking for increased fluorescence. This can be done

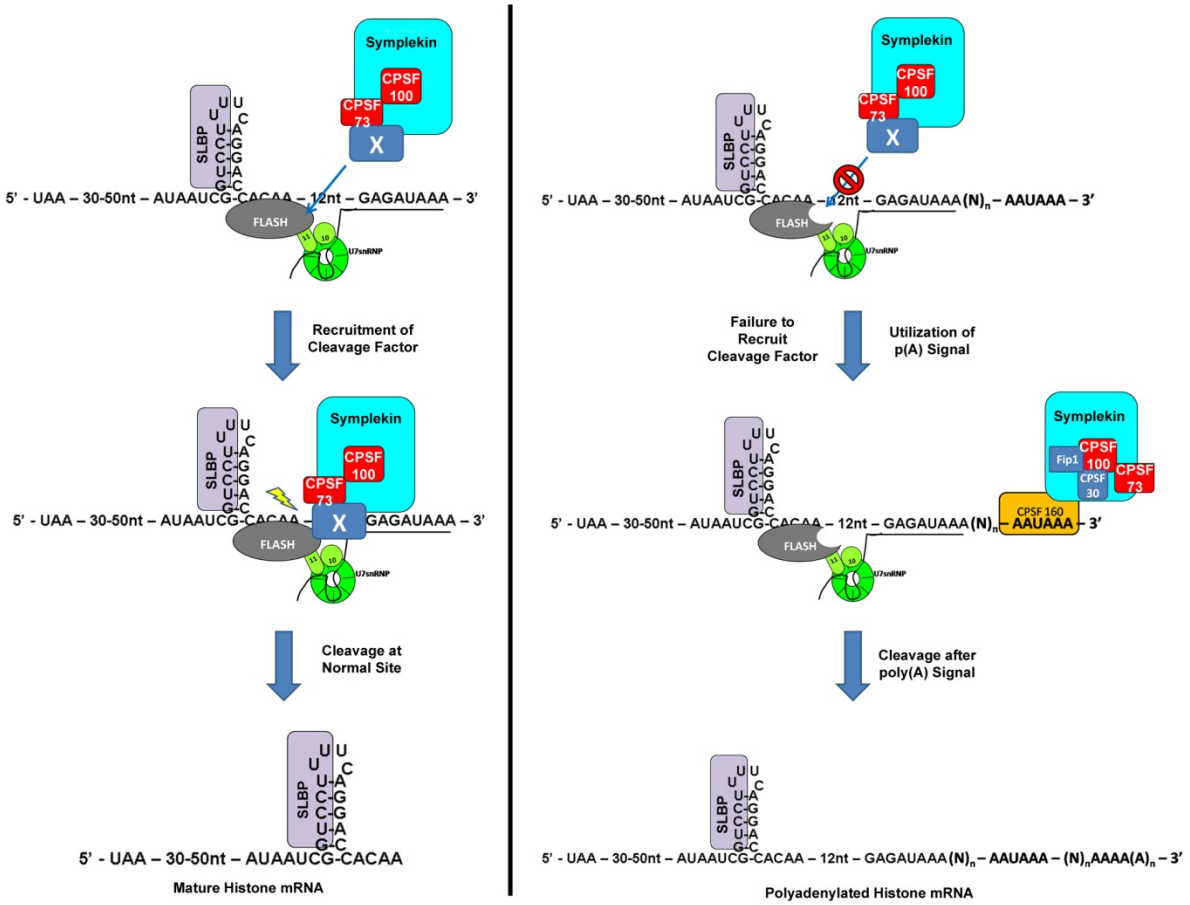
currently in a targeted fashion and is possible on a genome-wide scale given the proper overexpression library. There is an ongoing study by a graduate student in our lab, Ivan Sabbath, and Zbigniew Dominski that is investigating proteins that were found to interact with the N-terminus of hFLASH uisng the yeast-two-hybrid assay. To date, dsRNAs targeting several of these factors have been synthesized and screened using my processing reporter stable line. None of these factors scored as being required for the processing reaction. However, it remains possible that one or more of these factors may participate in inhibiting the cleavage reaction.

As an extension of these studies, it is now possible to screen thousands of chemical compounds to detect chemical inhibitors of histone pre-mRNA processing. This research could lead to new biotherapeutic compounds to be used in human health or alternatively, but not mutually exclusively, could identify compounds that researchers could use to inhibit the processing reaction in a targeted and more efficient way than is currently possible. At the present time, we do not have the capability to perform such targeted inhibition outside of RNAi depletion of required processing factors or overexpression of the dominant negative dFLASH construct. The identification of such an inhibitory compound would allow us to study the mechanisms that work to inhibit processing and the downstream implications of misprocessing histone mRNA without having to mutate or deplete any processing factors, thus eliminating possible indirect, or secondary effects.

In conclusion, the research presented here has helped fill in many gaps in our knowledge about RNA processing and, especially, histone pre-mRNA processing. It has also given us invaluable tools that we can now use to discover other aspects of the reaction, including a definition of all of the factors that are involved in promoting and inhibiting it.

Several questions have now been answered, but even more have been raised, giving me and others many possible avenues to pursue in our efforts to achieve a more complete understanding of this critical cellular process.

Figure 5-1. Model of dFLASH function in histone pre-mRNA processing. Shown is one of several possible models outlining the importance of the first 77 amino acids of dFLASH for histone pre-mRNA processing. Under normal conditions (left panel), dFLASH is recruited to the HLB, where it binds the U7 snRNP through interactions with Lsm11. This interaction requires amino acids 105-154 of dFLASH. dFLASH is then able to recruit an unknown factor "X," which, in this model, is associated with the cleavage factor, to near the site of cleavage. This interaction requires the first 77 amino acids of dFLASH. Once localized to the cleavage site, the cleavage factor is able to cleave the histone pre-mRNA to form a normal, mature histone mRNA with the stem loop and a 4 nt ssRNA tail at its 3' end. However, when the first 77 amino acids are deleted from the dFLASH N-terminus (right panel), FLASH is still recruited to the HLB, where it interacts with dLsm11, but is unable to interact with factor "X." As a result, the cleavage factor is not recruited to the cleavage site and the histone pre-mRNA is not cleaved properly. Instead, a downstream poly(A) signal is recognized by the canonical cleavage and polyadenylation machinery and the histone pre-mRNA is cleaved after this signal and polyadenylated. Abbreviations: HLB = Histone Locus Body; snRNP = small nuclear ribonucleoprotein; Lsm = Sm-like; nt = nucleotide; ssRNA = single-stranded RNA; poly(A) = polyadenylation.


Symplekin
CPSF 100
CPSF 73
X
SLBP
FLASH
U7snRNP
5' - UAA – 30-50nt – AUAAUCG-CACAA – 12nt – GAGAUAAA – 3'
Recruitment of Cleavage Factor
Cleavage at Normal Site
5' - UAA – 30-50nt – AUAAUCG-CACAA
Mature Histone mRNA
5' - UAA – 30-50nt – AUAAUCG-CACAA – 12nt – GAGAUAAA(N)n – AAUAAA – 3'
Failure to Recruit Cleavage Factor
Utilization of p(A) Signal
Fip1
CPSF 30
CPSF 160
Cleavage after poly(A) Signal
5' - UAA – 30-50nt – AUAAUCG-CACAA – 12nt – GAGAUAAA(N)n – AAUAAA – (N)nAAAA(A)n – 3'
Polyadenylated Histone mRNA

Appendix A: Initial hits from the genome-wide screen for histone pre-mRNA processing factors

This appendix contains a table of all factors that scored positively in the genome-wide screen for factors that participate in histone pre-mRNA processing, described in Chapter III. Those factors that were validated in the initial study (Wagner *et al.*, 2007) are highlighted in green. The *Drosophila* FLASH homolog, CG4616, which was validated after the initial study (Yang *et al.*, 2009), is highlighted in yellow. "Strength of Hit" is how strongly positive the factor was by visual inspection of GFP signal, with 3 being a strong hit, 2 being a moderate hit, and 1 being a weak hit. Each factor was graded independently by Dr. Eric J. Wagner and me and the results combined, followed by a final scoring based upon the combined list. Note that these designations were further refined for the validated factors in the final report (Wagner *et al.*, 2007; Chapter III), but this is not reflected here. "DRSC Amplicon" identifies the specific dsRNA from the genome-wide dsRNA library used at the *Drosophila* RNAi Screening Center. More information about each amplicon can be accessed online at http://www.flyrnai.org. "Gene" designates the CG number or abbreviated name of the postively-scoring gene, while "Gene Name" gives the full name. If a factor has a predicted human homolog, it is indicated in the "Human Homolog" column. Any protein domains indicated by bioinformatics are listed under "Protein Domains."

Strength of Hit	DRSC Amplicon	Gene	Gene Name	Human Homolog	Protein Domains
3	DRSC14146	CG1957	CG1957	CPSF2/CPSF100	RNA-metabolising metallo-beta-lactamase, Beta-lactamase-like
3	DRSC03555	kuz	kuzbanian	ADAM10	Peptidase M12B, ADAM/reprolysin, Disintegrin, Peptidase M, neutral zinc metallopeptidases, zinc-b
3	DRSC06250	CG12938	CG12938	Lsm10	
3	DRSC06238	CG12924	CG12924	Lsm11	
3	DRSC11726	CG14565	CG14565		
3	DRSC16764	Or85a	Odorant receptor 85a		Olfactory receptor, Drosophila
3	DRSC16393	CG8165	CG8165	JMJD1B	Transcription factor jumonji, jmjC, Cupin region
3	DRSC16291	CG7698	CG7698	CPSF3/CPSF73	Beta-lactamase-like, RNA-metabolising metallo-beta-lactamase
3	DRSC16863	Slbp	Stem-loop binding protein	SLBP	
3	DRSC20522	CG15450	CG15450		Phospholipid/glycerol acyltransferase
2	DRSC03223	CG9466	CG9466		Glycosyl hydrolases 38, C-terminal, Galactose mutarotase-like, Glycoside hydrolase, family 38
2	DRSC02771	CG4705	CG4705	WDR59	WD-40 repeat, Nitrous oxide reductase, N-terminal, RWD
2	DRSC01987	CG4778	CG4778		Chitin binding Peritrophin-A
2	DRSC02781	CG31743	CG31743	CHST11	Chondroitin 4-O-sulfotransferase
2	DRSC02112	CG10949	CG10949		MADF, HMG-I and HMG-Y, DNA-binding
2	DRSC02441	CG15141	CG15141	C14orf130	Zn-finger (putative), N-recognin, Zn-finger-like, PHD finger
2	DRSC08180	CG1135	CG1135	MCRS1	SMAD/FHA, Forkhead-associated
2	DRSC12301	CG2097	CG2097	SYMPK	HEAT, Armadillo-like helical
2	DRSC16702	His2Av	Histone H2A variant	H2AFV	Histone-fold, Histone core, Histone-fold/TFIID-TAF/NF-Y, Histone H2A

2	DRSC08603	CG7955	CG7955	ABCB7	ABC transporter, transmembrane region, type 1, AAA ATPase, ABC transporter, transmembrane region, ABC transporter related
2	DRSC10550	mib1	mind bomb 1	MIB1	Zn-finger, ZZ type, Zn-finger, RING, Ankyrin, Mib_herc2
2	DRSC09954	CG13071	CG13071		
2	DRSC11651	CG11310	CG11310		Insect cuticle protein
2	DRSC11873	Pc	Polycomb		Chromo
2	DRSC11850	CG32441	CG32441		
2	DRSC16537	CG9684	CG9684		Maternal tudor protein, Tudor, Zn-finger, MYND type
2	DRSC14365	CG31454	CG31454		
2	DRSC14180	MBD-R2	MBD-R2		Tudor, Zn-finger, C2H2 type, Zinc-finger protein THAP_DM3, Zn-finger-like, PHD finger, Methyl-CpG binding
2	DRSC15587	CG4509	CG4509		Sugar transporter superfamily, Cadherin
2	DRSC16015	Bruce	Bruce	BIRC6	Cytochrome cd1-nitrite reductase-like, C-terminal h, WD40-like, Proteinase inhibitor I32, inhibitor of apoptosis, Ubiquitin-conjugating enzymes
2	DRSC14531	dpr4	dpr4		
2	DRSC16914	bon	bonus		B-box, C-terminal, Zn-finger-like, PHD finger, Zn-finger, RING, Bromodomain, Zn-finger, B-box, FYVE/PHD zinc finger
2	DRSC15843	CG5611	CG5611		Enoyl-CoA hydratase/isomerase
2	DRSC17217	activin-beta	activin-beta		Transforming growth factor beta, Inhibin, alpha subunit
2	DRSC17861	sdt	stardust	MPP5	Variant SH3, Guanylate kinase/L-type calcium channel region, Guanylate kinase, L27, SH3, PDZ/DHR/GLGF
2	DRSC18774	flw	flap wing	PPP1CB	Serine/threonine-specific protein phosphatase and b, Metallophosphoesterase
2	DRSC19605	CG1492	CG1492		Gamma-glutamyltranspeptidase
2	DRSC17855	CG32694	CG32694		
2	DRSC07817	HDC07662			

2	DRSC09492	HDC16776, etc.			
2	DRSC12755	HDC12763			
2	DRSC12767	tws	twins		Protein phosphatase 2A regulatory subunit PR55, WD-40 repeat, WD40-like
2	DRSC12434	HDC12200			
2	DRSC13229	HDC14118			
2	DRSC21138	HDC20527			
2	DRSC20317	caz	cabeza	FUS	Eggshell protein, Zn-finger, Ran-binding, RNA-binding region RNP-1 (RNA recognition motif)
2	DRSC01832	HDC03525			
2	DRSC22111	CG32663	CG32663	BCAS3	WD40-like
2	DRSC21257				
2	DRSC23205	CG13894	CG13894		Homeodomain-like, Zinc-finger protein THAP_DM3, Centromere protein B, DNA-binding region
1	DRSC07116	CG8233	CG8233	FLJ10081	
1	DRSC08699	Shab	Shaker cognate b	KCNB1	Ion transport protein, Cation channel, non-ligand gated, Kv channel, Kv9 voltage-gated K+ channel, K+ channel tetramerisation, Voltage-dependent potassium channel, K+ channel, pore region
1	DRSC08450	Fit1	Fermitin 1	PLEKHC1	FERM, Pleckstrin-like, Band 4.1, Pleckstrin homology-type
1	DRSC10929	Ect4	Ect4		Armadillo-like helical, Sterile alpha motif homology 2, Sterile alpha motif homology, Sterile alpha motif SAM, TIR
1	DRSC09787	CG10724	CG10724	WDR1	Nitrous oxide reductase, N-terminal, WD-40 repeat
1	DRSC11324	asf1	anti-silencing factor 1	ASF1A	Anti-silencing protein, ASF1-like
1	DRSC10847	AGO2	Argonaute 2		Stem cell self-renewal protein Piwi, Argonaute and Dicer protein, PAZ
1	DRSC17089	trx	trithorax		Nuclear protein SET, Zn-finger, RING, Zn-finger-like, PHD finger, SET-related region, FY-rich, N-terminal, FY-rich, C-terminal
1	DRSC16253	CG7518	CG7518		Myb, DNA-binding
1	DRSC16731	MRG15	MRG15	MORF4L1	MRG

1	DRSC14613	Orct2	Organic cation transporter 2	SLC22A4	Major facilitator superfamily, Major facilitator superfamily MFS_1
1	DRSC14257	CG10562	CG10562		Protein of unknown function DUF227, Protein kinase-like
1	DRSC16791	Pkc98E	Protein C kinase 98E	PRKCE	Protein kinase-like, Serine/threonine protein kinase, active site, Serine/threonine protein kinase, Protein kinase C, phorbol ester/diacylglycerol bind, Protein kinase, C-terminal, Protein kinase
1	DRSC16676	Fur1	Furin 1	FURIN	Proteinase inhibitor, propeptide, Galactose-binding like, Furin-like repeat, Proprotein convertase, P, Peptidase S8 and S53, subtilisin, kexin, sedolisin, Growth factor, receptor
1	DRSC18568	CG14782	CG14782	PLEKHF2	Pleckstrin homology-type, FYVE/PHD zinc finger, Pleckstrin-like, Zn-finger, FYVE type
1	DRSC18512	CG32810	CG32810	KCTD5	BTB/POZ, K+ channel tetramerisation
1	DRSC15434	CstF-50	CstF-50	CSTF1	WD40-like, WD-40 repeat
1	DRSC17002	Dr	Drop		Homeodomain-like, Homeobox
1	DRSC16948	ferrochelatase	ferrochelatase	FECH	Ferrochelatase
1	DRSC17163	Crk	Crk	CRKL	SH2 motif, SH3, Variant SH3
1	DRSC18775	fs(1)K10	female sterile (1) K10		
1	DRSC17148	MED26	Mediator complex subunit 26		Transcription factors TFIIS, elongin A, CRSP70, con, Flagellar basal body rod protein
1	DRSC17202	CG11148	CG11148		
1	DRSC16671	Fer1HCH	Ferritin 1 heavy chain homologue		Ferritin, Ferritin and Dps, Ferritin-like, Ferritin/ribonucleotide reductase-like
1	DRSC19326	Actr13E	Actin-related protein 13E	ACTR6	Actin/actin-like
1	DRSC19442	CG12204	CG12204		
1	DRSC19906	CG32573	CG32573		
1	DRSC03405	Rack1	Receptor of activated protein kinase C 1	GNB2L1	Nitrous oxide reductase, N-terminal, WD-40 repeat

1	DRSC04547	Phk-3	Pherokine 3		Insect pheromone-binding protein A10/OS-D
1	DRSC06649	CG17048	CG17048		Zn-finger, RING
1	DRSC07695	san	separation anxiety	MAK3	GCN5-related N-acetyltransferase
1	DRSC07026	CG7639	CG7639	SAMM50	Bacterial surface antigen (D15)
1	DRSC06375	CG13338	CG13338		
1	DRSC07575	RacGAP50C	RacGAP50C	RACGAP1	Rho GTPase activation protein, Protein kinase C, phorbol ester/diacylglycerol bind, RhoGAP
1	DRSC06312	Rep3	Rep3		Caspase-activated nuclease CIDE-N
1	DRSC07715	vg	vestigial		Vestigial/tondu, Protein of unknown function TDU, Tubulin
1	DRSC07170	HDC06345			
1	DRSC06267	CG30089	CG30089		
1	DRSC07014	CG7097	CG7097	MAP4K3	Protein kinase, Citron-like, Serine/threonine protein kinase, Protein kinase-like
1	DRSC06008	CG10911	CG10911		
1	DRSC06928	CG5267	CG5267		EGF-like, Cysteine-rich TIL region
1	DRSC08374	CG34056	CG34056		
1	DRSC06393	l(2)05510	lethal (2) 05510		Endoglin/CD105 antigen
1	DRSC08240	CG12038	CG12038		
1	DRSC08379	bab1	bric a brac 1		Helix-turn-helix, Psq-like, Helix-turn-helix, Psq, HMG-I and HMG-Y, DNA-binding, BTB/POZ
1	DRSC06037	CG11110	CG11110	IMMP2L	Peptidase S24 and S26, C-terminal region, Peptidase S26A, signal peptidase I, Peptidase S24, S26A and S26B
1	DRSC08195	CG32269	CG32269		Peptidase, trypsin-like serine and cysteine proteas, Peptidase S1A, chymotrypsin, Peptidase S1, chymotrypsin
1	DRSC05988	CG33786	CG33786		SUA5/yciO/yrdC, N-terminal, Sua5/YciO/YrdC/YwlC
1	DRSC08619	CG9094	CG9094		
1	DRSC08674	LysC, LysD	LysC: Lysozyme C, LysD: Lysozyme D	LYZ	
1	DRSC06403	CG13443	CG13443		

1	DRSC08222	CG12014	CG12014	IDS	Sulfatase
1	DRSC07772	CG17510	CG17510		
1	DRSC08334	CG13800	CG13800		
1	DRSC08444	Gr64b	Gustatory receptor 64b		Trehalose receptor
1	DRSC08739	trh	trachealess	NPAS3	Helix-loop-helix DNA-binding, Basic helix-loop-helix dimerization region bHLH, Legume lectin, beta domain, PAS
1	DRSC08362	Cypl	Cyclophilin-like	PPIL1	Peptidyl-prolyl cis-trans isomerase, cyclophilin ty
1	DRSC08359	CG13889	CG13889	KIAA0373	Haem peroxidase, plant/fungal/bacterial
1	DRSC08198	CG32271	CG32271		Peptidase, trypsin-like serine and cysteine proteas, Peptidase S1, chymotrypsin, Peptidase S1A, chymotrypsin
1	DRSC08590	CG7447	CG7447	EGFL7	EMI, EGF-like calcium-binding, Aspartic acid and asparagine hydroxylation site, EGF-like, EGF-like, subtype 2
1	DRSC11020	CG8620	CG8620		
1	DRSC10672	CG6600	CG6600		Major facilitator superfamily MFS_1, Major facilitator superfamily
1	DRSC10745	CG6902	CG6902		
1	DRSC08321	CG13712	CG13712		
1	DRSC10567	Nelf-E	Negative elongation factor E		RNA-binding region RNP-1 (RNA recognition motif)
1	DRSC09775	CG10674	CG10674	PTD008	Protein of unknown function UPF0139
1	DRSC10663	CG6576	CG6576		
1	DRSC09765	CG32423	CG32423	RBMS3	RNA-binding region RNP-1 (RNA recognition motif), Paraneoplastic encephalomyelitis antigen
1	DRSC11329	bol	boule	BOLL	RNA-binding region RNP-1 (RNA recognition motif)
1	DRSC10420	CG4446, Klp67A	Klp67A: Kinesin-like protein at 67A, CG4446: CG4446	PDXK	Klp67A: Kinesin, motor region, CG4446: PfkB, CG4446: Pyridoxal kinase

1	DRSC09740	sti	sticky	CIT	Protein kinase, Protein kinase, C-terminal, Citron-like, Protein kinase C, phorbol ester/diacylglycerol bind, Serine/threonine protein kinase, Serine/threonine protein kinase, active site, Protein kinase-like
1	DRSC10579	CG6071	CG6071		Peptidase M1, membrane alanine aminopeptidase
1	DRSC10642	CG33493	CG33493		
1	DRSC10148	CG32070	CG32070		
1	DRSC11240	Or67a	Odorant receptor 67a		Olfactory receptor, Drosophila, PAP/25A core
1	DRSC09831	CG11262	CG11262		Citrate transporter
1	DRSC10791	CG7257	CG7257		AAA ATPase, central region, 26S proteasome subunit P45, AAA-protein subdomain, AAA ATPase, Nucleic acid-binding OB-fold
1	DRSC10788	CG7248	CG7248		Chitin binding Peritrophin-A
1	DRSC10260	CG17300	CG17300		
1	DRSC10166	CG14165	CG14165		
1	DRSC10697	CG6707	CG6707	TMEM55B	
1	DRSC09784	CG10710	CG10710		
1	DRSC10584	CG33270	CG33270		
1	DRSC10266	CG17361	CG17361		Zn-finger, C2H2 type
1	DRSC10027	CG32138	CG32138	FMNL2	Diaphanous GTPase-binding, Diaphanous FH3, Actin-binding FH2
1	DRSC11379	nuf	nuclear fallout	RAB11FIP4	
1	DRSC10435	CG4613	CG4613		Peptidase, trypsin-like serine and cysteine proteas, Peptidase S1A, chymotrypsin, Peptidase S1, chymotrypsin
1	DRSC11129	shd	shade		Cytochrome P450, E-class P450, group I
1	DRSC10018	CG13474	CG13474		
1	DRSC09914	CG32159	CG32159		
1	DRSC11115	CkIIalpha-i1	CKII-alpha subunit interactor-1		Zn-finger, C2H2 type
1	DRSC10016	CG13472	CG13472	TDRD3	
1	DRSC10521	CG32187	CG32187		Calycin
1	DRSC10225	CG16775	CG16775		
1	DRSC09889	BobA	Brother of Bearded A		

1	DRSC09912	CG13032	CG13032		
1	DRSC11128	Cyp312a1	Cyp312a1		E-class P450, group IV, Cytochrome P450
1	DRSC11140	Dab	Disabled		Pleckstrin homology-type, Phosphotyrosine interaction region, Peptidase, eukaryotic cysteine peptidase active sit, Fumarate lyase
1	DRSC10494	CG5235	CG5235		PHM/PNGase F Fold, Copper type II, ascorbate-dependent monooxygenase, DOMON, Dopamine-beta-monooxygenase
1	DRSC10503	CG5389	CG5389		H+-transporting two-sector ATPase, alpha/beta subun, ATP synthase F1, beta subunit, H+-transporting two-sector ATPase, alpha/beta subun, H+-transporting two-sector ATPase, alpha/beta subun, AAA ATPase
1	DRSC10008	CG13461	CG13461		
1	DRSC10334	CG18649	CG18649		
1	DRSC10048	CG13699	CG13699		
1	DRSC11221	Met75Cb, Met75Ca	Met75Cb: Met75Cb, Met75Ca: Met75Ca		
1	DRSC10922	CG7841	CG7841		
1	DRSC10074	CG14057	CG14057		
1	DRSC10662	HDC10619			
1	DRSC10730	CG6856	CG6856	DTNBP1	
1	DRSC10805	CG7306	CG7306		Chitin binding Peritrophin-A, Blue (type 1) copper domain
1	DRSC09849	CG11619	CG11619		
1	DRSC11707	CG14450	CG14450		
1	DRSC11706	SPoCk	Secretory Pathway Calcium atpase	ATP2C1	Haloacid dehalogenase-like hydrolase, Cation transporting ATPase, C-terminal, Calcium-transporting P-type ATPase, PMR1-type, ATPase, E1-E2 type, Cation transporting ATPase, N-terminal, H+ transporting ATPase, proton pump, E1-E2 ATPase-associated region
1	DRSC10083	CG14079	CG14079		
1	DRSC11605	Aef1	Adult enhancer factor 1		Zn-finger, C2H2 type

1	DRSC10092	CG14088	CG14088		Peptidase S1, chymotrypsin, Peptidase, trypsin-like serine and cysteine proteas
1	DRSC11695	CG13247	CG13247		
1	DRSC11784	CG5847	CG5847		Endoglin/CD105 antigen, YLP motif
1	DRSC11855				
1	DRSC11843	CG7634	CG7634		
1	DRSC11757	CG4825	CG4825	PTDSS1	Phosphatidyl serine synthase
1	DRSC11737	ORMDL	ORMDL	ORMDL3	ORMDL
1	DRSC11645	CG11249	CG11249		Pyruvate kinase, Pyruvate kinase, beta-barrel-like
1	DRSC11068	CG9372	CG9372		Peptidase S1A, chymotrypsin, Disulfide knot CLIP, Peptidase, trypsin-like serine and cysteine proteas, Peptidase S1, chymotrypsin
1	DRSC11856	CG11489	CG11489		Protein kinase, Serine/threonine protein kinase, active site, Protein kinase-like
1	DRSC11659	CG11437	CG11437		Acid phosphatase/vanadium-dependent haloperoxidase, Phosphoesterase, PA-phosphatase related
1	DRSC11900	Trxr-2	thioredoxin reductase 2	TXNRD2	FAD-dependent pyridine nucleotide-disulfide oxidore, Pyridine nucleotide-disulfide oxidoreductase dimeri, Thioredoxin and glutathione reductase selenoprotein, Pyridine nucleotide-disulfide oxidoreductase, class, Mercuric reductase, NAD-binding site
1	DRSC11716	CG33766	CG33766		
1	DRSC12215	CG12589	CG12589		
1	DRSC16451	CG8866	CG8866	SCAMP2	Protein kinase-like, Serine/threonine protein kinase, active site, MIT, Serine/threonine protein kinase, Tyrosine protein kinase, Protein kinase
1	DRSC14959	CG14739	CG14739		Ubiquitin-conjugating enzymes
1	DRSC15649	Tctp	Translationally controlled tumor protein	TPT1	Mss4-like, Translationally controlled tumor protein, Transposase, IS4
1	DRSC15253	CG17721	CG17721		

1	DRSC14179	CG10041	CG10041		Peptidase, trypsin-like serine and cysteine proteas, Peptidase S1A, chymotrypsin, Peptidase S1, chymotrypsin
1	DRSC15346	GstD10	Glutathione S transferase D10		Glutathione S-transferase, N-terminal, Glutathione S-transferase, C-terminal-like, Glutathione S-transferase, C-terminal
1	DRSC16688	GstD7	Glutathione S transferase D7		Glutathione S-transferase, C-terminal-like, Glutathione S-transferase, C-terminal, Glutathione S-transferase, N-terminal
1	DRSC17072	stich1	sticky ch1		Helix-loop-helix DNA-binding, Orange, Basic helix-loop-helix dimerization region bHLH
1	DRSC14927				
1	DRSC15390				
1	DRSC16415	Pnn	Pinin	PNN	Pinin/SDK/memA protein
1	DRSC15163	CG16908	CG16908	KIAA0406	
1	DRSC14940	CG14717	CG14717		Esterase/lipase/thioesterase, Alpha/beta hydrolase fold
1	DRSC14331	CG11598	CG11598		Alpha/beta hydrolase fold, AB-hydrolase associated lipase region, Esterase/lipase/thioesterase, Lipase, active site
1	DRSC15861	CG5724	CG5724		UDP-glucuronosyl/UDP-glucosyltransferase
1	DRSC14845	beat-Vb	beat-Vb		Immunoglobulin-like, Immunoglobulin subtype
1	DRSC15011	CG14889	CG14889		Collagen helix repeat, Collagen triple helix repeat
1	DRSC16854	Scp2	Sarcoplasmic calcium-binding protein 2		EF-Hand type, Calcium-binding EF-hand
1	DRSC14538	CG12783	CG12783		Major facilitator superfamily, General substrate transporter
1	DRSC14106	Actn3	alpha actinin 3		Calponin-like actin-binding
1	DRSC16423	wntD	wnt inhibitor of Dorsal		Secreted growth factor Wnt protein, Wnt superfamily
1	DRSC16333	CG7886	CG7886		
1	DRSC14403	CG31183	CG31183	NPR2	Natriuretic peptide receptor, N-terminal, Protein kinase-like, Extracellular ligand-binding receptor, Protein kinase, Guanylate cyclase

1	DRSC16821	Rh2	Rhodopsin 2	OPN4	Rhodopsin-like GPCR superfamily, Opsin, Opsin RH1/RH2
1	DRSC14772	CG14292	CG14292		
1	DRSC15283	CG17836	CG17836		HMG-I and HMG-Y, DNA-binding
1	DRSC14798	tinc	tincar		
1	DRSC15109	CG31209	CG31209		
1	DRSC15708	CG5060	CG5060		
1	DRSC16224	Rim	Rim		C2 calcium/lipid-binding region, CaLB, PDZ/DHR/GLGF, C2
1	DRSC15740	CG5217	CG5217		
1	DRSC15714	Sirt2	Sirt2	SIRT2	Silent information regulator protein Sir2
1	DRSC17090	tsl	torso-like		Membrane attack complex component/perforin/compleme
1	DRSC14760	CG31212	CG31212	INOC1	DEAD/DEAH box helicase, N-terminal, SNF2-related, Helicase, C-terminal
1	DRSC16371	CG31241	CG31241	NCOA6IP	
1	DRSC14506	CG18600	CG18600		
1	DRSC15113	CG31203	CG31203		
1	DRSC14701	CG31160	CG31160		FLYWCH Zn-finger, BTB/POZ
1	DRSC15220	Obp93a	Odorant-binding protein 93a		Pheromone/general odorant binding protein, PBP/GOBP
1	DRSC16149	CG6954	CG6954	ANKFN1	Ankyrin, Fibronectin, type III, RA
1	DRSC16781	OstStt3	Oligosaccharyl transferase 3	SIMP	Phosphopantetheine attachment site, Oligosaccharyl transferase, STT3 subunit
1	DRSC14678	CG13828	CG13828		
1	DRSC17084	tld	tolloid		Peptidase M, neutral zinc metallopeptidases, zinc-b, EGF-like, Peptidase, metallopeptidases, CUB, Peptidase M12A, astacin, Aspartic acid and asparagine hydroxylation site, EGF-like, subtype 2, EGF-like calcium-binding
1	DRSC14437	CG11913	CG11913		NADH dehydrogenase I, D subunit, NADH-ubiquinone oxidoreductase, chain 49kDa, Pectinesterase
1	DRSC15844	CG5612	CG5612		
1	DRSC14735	CG31080	CG31080		

1	DRSC15687	CG4963	CG4963	SLC25A37	Mitochondrial carrier protein, Mitochondrial substrate carrier
1	DRSC14950	PH4alphaEFB	prolyl-4-hydroxylase-alpha EFB	P4HA1	Prolyl 4-hydroxylase, alpha subunit, 2OG-Fe(II) oxygenase
1	DRSC15288	CG17856	CG17856	RP11-278E11.2	Cytochrome bd ubiquinol oxidase, 14 kDa subunit
1	DRSC14864	CG14512	CG14512	GLT28D1	
1	DRSC15505	CG3669	CG3669		Carbonic anhydrase, eukaryotic
1	DRSC15104	CG15566	CG15566		
1	DRSC18782	gt	giant		Basic leucine zipper, Basic-leucine zipper (bZIP) transcription factor
1	DRSC18566	CG14780	CG14780		Peptidase, trypsin-like serine and cysteine proteas, Peptidase S1A, chymotrypsin, Peptidase S1, chymotrypsin
1	DRSC18832	sgg	shaggy	GSK3B	Protein kinase, Serine/threonine protein kinase, active site, Protein kinase-like
1	DRSC18638	CG14417	CG14417		
1	DRSC18180	mus81	mus81	MUS81	
1	DRSC18642	CG3699	CG3699		Glucose/ribitol dehydrogenase, Short-chain dehydrogenase/reductase SDR
1	DRSC18472	CG14424	CG14424		
1	DRSC18090	CG15784	CG15784		
1	DRSC18045	CG15375	CG15375		
1	DRSC18014	CG15330	CG15330		
1	DRSC17753	CG10761	CG10761		
1	DRSC18691	Or7a	Odorant receptor 7a		Olfactory receptor, Drosophila
1	DRSC18341	CG4064	CG4064		
1	DRSC18795	mof	males absent on the first	MYST1	Winged helix repressor DNA-binding, MOZ/SAS-like protein
1	DRSC18357	Ca-alpha1T	Ca-alpha1T		Ion transport protein, Cation channel, non-ligand gated, Carbamoyl-phosphate synthase L chain, ATP-binding, T-type voltage-dependent calcium channel alpha 1 su, Cation (not K+) channel, TM region, Glyceraldehyde 3-phosphate dehydrogenase, Ca2+/Na+ channel, pore region
1	DRSC17966	CG15034	CG15034		

1	DRSC17939	CG14446	CG14446	DKFZp761O2018	
1	DRSC18330	CG3898	CG3898		
1	DRSC18430	fh	frataxin-like	FXN	Frataxin, Frataxin-like
1	DRSC19796	CG1824	CG1824	ABCB8	ABC transporter, transmembrane region, AAA ATPase, ABC transporter, transmembrane region, type 1, ABC transporter related
1	DRSC17740	Hmr	Hybrid male rescue		MADF
1	DRSC18439	CG9686	CG9686		
1	DRSC19604	CG1490	CG1490	USP7	TRAF-like, Peptidase C19, ubiquitin carboxyl-terminal hydrolas, MATH
1	DRSC19867	CG2750	CG2750		
1	DRSC19862	CG2574	CG2574		Ubiquitin-conjugating enzymes
1	DRSC19393	NFAT	NFAT		RHD, p53-like transcription factor, DNA-binding, Nuclear factor of activated T cells (NFAT), Cell surface receptor IPT/TIG
1	DRSC19742	CG15927	CG15927		
1	DRSC19473	CG12609	CG12609		
1	DRSC20472	CG10918	CG10918		
1	DRSC20037	CG7326	CG7326		
1	DRSC21057				
1	DRSC19781	CG17757	CG17757		
1	DRSC19545	CG14190	CG14190		
1	DRSC19546	CG14191	CG14191		
1	DRSC20527	CG15456	CG15456		
1	DRSC20494	CG1304	CG1304		Peptidase, trypsin-like serine and cysteine proteas, Peptidase S1A, chymotrypsin, Peptidase S1, chymotrypsin
1	DRSC00105	HDC00355			
1	DRSC00205	HDC00659			
1	DRSC00128	HDC00401			
1	DRSC00217	HDC00689			
1	DRSC00906	CG31643	CG31643	FLJ21901	FAST kinase leucine-rich
1	DRSC00241	HDC00760			
1	DRSC00922	HDC01071			
1	DRSC00064	HDC00239			
1	DRSC00960	xl6	xl6	SFRS7	Zn-finger, CCHC type, RNA-binding region RNP-1 (RNA recognition motif)
1	DRSC00964	HDC01229			

1	DRSC01261	HDC02204			
1	DRSC01090	HDC01659			
1	DRSC01003	HDC01363			
1	DRSC01101	CG33300	CG33300		
1	DRSC01224	HDC02112			
1	DRSC01345	HDC02404			
1	DRSC01253	HDC02192			
1	DRSC01351	HDC02434			
1	DRSC01454	HDC02642			
1	DRSC01363	HDC02478			
1	DRSC01493	HDC02700			
1	DRSC01687	HDC03210			
1	DRSC01408	HDC02555			
1	DRSC01423	HDC02581			
1	DRSC01436	HDC02612			
1	DRSC01634	HDC03051			
1	DRSC01731	HDC03321			
1	DRSC03903	HDC03986			
1	DRSC01760	HDC04718, etc.			
1	DRSC03920	HDC04033			
1	DRSC03932	HDC04086			
1	DRSC03963	HDC04202			
1	DRSC03988	HDC04256			
1	DRSC03997	HDC04272			
1	DRSC05135	Tom7	Translocase of outer membrane 7		
1	DRSC04032				
1	DRSC04753	HDC04678			
1	DRSC05317	HDC05864			
1	DRSC05222	lola	longitudinals lacking		Aspartate carbamoyltransferase, regulatory chain, C, Zn-finger, C2H2 type, Calcium-binding EF-hand, Flagellar basal body rod protein, BTB/POZ
1	DRSC05406	HDC06176			
1	DRSC05259	HDC05705			
1	DRSC05362	HDC06052			
1	DRSC05606	HDC06739			
1	DRSC05702	HDC07069			

1	DRSC05621	CG30463	CG30463		Ricin B-related lectin, Glycosyl transferase, family 2, Ricin B lectin, Putative DNA binding
1	DRSC05717	HDC07110			
1	DRSC05635	HDC06835			
1	DRSC05826	HDC07406			
1	DRSC05647	HDC06896			
1	DRSC05761	HDC07242			
1	DRSC07806	HDC07650			
1	DRSC07928	trio	trio	TRIO	Pleckstrin homology-type, Spectrin repeat, Pleckstrin-like, SH3, Cellular retinaldehyde-binding)/triple function, C-, DH
1	DRSC07810	CG17082	CG17082		
1	DRSC08782	CG13291	CG13291		
1	DRSC08815	HDC08760			
1	DRSC09098	HDC09594			
1	DRSC09100	HDC17427, etc.			
1	DRSC09104	HDC09602			
1	DRSC09033	HDC09444			
1	DRSC08859	HDC08829			
1	DRSC08889	HDC08951			
1	DRSC09081	HDC09528			
1	DRSC09185	HDC09845			
1	DRSC09188	HDC09850			
1	DRSC09541	HDC10807			
1	DRSC12745	HDC12669			
1	DRSC12784	HDC12851			
1	DRSC12088	HDC12016			
1	DRSC12105	HDC12040			
1	DRSC12938	HDC13300			
1	DRSC12889	HDC13149			
1	DRSC13085	HDC13758			
1	DRSC17699	ptr	proximal to raf		
1	DRSC17245	HDC17203			
1	DRSC14069	HDC16604			
1	DRSC20530	CG15459	CG15459		
1	DRSC21158	HDC20545			
1	DRSC08997	CG8177	CG8177	SLC4A3	HCO3- transporter, eukaryote, Anion exchange protein, HCO3-transporter

1	DRSC06872	CG4616	CG4616	FLASH	
1	DRSC21826	Bap170	Brahma associated protein 170kD	ARID2	Winged helix repressor DNA-binding, Zn-finger, C2H2 type, AT-rich interaction region
1	DRSC22142	CG30058	CG30058		
1	DRSC21874	gfzf	GST-containing FLYWCH zinc-finger protein		FLYWCH Zn-finger
1	DRSC22354	CG32299	CG32299		
1	DRSC22112	HDC02243			
1	DRSC04716	slbo	slow border cells		Basic leucine zipper, Basic-leucine zipper (bZIP) transcription factor
1	DRSC06545	CG18367	CG18367		
1	DRSC18320	CG3726	CG3726		Helix-turn-helix, Fis-type, HMG-I and HMG-Y, DNA-binding, BTB/POZ, Helix-turn-helix, Psq, Helix-turn-helix, Psq-like
1	DRSC23345	CycG	Cyclin G	CCNG2	Cyclin-like, Cyclin, N-terminal, Cyclin

References

1. Ach,R.A., Weiner,A.M. (1987). The highly conserved U small nuclear RNA 3'-end formation signal is quite tolerant to mutation. Molecular and Cellular Biology *7*, 2070-2079.

2. Adamson,T.E., Price,D.H. (2003). Cotranscriptional processing of Drosophila histone mRNAs. Mol.Cell Biol. *23* , 4046-4055.

3. Ahmad,K., Henikoff,S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol.Cell *9*, 1191-1200.

4. Allfrey,V.G., Faulkner,R., Mirsky,A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc.Natl.Acad.Sci.U.S.A *51*, 786-794.

5. Antequera,F. (2004). Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. EMBO Journal *23*, 4365-4370.

6. Aparicio,O.M., Stout,A.M., Bell,S.P. (1999). Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc.Natl.Acad.Sci.U.S.A *96*, 9130-9135.

7. Aravind,L. (1999). An evolutionary classification of the metallo-beta-lactamase fold proteins. In Silico.Biol. *1*, 69-91.

8. Ardehali,M.B., Yao,J., Adelman,K., Fuda,N.J., Petesch,S.J., Webb,W.W., Lis,J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO Journal *28*, 1067-1077.

9. Armknecht,S., Boutros,M., Kiger,A., Nybakken,K., Mathey-Prevot,B., Perrimon,N. (2005). High-throughput RNA interference screens in Drosophila tissue culture cells. Methods Enzymol. *392*, 55-73.

10. Aves,S.J., Tongue,N., Foster,A.J., Hart,E.A. (1998). The essential schizosaccharomyces pombe cdc23 DNA replication gene shares structural and functional homology with the Saccharomyces cerevisiae DNA43 (MCM10) gene. Curr.Genet. *34*, 164-171.

11. Azzouz,T.N., Schumperli,D. (2003). Evolutionary conservation of the U7 small nuclear ribonucleoprotein in Drosophila melanogaster. RNA. *9*, 1532-1541.

12. Azzouz,T.N., Gruber,A., Schumperli,D. (2005). U7 snRNP-specific Lsm11 protein: dual binding contacts with the 100 kDa zinc finger processing factor (ZFP100) and a

ZFP100-independent function in histone RNA 3' end processing. Nucleic Acids Research *33*, 2106-2117.

13. Bachi,A., Braun,I.C., Rodrigues,J.P., Pante,N., Ribbeck,K., von Kobbe,C., Kutay,U., Wilm,M., Gorlich,D., Carmo-Fonseca,M., Izaurralde,E. (2000). The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. RNA. *6*, 136-158.

14. Baillat,D., Hakimi,M.A., Naar,A.M., Shilatifard,A., Cooch,N., Shiekhattar,R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265-276.

15. Barcaroli,D., Dinsdale,D., Neale,M.H., Bongiorno-Borbone,L., Ranalli,M., Munarriz,E., Sayan,A.E., McWilliam,J.M., Smith,T.M., Fava,E., Knight,R.A., Melino,G., De,L., V (2006). FLASH is an essential component of Cajal bodies. Proc.Natl.Acad.Sci.U.S.A *103*, 14802-14807.

16. Barcaroli,D., Bongiorno-Borbone,L., Terrinoni,A., Hofmann,T.G., Rossi,M., Knight,R.A., Matera,A.G., Melino,G., De,L., V (2006). FLASH is required for histone transcription and S-phase progression. Proc.Natl.Acad.Sci.U.S.A *103*, 14808-14812.

17. Baumbach,L.L., Stein,G.S., Stein,J.L. (1987). Regulation of human histone gene expression: transcriptional and posttranscriptional control in the coupling of histone messenger RNA stability with DNA replication. Biochemistry *26*, 6178-6187.

18. Bell,S.P., Stillman,B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature(London) *357*, 128-134.

19. Bell,S.P., Mitchell,J., Leber,J., Kobayashi,R., Stillman,B. (1995). The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. Cell *83*, 563-568.

20. Bell,S.P. (2002). The origin recognition complex: from simple origins to complex functions. Genes and Development *16*, 659-672.

21. Benecke,A. (2006). Chromatin code, local non-equilibrium dynamics, and the emergence of transcription regulatory programs. Eur.Phys.J.E.Soft.Matter *19*, 353-366.

22. Bentley,D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. Current Opinion in Cell Biology *14*, 336-342.

23. Birchmeier,C., Grosschedl,R., Birnstiel,M.L. (1982). Generation of authentic 3' termini of an H2A mRNA in vivo is dependent on a short inverted DNA repeat and on spacer sequences. Cell *28* , 739-745.

24. Birchmeier,C., Folk,W., 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.

25. Birchmeier,C., Schümperli,D., Sconzo,G., Birnstiel,M.L. (1984). 3' editing of mRNAs: sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proceedings of the National Academy of Sciences of the United States of America *81*, 1057-1061.

26. Birnstiel,M.L., Busslinger,M., Strub,K. (1985). Transcription termination and 3' processing: the end is in site! Cell *41*, 349-359.

27. Blow,J.J., Dutta,A. (2005). Preventing re-replication of chromosomal DNA. Nat.Rev.Mol.Cell Biol. *6*, 476-486.

28. Bogan,J.A., Natale,D.A., DePamphilis,M.L. (2000). Initiation of eukaryotic DNA replication: Conservative or liberal? Journal of Cellular Physiology *184*, 139-150.

29. Bousset,K., Diffley,J.F. (1998). The Cdc7 protein kinase is required for origin firing during S phase. Genes and Development *12*, 480-490.

30. Boutros,M., Kiger,A.A., Armknecht,S., Kerr,K., Hild,M., Koch,B., Haas,S.A., Consortium,H.F., Paro,R., Perrimon,N. (2004). Genome-wide RNAi analysis of growth and viability in Drosophila cells. Science *303*, 832-835.

31. Bowman,T.L., Hurt,M.M. (1995). The coding sequences of mouse H2A and H3 histone genes contains a conserved seven nucleotide element that interacts with nuclear factors and is necessary for normal expression. Nucleic Acids Research *23*, 3083-3092.

32. Brewer,B.J., Fangman,W.L. (1987). The localization of replication origins on ARS plasmids in S. cerevisiae. Cell *51*, 463-471.

33. Brown,K.M., Gilmartin,G.M. (2003). A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol.Cell *12*, 1467-1476.

34. Brush,D., Dodgson,J.B., Choi,O.R., Stevens,P.W., Engel,J.D. (1985). Replacement variant histone genes contain intervening sequences. Molecular and Cellular Biology *5*, 1307-1317.

35. Burzio,L.O., Riquelme,P.T., Koide,S.S. (1979). ADP ribosylation of rat liver nucleosomal core histones. Journal of Biological Chemistry *254*, 3029-3037.

36. Byrnes,J.J., Downey,K.M., Black,V.L., So,A.G. (1976). A new mammalian DNA polymerase with 3' to 5' exonuclease activity: DNA polymerase delta. Biochemistry *15*, 2817-2823.

37. Cakmakci,N.G., Lerner,R.S., Wagner,E.J., Zheng,L.-X., Marzluff,W.F. (2008). SLIP1, A factor required for activation of histone mRNA translation by the stem-loop binding protein. Mol.Cell Biol. *28*, 1182-1194.

38. Callebaut,I., Moshous,D., Mornon,J.P., de Villartay,J.P. (2002). Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. Nucleic Acids Research *30*, 3592-3601.

39. Camporeale,G., Shubert,E.E., Sarath,G., Cerny,R., Zempleni,J. (2004). K8 and K12 are biotinylated in human histone H4. Eur.J.Biochem. *271*, 2257-2263.

40. Caplen,N.J., Fleenor,J., Fire,A., Morgan,R.A. (2000). dsRNA-mediated gene silencing in cultured Drosophila cells: a tissue culture model for the analysis of RNA interference. Gene *252*, 95-105.

41. Carpenter,P.B., Mueller,P.R., Dunphy,W.G. (1996). Role for a *Xenopus* Orc2-related protein in controlling DNA replication. Nature(London) *379*, 357-360.

42. Choi,Y.H., Kim,K.B., Kim,H.H., Hong,G.S., Kwon,Y.K., Chung,C.W., Park,Y.M., Shen,Z.J., Kim,B.J., Lee,S.Y., Jung,Y.K. (2001). FLASH coordinates NF-kappa B activity via TRAF2. Journal of Biological Chemistry *276*, 25073-25077.

43. Chong,J.P.J., Mahbubani,H.M., Khoo,C.-Y., Blow,J.J. (1995). Purification of an MCM-containing complex as a component of the DNA replication licensing system. Nature(London) *375*, 418-421.

44. Clemens,J.C., Worby,C.A., Simonson-Leff,N., Muda,M., Maehama,T., Hemmings,B.A., Dixon,J.E. (2000). Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proc.Natl.Acad.Sci.U.S.A *97*, 6499-6503.

45. Coleman,T.R., Carpenter,P.B., Dunphy,W.G. (1996). The Xenopus Cdc6 protein is essential for the initiation of a single round of DNA replication in cell-free extracts. Cell *87*, 53-63.

46. Conaway,R.C., Lehman,I.R. (1982). A DNA primase activity associated with DNA polymerase alpha from Drosophila melanogaster embryos. Proc.Natl.Acad.Sci.U.S.A *79*, 2523-2527.

47. Cotten,M., Gick,O., Vasserot,A., Schaffner,G., Birnstiel,M.L. (1988). Specific contacts between mammalian U7 snRNA and histone precursor RNA are indispensable for the in vitro 3' RNA processing reaction. EMBO Journal *7*, 801-808.

48. Cramer,P., Srebrow,A., Kadener,S., Werbajh,S., De la Mata,M., Melen,G., Nogués,G., Kornblihtt,A.R. (2001). Coordination between transcription and pre-mRNA processing. FEBS Letters *498*, 179-182.

49. Cvetic,C., Walter,J.C. (2005). Eukaryotic origins of DNA replication: could you please be more specific? Semin.Cell Dev.Biol. *16*, 343-353.

50. Davidovic,L., Bechara,E., Gravel,M., Jaglin,X.H., Tremblay,S., Sik,A., Bardoni,B., Khandjian,E.W. (2006). The nuclear microspherule protein 58 is a novel RNA-binding protein that interacts with fragile X mental retardation protein in polyribosomal mRNPs from neurons. Hum.Mol.Genet. *15*, 1525-1538.

51. De Vries,H., Rüegsegger,U., Hübner,W., Friedlein,A., Langen,H., Keller,W. (2000). Human pre-mRNA cleavage factor II_m contains homologs of yeast proteins and bridges two other cleavage factors. EMBO Journal *19*, 5895-5904.

52. DeLisle,A.J., Graves,R.A., Marzluff,W.F., Johnson,L.F. (1983). Regulation of histone mRNA production and stability in serum-stimulated mouse 3T6 fibroblasts. Mol.Cell Biol. *3*, 1920-1929.

53. DePamphilis,M.L. (1993). Origins of DNA replication in metazoan chromosomes. Journal of Biological Chemistry *268*, 1-4.

54. Diffley,J.F.X., Cocker,J.H. (1992). Protein-DNA interactions at a yeast replication origin. Nature(London) *357*, 169-172.

55. Dominski,Z., Sumerel,J., Hanson,R.J., Marzluff,W.F. (1995). The polyribosomal protein bound to the 3' end of histone mRNA can function in histone pre-mRNA processing. RNA *1*, 915-923.

56. Dominski,Z., Zheng,L.-X., Sanchez,R., Marzluff,W.F. (1999). The stem-loop binding protein facilitates 3' end formation by stabilizing U7 snRNP binding to the histone pre-mRNA. Molecular and Cellular Biology *19*, 3561-3570.

57. Dominski,Z., Erkmann,J.A., Yang,X., Sanchez,R., Marzluff,W.F. (2002). A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing. Genes Dev. *16*, 58-71.

58. Dominski,Z., Yang,X., Kaygun,H., Marzluff,W.F. (2003). A 3' exonuclease that specifically interacts with the 3' end of histone mRNA. Molecular Cell *12*, 295-305.

59. Dominski,Z., Yang,X.C., Purdy,M., Marzluff,W.F. (2003). Cloning and characterization of the Drosophila U7 small nuclear RNA. Proc.Natl.Acad.Sci.U.S.A *100*, 9422-9427.

60. Dominski,Z., Yang,X.C., Marzluff,W.F. (2005). The polyadenylation factor CPSF-73 is involved in histone pre-mRNA processing. Cell *123*, 37-48.

61. Dominski,Z. (2007). Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. Crit Rev.Biochem.Mol.Biol. *42*, 67-93.

62. Donaldson,A.D., Fangman,W.L., Brewer,B.J. (1998). Cdc7 is required throughout the yeast S phase to activate replication origins. Genes and Development *12*, 491-501.

63. Dou,Y., Gorovsky,M.A. (2000). Phosphorylation of linker histone H1 regulates gene expression in vivo by creating a charge patch. Mol.Cell *6*, 225-231.

64. Dutta,A., Bell,S.P. (1997). Initiation of DNA replication in eukaryotic cells. Annual Review of Cell and Developmental Biology *13*, 293-332.

65. Echeverri,C.J., Perrimon,N. (2006). High-throughput RNAi screening in cultured cells: a user's guide. Nat.Rev.Genet. *7*, 373-384.

66. Eckner,R., Ellmeier,W., Birnstiel,M.L. (1991). Mature mRNA 3' end formation stimulates RNA export from the nucleus. EMBO Journal *10*, 3513-3522.

67. Egloff,S., O'Reilly,D., Chapman,R.D., Taylor,A., Tanzhaus,K., Pitts,L., Eick,D., Murphy,S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science *318*, 1777-1779.

68. Egloff,S., O'Reilly,D., Murphy,S. (2008). Expression of human snRNA genes from beginning to end. Biochem.Soc.Trans. *36*, 590-594.

69. Ekholm,S.V., Reed,S.I. (2000). Regulation of G(1) cyclin-dependent kinases in the mammalian cell cycle. Curr.Opin.Cell Biol. *12*, 676-684.

70. Engel,J.D., Sugarman,R.A., Dodgson,J.B. (1982). A chicken histone H3 gene contains intervening sequences. Nature(London) *297*, 434-436.

71. Erkmann,J.A., Sanchez,R., Treichel,N., Marzluff,W.F., Kutay,U. (2005). Nuclear export of metazoan replication-dependent histone mRNAs is dependent on RNA length and is mediated by TAP. RNA *11*, 45-58.

72. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E., Mello,C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans [see comments]. Nature(London) *391*, 806-811.

73. Foley,E., O'Farrell,P.H. (2004). Functional dissection of an innate immune response by a genome-wide RNAi screen. PLoS.Biol. *2*, E203.

74. Fraser,A.G., Kamath,R.S., Zipperlen,P., Martinez-Campos,M., Sohrmann,M., Ahringer,J. (2000). Functional genomic analysis of *C-elegans* chromosome I by systematic RNA interference. Nature(London) *408*, 325-330.

75. Frey,M.R., Matera,A.G. (1995). Coiled bodies contain U7 small nuclear RNA and associate with specific DNA sequences in interphase human cells. Proceedings of the National Academy of Sciences of the United States of America *92*, 5915-5919.

76. Gallie,D.R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev. *5*, 2108-2116.

77. Gao,G., Bracken,A.P., Burkard,K., Pasini,D., Classon,M., Attwooll,C., Sagara,M., Imai,T., Helin,K., Zhao,J. (2003). NPAT expression is regulated by E2F and is essential for cell cycle progression. Mol.Cell Biol. *23*, 2821-2833.

78. Georgiev,O., Birnstiel,M.L. (1985). The conserved CAAGAAAGA spacer sequence is an essential element for the formation of 3' termini of the sea urchin H3 histone mRNA by RNA processing. EMBO Journal *4*, 481-489.

79. Ghule,P.N., Dominski,Z., Yang,X.C., Marzluff,W.F., Becker,K.A., Harper,J.W., Lian,J.B., Stein,J.L., Van Wijnen,A.J., Stein,G.S. (2008). Staged assembly of histone gene expression machinery at subnuclear foci in the abbreviated cell cycle of human embryonic stem cells. Proc.Natl.Acad.Sci.U.S.A *44*, 16964-16969.

80. Gick,O., Krämer,A., Vasserot,A., Birnstiel,M.L. (1987). Heat-labile regulatory factor is required for 3' processing of histone precursor mRNAs. Proceedings of the National Academy of Sciences of the United States of America *84*, 8937-8940.

81. Girard,F., Strausfeld,U., Fernandez,A., Lamb,N.J.C. (1991). Cyclin A is required for the onset of DNA replication in mammalian fibroblasts. Cell *67*, 1169-1179.

82. Godfrey,A.C., Kupsco,J.M., Burch,B.D., Zimmerman,R.M., Dominski,Z., Marzluff,W.F., Duronio,R.J. (2006). U7 snRNA mutations in Drosophila block histone pre-mRNA processing and block oogenesis. RNA *12*, 396-409.

83. Godfrey,A.C., White,A.E., Tatomer,D.C., Marzluff,W.F., Duronio,R.J. (2009). The Drosophila U7 snRNP proteins Lsm10 and Lsm11 play an essential role in development independent of histone pre-mRNA processing. RNA *15*, 1661-1672.

84. Graves,R.A., Pandey,N.B., Chodchoy,N., Marzluff,W.F. (1987). Translation is required for regulation of histone mRNA degradation. Cell *48*, 615-626.

85. Green,L., Van Antwerpen,R., Stein,J., Stein,G., Tripputi,P., Emanuel,B., Selden,J., Croce,C. (1984). A major human histone gene cluster on the long arm of chromosome 1. Science *226*, 838-840.

86. Grunstein,M. (1997). Histone acetylation in chromatin structure and transcription. Nature(London) *389*, 349-352.

87. Hammond,S.M., Bernstein,E., Beach,D., Hannon,G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature(London) *404*, 293-296.

88. Hanson,R.J., Sun,J.-H., Willis,D.G., Marzluff,W.F. (1996). Efficient extraction and partial purification of the polyribosomal-associated stem-loop binding protein bound to the 3' end of histone mRNA. Biochemistry *35*, 2146-2156.

89. Hargous,Y., Hautbergue,G.M., Tintaru,A.M., Skrisovska,L., Golovanov,A.P., Stevenin,J., Lian,L.Y., Wilson,S.A., Allain,F.H. (2006). Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. EMBO Journal *25*, 5126-5137.

90. Harris,M.E., Böhni,R., Schneiderman,M.H., Ramamurthy,L., Schümperli,D., Marzluff,W.F. (1991). Regulation of histone mRNA in the unperturbed cell cycle: Evidence suggesting control at two posttranscriptional steps. Molecular and Cellular Biology *11*, 2416-2424.

91. Heintz,N., Sive,H.L., Roeder,R.G. (1983). Regulation of human histone gene expression: kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle. Molecular and Cellular Biology *3*, 539-550.

92. Heintz,N. (1991). The regulation of histone gene expression during the cell cycle. Biochimica et Biophysica Acta: Gene Structure and Expression *1088*, 327-339.

93. Henry,K.W., Wyce,A., Lo,W.S., Duggan,L.J., Emre,N.C., Kao,C.F., Pillus,L., Shilatifard,A., Osley,M.A., Berger,S.L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev. *17*, 2648-2663.

94. Hernandez,N., Weiner,A.M. (1986). Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. Cell *47*, 249-258.

95. Hirohashi,Y., Wang,Q., Liu,Q., Du,X., Zhang,H., Sato,N., Greene,M.I. (2006). p78/MCRS1 forms a complex with centrosomal protein Nde1 and is essential for cell viability. Oncogene *25*, 4937-4946.

96. Hofmann,J.F., Beach,D. (1994). cdt1 is an essential target of the Cdc10/Sct1 transcription factor: requirement for DNA replication and inhibition of mitosis. EMBO Journal *13*, 425-434.

97. Homesley,L., Lei,M., Kawasaki,Y., Sawyer,S., Christensen,T., Tye,B.K. (2000). Mcm10 and the MCM2-7 complex interact to initiate DNA synthesis and to release replication factors from origins. Genes Dev. *14*, 913-926.

98. Huang,Y., Steitz,J.A. (2001). Splicing Factors SRp20 and 9G8 Promote the Nucleocytoplasmic Export of mRNA. Mol.Cell *7*, 899-905.

99. Huang,Y., Gattoni,R., Stevenin,J., Steitz,J.A. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. Mol.Cell *11*, 837-843.

100. Huberman,J.A., Zhu,J.G., Davis,L.R., Newlon,C.S. (1988). Close association of a DNA replication origin and an ARS element on chromosome III of the yeast, Saccharomyces cerevisiae. Nucleic Acids Research *16*, 6373-6384.

101. Hurt,M.M., Bowman,T.L., Marzluff,W.F. (1991). A common transcriptional activator is located in the coding region of two replication-dependent mouse histone genes. Molecular and Cellular Biology *11*, 2929-2936.

102. Hwang,W.W., Venkatasubrahmanyam,S., Ianculescu,A.G., Tong,A., Boone,C., Madhani,H.D. (2003). A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. Mol.Cell *11*, 261-266.

103. Imai,Y., Kimura,T., Murakami,A., Yajima,N., Sakamaki,K., Yonehara,S. (1999). The CED-4-homologous protein FLASH is involved in Fas-mediated activation of caspase-8 during apoptosis. Nature(London) *398*, 777-785.

104. Imataka,H., Gradi,A., Sonenberg,N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. EMBO Journal *17*, 7480-7489.

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

106. Jackson,P.K., Chevalier,S., Philippe,M., Kirschner,M.W. (1995). Early events in DNA replication require cyclin E and are blocked by $p21^{CIP1}$. Journal of Cell Biology *130*, 755-769.

107. Jacobs,E.Y., Ogiwara,I., Weiner,A.M. (2004). Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. Mol.Cell Biol. *24*, 846-855.

108. Jenuwein,T., Allis,C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

109. Kamath,R.S., Ahringer,J. (2003). Genome-wide RNAi screening in Caenorhabditis elegans. Methods *30*, 313-321.

110. Kamimura,Y., Masumoto,H., Sugino,A., Araki,H. (1998). Sld2, which interacts with Dpb11 in Saccharomyces cerevisiae, is required for chromosomal DNA replication. Mol.Cell Biol. *18*, 6102-6109.

111. Katahira,J., Strässer,K., Podtelejnikov,A., Mann,M., Jung,J.U., Hurt,E. (1999). The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. EMBO Journal *18*, 2593-2609.

112. Kaufmann,I., Martin,G., Friedlein,A., Langen,H., Keller,W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO Journal *23*, 616-626.

113. Kaygun,H., Marzluff,W.F. (2005). Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. Nat.Struct.Mol.Biol. *12*, 794-800.

114. Kaygun,H., Marzluff,W.F. (2005). Translation termination is involved in histone mRNA degradation when DNA replication is inhibited. Mol.Cell Biol. *25*, 6879-6888.

115. Keon,B.H., Schafer,S., Kuhn,C., Grund,C., Franke,W.W. (1996). Symplekin, a novel type of tight junction plaque protein. Journal of Cell Biology *134*, 1003-1018.

116. Kephart,D.D., Marshall,N.F., Price,D.H. (1992). Stability of Drosophila RNA polymerase II elongation complexes in vitro. Molecular and Cellular Biology *12*, 2067-2077.

117. Kiger,A.A., Baum,B., Jones,S., Jones,M.R., Coulson,A., Echeverri,C., Perrimon,N. (2003). A functional genomic analysis of cell morphology using RNA interference. J.Biol. *2*, 27.

118. Kino,T., Chrousos,G.P. (2003). Tumor necrosis factor alpha receptor- and Fas-associated FLASH inhibit transcriptional activity of the glucocorticoid receptor by binding to and interfering with its interaction with p160 type nuclear receptor coactivators. Journal of Biological Chemistry *278*, 3023-3029.

119. Kino,T., Ichijo,T., Chrousos,G.P. (2004). FLASH interacts with p160 coactivator subtypes and differentially suppresses transcriptional activity of steroid hormone receptors. J.Steroid Biochem.Mol.Biol. *92*, 357-363.

120. Kiriyama,M., Kobayashi,Y., Saito,M., Ishikawa,F., Yonehara,S. (2009). Interaction of FLASH with arsenite resistance protein 2 is involved in cell cycle progression at S phase. Mol.Cell Biol. *29*, 4729-4741.

121. Kleiman,F.E., Manley,J.L. (2001). The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. Cell *104*, 743-753.

122. Knoblich,J.A., Sauer,K., Jones,L., Richardson,H., Saint,R., Lehner,C.F. (1994). Cyclin E controls S phase progression and its down-regulation during Drosophila embryogenesis is required for the arrest of cell proliferation. Cell *77*, 107-120.

123. Kolev,N.G., Steitz,J.A. (2006). In vivo assembly of functional U7 snRNP requires RNA backbone flexibility within the Sm-binding site. Nat.Struct.Mol.Biol. *13*, 347-353.

124. Kolev,N.G., Yario,T.A., Benson,E., Steitz,J.A. (2008). Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. EMBO Rep.

125. Koonin,E.V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. Nucleic Acids Research *21*, 2541-2547.

126. Koseoglu,M.M., Graves,L.M., Marzluff,W.F. (2008). Phosphorylation of threonine 61 by cyclin A/Cdk1 triggers degradation of stemloop binding protein at the end of S-phase. Molecular & Cellular Biology *28*, 4469-4479.

127. Krieg,P.A., Melton,D.A. (1984). Formation of the 3' end of histone mRNA by post-transcriptional processing. Nature(London) *308*, 203-206.

128. Kubota,Y., Mimura,S., Nishimoto,S., Masuda,T., Nojima,H., Takisawa,H. (1997). Licensing of DNA replication by a multi-protein complex of MCM/P1 proteins in Xenopus eggs. EMBO Journal *16*, 3320-3331.

129. Kumagai,A., Lee,J., Yoo,H.Y., Dunphy,W.G. (2006). TopBP1 activates the ATR-ATRIP complex. Cell *124*, 943-955.

130. Kuo,M.H., Zhou,J., Jambeck,P., Churchill,M.E., Allis,C.D. (1998). Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. Genes Dev. *12*, 627-639.

131. LaBella,F., Sive,H.L., Roeder,R.G., Heintz,N. (1988). Cell-cycle regulation of a human histone H2b gene is mediated by the H2b subtype-specific consensus element. Genes Dev. *2*, 32-39.

132. Lanzotti,D.J., Kaygun,H., Yang,X., Duronio,R.J., Marzluff,W.F. (2002). Developmental control of histone mRNA and dSLBP synthesis during Drosophila embryogenesis and the role of dSLBP in histone mRNA 3' processing in vivo. Mol.Cell Biol. *22*, 2267-2282.

133. Lanzotti,D.J., Kupsco,J.M., Yang,X.C., Dominski,Z., Marzluff,W.F., Duronio,R.J. (2004). Drosophila stem-loop binding protein intracellular localization is mediated by phosphorylation and is required for cell cycle-regulated histone mRNA expression. Mol.Biol.Cell *15*, 1112-1123.

134. Lee,D.G., Bell,S.P. (1997). Architecture of the yeast origin recognition complex bound to origins of DNA replication. Molecular and Cellular Biology *17*, 7159-7168.

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

136. Levine,B.J., Chodchoy,N., Marzluff,W.F., Skoultchi,A.I. (1987). Coupling of replication type histone mRNA levels to DNA synthesis requires the stem-loop sequence at the 3' end of the mRNA. Proceedings of the National Academy of Sciences of the United States of America *84*, 6189-6193.

137. Li,A., Blow,J.J. (2004). Non-proteolytic inactivation of geminin requires CDK-dependent ubiquitination. Nature Cell Biology *6*, 260-267.

138. Lifton,R.P., Goldberg,M.L., Karp,R.W., Hogness,D.S. (1978). The organization of the histone genes in drosophila melanogaster: functional and evolutionary implications. Cold Spring Harbor Symposia on Quantitative Biology *42*, 1047-1051.

139. Liu,J.L., Murphy,C., Buszczak,M., Clatterbuck,S., Goodman,R., Gall,J.G. (2006). The Drosophila melanogaster Cajal body. Journal of Cell Biology *172*, 875-884.

140. Loidl,P. (1994). Histone acetylation: Facts and questions. Chromosoma *103*, 441-449.

141. Lopez-Rodas,G., Brosch,G., Georgieva,E.I., Sendra,R., Franco,L., Loidl,P. (1993). Histone deacetylase. A key enzyme for the binding of regulatory proteins to chromatin. FEBS Letters *317*, 175-180.

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

143. Macalpine,D.M., Bell,S.P. (2005). A genomic view of eukaryotic DNA replication. Chromosome.Res. *13*, 309-326.

144. MacDonald,C.C., Wilusz,J., Shenk,T. (1994). The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Molecular and Cellular Biology *14*, 6647-6654.

145. Mandel,C.R., Kaneko,S., Zhang,H., Gebauer,D., Vethantham,V., Manley,J.L., Tong,L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature(London) *444*, 953-956.

146. Marahrens,Y., Stillman,B. (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science *255*, 817-823.

147. Marheineke,K., Hyrien,O., Krude,T. (2005). Visualization of bidirectional initiation of chromosomal DNA replication in a human cell free system. Nucleic Acids Research *33*, 6931-6941.

148. Marshall,N.F., Price,D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Molecular and Cellular Biology *12*, 2078-2090.

149. Marzluff,W.F. (1992). Histone 3' ends: essential and regulatory functions. Gene Expression *2*, 93-97.

150. McGarry,T.J., Kirschner,M.W. (1998). Geminin, an inhibitor of DNA replication, is degraded during mitosis. Cell *93*, 1043-1053.

151. Medlin,J., Scurry,A., Taylor,A., Zhang,F., Peterlin,B.M., Murphy,S. (2005). P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. EMBO Journal *24*, 4154-4165.

152. Medlin,J.E., Uguen,P., Taylor,A., Bentley,D.L., Murphy,S. (2003). The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. EMBO Journal *22*, 925-934.

153. Meeks-Wagner,D., Hartwell,L.H. (1986). Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. Cell *44*, 43-52.

154. Melin,L., Soldati,D., Mital,R., Streit,A., Schümperli,D. (1992). Biochemical demonstration of complex formation of histone pre- mRNA with U7 small nuclear ribonucleoprotein and hairpin binding factors. EMBO Journal *11*, 691-697.

155. Merchant,K., Rill,R.L. (1997). DNA length and concentration dependencies of anisotropic phase transitions of DNA solutions. Biophysical Journal *73*, 3154-3163.

156. Mifflin,R.C., Kellems,R.E. (1991). Coupled transcription-polyadenylation in a cell-free system. Journal of Biological Chemistry *266*, 19593-19598.

157. Mitra,P., Xie,R.L., Medina,R., Hovhannisyan,H., Zaidi,S.K., Wei,Y., Harper,J.W., Stein,J.L., Van Wijnen,A.J., Stein,G.S. (2003). Identification of HiNF-P, a key activator of cell cycle-controlled histone H4 genes at the onset of S phase. Mol.Cell Biol. *23*, 8110-8123.

158. Mizzen,C., Kuo,M.H., Smith,E., Brownell,J., Zhou,J., Ohba,R., Wei,Y., Monaco,L., Sassone-Corsi,P., Allis,C.D. (1998). Signaling to chromatin through histone modifications: How clear is the signal? Cold Spring Harbor Symposia on Quantitative Biology *63*, 469-481.

159. Morgan,D.O. (1997). Cyclin-dependent kinases: Engines, clocks, and microprocessors. Annual Review of Cell and Developmental Biology *13*, 261-291.

160. Mowry,K.L., 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.

161. Mowry,K.L., Oh,R., Steitz,J.A. (1989). Each of the conserved sequence elements flanking the cleavage site of mammalian histone pre-mRNAs has a distinct role in the 3'-end processing reaction. Molecular and Cellular Biology *9*, 3105-3108.

162. Mullen,T.E., Marzluff,W.F. (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. Genes Dev. *22*, 50-65.

163. Muller,B., Blackburn,J., Feijoo,C., Zhao,X., Smythe,C. (2007). DNA-activated protein kinase functions in a newly observed S phase checkpoint that links histone mRNA abundance with DNA replication. Journal of Cell Biology *179*, 1385-1398.

164. Murthy,K.G.K., Manley,J.L. (1995). The 160-kD subunit of human cleavage polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes and Development *9*, 2672-2683.

165. Narita,T., Yung,T.M., Yamamoto,J., Tsuboi,Y., Tanabe,H., Tanaka,K., Yamaguchi,Y., Handa,H. (2007). NELF Interacts with CBC and Participates in 3' End Processing of Replication-Dependent Histone mRNAs. Mol.Cell *26*, 349-365.

166. Nelson,D.M., Ye,X., Hall,C., Santos,H., Ma,T., Kao,G.D., Yen,T.J., Harper,J.W., Adams,P.D. (2002). Coupling of DNA synthesis and histone synthesis in S-phase independent of cyclin/cdk2 activity. Molecular and Cellular Biology *22*, 7459-7472.

167. Neuman de Vegvar,H.E., Lund,E., Dahlberg,J.E. (1986). 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. Cell *47*, 259-266.

168. Nishitani,H., Lygerou,Z., Nishimoto,T., Nurse,P. (2000). The Cdt1 protein is required to license DNA for replication in fission yeast. Nature(London) *404*, 625-626.

169. Obradovic,D., Tirard,M., Nemethy,Z., Hirsch,O., Gronemeyer,H., Almeida,O.F. (2004). DAXX, FLASH, and FAF-1 modulate mineralocorticoid and glucocorticoid receptor-mediated transcription in hippocampal cells--toward a basis for the opposite actions elicited by two nuclear receptors? Mol.Pharmacol. *65*, 761-769.

170. Paddison,P.J., Silva,J.M., Conklin,D.S., Schlabach,M., Li,M., Aruleba,S., Balija,V., O'Shaughnessy,A., Gnoj,L., Scobie,K., Chang,K., Westbrook,T., Cleary,M., Sachidanandam,R., McCombie,W.R., Elledge,S.J., Hannon,G.J. (2004). A resource for large-scale RNA-interference-based screens in mammals. Nature(London) *428*, 427-431.

171. Palko,L., Bass,H.W., Beyrouthy,M.J., Hurt,M.M. (2004). The Yin Yang-1 (YY1) protein undergoes a DNA-replication-associated switch in localization from the cytoplasm to the nucleus at the onset of S phase. Journal of Cell Science *117*, 465-476.

172. Pandey,N.B., Marzluff,W.F. (1987). The stem-loop structure at the 3' end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability. Molecular and Cellular Biology *7*, 4557-4559.

173. Pandey,N.B., Williams,A.S., Sun,J.-H., Brown,V.D., Bond,U., Marzluff,W.F. (1994). Point mutations in the stem-loop at the 3' end of mouse histone mRNA reduce expression by reducing the efficiency of 3' end formation. Molecular and Cellular Biology *14*, 1709-1720.

174. Peters,A.H.F.M., O'Carroll,D., Scherthan,H., Mechtler,K., Sauer,S., Schöfer,C., Weipoltshammer,K., Pagani,M., Lachner,M., Kohlmaier,A., Opravil,S., Doyle,M., Sibilia,M., Jenuwein,T. (2001). Loss of the *Suv39h* histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.

175. Peterson,C.L., Laniel,M.A. (2004). Histones and histone modifications. Current Biology *14*, R546-R551.

176. Pilch,D.R., Marzluff,W.F. (1991). Expression of histone-U1 snRNA genes: U1 promoters are compatible with histone 3' end formation. Gene Expression *1*, 41-53.

177. Pillai,R.S., Will,C.L., Lührmann,R., Schümperli,D., 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 Journal *20*, 5470-5479.

178. Pillai,R.S., Grimmler,M., Meister,G., Will,C.L., Luhrmann,R., Fischer,U., Schumperli,D. (2003). Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev. *17*, 2321-2333.

179. Pospiech,H., Grosse,F., Pisani,F.M. (2010). The initiation step of eukaryotic DNA replication. Subcell.Biochem. *50*, 79-104.

180. Proudfoot,N.J., Furger,A., Dye,M.J. (2002). Integrating rnRNA processing with transcription. Cell *108*, 501-512.

181. Ramadan,N., Flockhart,I., Booker,M., Perrimon,N., Mathey-Prevot,B. (2007). Design and implementation of high-throughput RNAi screens in cultured Drosophila cells. Nat.Protoc. *2*, 2245-2264.

182. Ramamurthy,L., Ingledue,T.C., Pilch,D.R., Kay,B.K., Marzluff,W.F. (1996). Increasing the distance between the snRNA promoter and the 3' box decreases the efficiency of snRNA 3'-end formation. Nucleic Acids Research *24*, 4525-4534.

183. Ren,Q.H., Gorovsky,M.A. (2001). Histone H2A.Z acetylation modulates an essential charge patch. Molecular Cell *7*, 1329-1335.

184. Ricke,R.M., Bielinsky,A.K. (2004). Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. Mol.Cell *16*, 173-185.

185. Riquelme,P.T., Burzio,L.O., Koide,S.S. (1979). ADP ribosylation of rat liver lysine-rich histone in vitro. Journal of Biological Chemistry *254*, 3018-3028.

186. Robbins,E., Borun,T.W. (1967). THE CYTOPLASMIC SYNTHESIS OF HISTONES IN HELA CELLS AND ITS TEMPORAL RELATIONSHIP TO DNA REPLICATION. Proc.Natl.Acad.Sci.U.S.A *57*, 409-416.

187. Roberts,J.M., Sherr,C.J. (2003). Bared essentials of CDK2 and cyclin E. Nat.Genet. *35*, 9-10.

188. Robzyk,K., Recht,J., Osley,M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science *287* , 501-504.

189. Romanowski,P., Madine,M.A., Rowles,A., Blow,J.J., Laskey,R.A. (1996). The Xenopus origin recognition complex is essential for DNA replication and MCM binding to chromatin. Current Biology *6*, 1416-1425.

190. Roth,S.Y., Allis,C.D. (1992). Chromatin condensation: Does histone H1 dephosphorylation play a role. Trends Biochem.Sci. *17*, 93-98.

191. Rowles,A., Chong,J.P.J., Brown,L., Howell,M., Evan,G.I., Blow,J.J. (1996). Interaction between the origin recognition complex and the replication licensing system in Xenopus. Cell *87*, 287-296.

192. Ruegsegger,U., Beyer,K., Keller,W. (1996). Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. Journal of Biological Chemistry *271*, 6107-6113.

193. Ruegsegger,U., Blank,D., Keller,W. (1998). Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Molecular Cell *1*, 243-253.

194. Ruhf,M.L., Braun,A., Papoulas,O., Tamkun,J.W., Randsholt,N., Meister,M. (2001). The domino gene of Drosophila encodes novel members of the SWI2/SNF2 family of DNA-dependent ATPases, which contribute to the silencing of homeotic genes. Development *128*, 1429-1441.

195. Saleh,M.C., van Rij,R.P., Hekele,A., Gillis,A., Foley,E., O'Farrell,P.H., Andino,R. (2006). The endocytic pathway mediates cell entry of dsRNA to induce RNAi silencing. Nature Cell Biology *8*, 793-802.

196. Sanchez,R., Marzluff,W.F. (2002). The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. Molecular and Cellular Biology *22*, 7093-7104.

197. Scharl,E.C., Steitz,J.A. (1994). The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. EMBO Journal *13*, 2432-2440.

198. Scharl,E.C., Steitz,J.A. (1996). Length suppression in histone messenger RNA 3'-end maturation: Processing defects of insertion mutant premessenger RNAs can be compensated by insertions into the U7 small nuclear RNA. Proceedings of the National Academy of Sciences of the United States of America *93*, 14659-14664.

199. Schochetman,G., Perry,R.P. (1972). Early appearance of histone messenger RNA in polyribosomes of cultured L cells. Journal of Molecular Biology *63*, 591-596.

200. Schwob,E., Böhm,T., Mendenhall,M.D., Nasmyth,K. (1994). The B-type cyclin kinase inhibitor p40^{SIC1} controls the G1 to S transition in S. cerevisiae. Cell *79*, 233-244.

201. Shiio,Y., Eisenman,R.N. (2003). Histone sumoylation is associated with transcriptional repression. Proc.Natl.Acad.Sci.U.S.A *100*, 13225-13230.

202. Shimono,K., Shimono,Y., Shimokata,K., Ishiguro,N., Takahashi,M. (2005). Microspherule protein 1, Mi-2beta, and RET finger protein associate in the nucleolus and up-regulate ribosomal gene transcription. Journal of Biological Chemistry *280*, 39436-39447.

203. Silva,J.M., Li,M.Z., Chang,K., Ge,W., Golding,M.C., Rickles,R.J., Siolas,D., Hu,G., Paddison,P.J., Schlabach,M.R., Sheth,N., Bradshaw,J., Burchard,J., Kulkarni,A., Cavet,G., Sachidanandam,R., McCombie,W.R., Cleary,M.A., Elledge,S.J., Hannon,G.J. (2005). Second-generation shRNA libraries covering the mouse and human genomes. Nat.Genet. *37*, 1281-1288.

204. Sittman,D.B., Graves,R.A., Marzluff,W.F. (1983). Histone mRNA concentrations are regulated at the level of transcription and mRNA degradation. Proceedings of the National Academy of Sciences of the United States of America *80*, 1849-1853.

205. Song,H., Li,Y., Chen,G., Xing,Z., Zhao,J., Yokoyama,K.K., Li,T., Zhao,M. (2004). Human MCRS2, a cell-cycle-dependent protein, associates with LPTS/PinX1 and reduces the telomere length. Biochem.Biophys.Res.Commun. *316*, 1116-1123.

206. Spadari,S., Weissbach,A. (1975). RNA-primed DNA synthesis: specific catalysis by HeLa cell DNA polymerase alpha. Proc.Natl.Acad.Sci.U.S.A *72*, 503-507.

207. Spalding,J., Kajiwara,K., Mueller,G.C. (1966). The metabolism of basic proteins in HeLa cell nuclei. Proc.Natl.Acad.Sci.U.S.A *56*, 1535-1542.

208. Spycher,C., Streit,A., Stefanovic,B., Albrecht,D., Koning,T.H.W., Schümperli,D. (1994). 3' end processing of mouse histone pre-mRNA: Evidence for additional base-pairing between U7 snRNA and pre-mRNA. Nucleic Acids Research *22*, 4023-4030.

209. Stanley,J.S., Griffin,J.B., Zempleni,J. (2001). Biotinylation of histones in human cells. Effects of cell proliferation. Eur.J.Biochem. *268*, 5424-5429.

210. Strahl,B.D., Allis,C.D. (2000). The language of covalent histone modifications. Nature(London) *403*, 41-45.

211. Streit,A., Koning,T.W., Soldati,D., Melin,L., Schümperli,D. (1993). Variable effects of the conserved RNA hairpin element upon 3' end processing of histone pre-mRNA *in vitro*. Nucleic Acids Research *21*, 1569-1575.

212. Su,C., Gao,G., Schneider,S., Helt,C., Weiss,C., O'Reilly,M.A., Bohmann,D., Zhao,J. (2004). DNA damage induces downregulation of histone gene expression through the G1 checkpoint pathway. EMBO Journal *23*, 1133-1143.

213. Sullivan,E., Santiago,C., Parker,E.D., Dominski,Z., Yang,X., Lanzotti,D.J., Ingledue,T.C., Marzluff,W.F., Duronio,R.J. (2001). Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. Genes Dev. *15*, 173-187.

214. Sullivan,K.D., Steiniger,M., Marzluff,W.F. (2009). A core complex of CPSF73, CPSF100 and symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. Molecular Cell *34*, 322-332.

215. Sullivan,K.D., Mullen,T.E., Marzluff,W.F., Wagner,E.J. (2009). Knockdown of SLBP results in nuclear retention of histone mRNA. RNA *15*, 459-472.

216. Sun,J.-H., Pilch,D.R., Marzluff,W.F. (1992). The histone mRNA 3' end is required for localization of histone mRNA to polyribosomes. Nucleic Acids Research *20*, 6057-6066.

217. Swaminathan,J., Baxter,E.M., Corces,V.G. (2005). The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin. Genes Dev. *19* , 65-76.

218. Syvaoja,J.E. (1990). DNA polymerase epsilon: the latest member in the family of mammalian DNA polymerases. Bioessays *12*, 533-536.

219. Takagaki,Y., Manley,J.L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol.Cell Biol. *20*, 1515-1525.

220. Takayama,Y., Kamimura,Y., Okawa,M., Muramatsu,S., Sugino,A., Araki,H. (2003). GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. Genes Dev. *17*, 1153-1165.

221. Tordera,V., Sendra,R., Perez-Ortin,J.E. (1993). The role of histones and their modifications in the informative content of chromatin. Experientia *49*, 780-788.

222. Triputti,P., Emanuel,B.S., Croce,C.M., Green,L.G., Stein,G.S., Stein,J.L. (1986). Human histone genes map to multiple chromosomes. Proceedings of the National Academy of Sciences of the United States of America *83*, 3185-3188.

223. Turner,B.M. (1993). Decoding the nucleosome. Cell *75*, 5-8.

224. Ueda,K., Omachi,A., Kawaichi,M., Hayaishi,O. (1975). Natural occurrence of poly(ADP-ribosyl) histones in rat liver. Proc.Natl.Acad.Sci.U.S.A *72*, 205-209.

225. Vasserot,A.P., Schaufele,F.J., Birnstiel,M.L. (1989). Conserved terminal hairpin sequences of histone mRNA precursors are not involved in duplex formation with the U7 RNA but act as a target site for a distinct processing factor. Proceedings of the National Academy of Sciences of the United States of America *86*, 4345-4349.

226. Venkataraman,K., Brown,K.M., Gilmartin,G.M. (2005). Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. *19*, 1315-1327.

227. Wagner,E.J., Ospina,J.K., Hu,Y., Dundr,M., Matera,A.G., Marzluff,W.F. (2006). Conserved zinc fingers mediate multiple function of ZFP100, a U7 snRNP associated protein. RNA *12*, 1206-1218.

228. Wagner,E.J., Burch,B.D., Godfrey,A.C., Salzer,H.R., Duronio,R.J., Marzluff,W.F. (2007). A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing. Molecular Cell *28*, 692-699.

229. Walter,J., Newport,J. (2000). Initiation of eukaryotic DNA replication: origin unwinding and sequential chromatin association of Cdc45, RPA, and DNA polymerase alpha. Mol.Cell *5*, 617-627.

230. Wang,Z.-F., Whitfield,M.L., Ingledue,T.I., Dominski,Z., Marzluff,W.F. (1996). The protein which binds the 3' end of histone mRNA: a novel RNA- binding protein required for histone pre-mRNA processing. Genes and Development *10*, 3028-3040.

231. Wang,Z.-F., Ingledue,T.C., Dominski,Z., Sanchez,R., Marzluff,W.F. (1999). Two Xenopus proteins that bind the 3' end of histone mRNA: Implications for translational control of histone synthesis during oogenesis. Molecular and Cellular Biology *19*, 835-845.

232. Wei,Y., Jin,J., Harper,J.W. (2003). The cyclin E/Cdk2 substrate and Cajal body component p220(NPAT) activates histone transcription through a novel LisH-like domain. Mol.Cell Biol. *23*, 3669-3680.

233. Wells,S.E., Hillner,P.E., Vale,R.D., Sachs,A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Molecular Cell *2*, 135-140.

234. Wendelburg,B.W., Marzluff,W.F. (1992). Formation of the 3' end of sea urchin U1 small nuclear RNA occurs independently of the conserved 3' box and on transcripts initiated from a histone promoter. Molecular and Cellular Biology *12*, 4132-4141.

235. Wendelburg,B.W., Marzluff,W.F. (1992). Two promoter elements are necessary and sufficient for expression of the sea urchin U1 snRNA gene. Nucleic Acids Research *20*, 3743-3751.

236. White,A.E., Leslie,M.E., Calvi,B.R., Marzluff,W.F., Duronio,R.J. (2007). Developmental and cell cycle regulation of the Drosophila histone locus body. Mol.Biol.Cell *18*, 2491-2502.

237. Whitfield,M.L., Zheng,L.-X., Baldwin,A., Ohta,T., Hurt,M.M., 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. Molecular and Cellular Biology *20*, 4188-4198.

238. Whitfield,M.L., Kaygun,H., Erkmann,J.A., Townley-Tilson,W.H.D., Dominski,Z., Marzluff,W.F. (2004). SLBP is associated with histone mRNA on polyribosomes as a component of histone mRNP. Nucleic Acids Research *32*, 4833-4842.

239. Williams,A.S., Ingledue,T.C., III, Kay,B.K., Marzluff,W.F. (1994). Changes in the stem-loop at the 3' terminus of histone mRNA affects its nucleocytoplasmic transport and cytoplasmic regulation. Nucleic Acids Research *22*, 4660-4666.

240. Williams,B.R. (1999). PKR; a sentinel kinase for cellular stress. Oncogene *18*, 6112-6120.

241. Wohlschlegel,J.A., Dwyer,B.T., Dhar,S.K., Cvetic,C., Walter,J.C., Dutta,A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science *290*, 2309-2312.

242. Wood,A., Schneider,J., Dover,J., Johnston,M., Shilatifard,A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. Journal of Biological Chemistry *278*, 34739-34742.

243. Woudt,L.P., Patink,A., Kempers-Veenstra,A.E., Jansen,A.E.M., Mager,W.H., Planta,R.J. (1983). The genes encoding histone H3 and H4 in Neurospora crassa are unique and contain intervening sequences. Nucleic Acids Research *11*, 5347.

244. Wu,C.H., Lee,C., Fan,R., Smith,M.J., Yamaguchi,Y., Handa,H., Gilmour,D.S. (2005). Molecular characterization of Drosophila NELF. Nucleic Acids Research *33*, 1269-1279.

245. Wu,R.S., Bonner,W.M. (1981). Separation of basal histone synthesis from S phase histone synthesis in dividing cells. Cell *27*, 321-330.

246. Yang,X.C., Purdy,M., Marzluff,W.F., Dominski,Z. (2006). Characterization of 3'hExo, a 3' exonuclease specifically interacting with the 3' end of histone mRNA. Journal of Biological Chemistry *281*, 30447-30454.

247. Yang,X.C., Burch,B.D., Yan,Y., Marzluff,W.F., Dominski,Z. (2009). FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3' end processing of histone pre-mRNAs. Molecular Cell *36*, 267-278.

248. Yang,X.C., Torres,M.P., Marzluff,W.F., Dominski,Z. (2009). Three proteins of the U7-specific Sm rIng function as the molecular ruler to determine the site of 3' end processing in mammalian histone pre-mRNA. Molecular and Cellular Biology *29*, 4045-4056.

249. Ye,X., Franco,A.A., Santos,H., Nelson,D.M., Kaufman,P.D., Adams,P.D. (2003). Defective S phase chromatin assembly causes DNA damage, activation of the S phase checkpoint, and S phase arrest. Mol.Cell *11*, 341-351.

250. Zanier,K., Luyten,I., Crombie,C., Müller,B., Schümperli,D., Linge,J.P., Nilges,M., Sattler,M. (2002). Structure of the histone mRNA hairpin required for cell cycle regulation of histone gene expression. RNA-A Publication of the RNA Society *8*, 29-46.

251. Zhang,J., Liu,J., Li,X., Li,F., Wang,L., Zhang,J., Liu,X., Shen,L., Liu,N., Deng,Y., Yang,A., Han,H., Zhao,M., Yao,L. (2007). The physical and functional interaction of NDRG2 with MSP58 in cells. Biochem.Biophys.Res.Commun. *352*, 6-11.

252. Zhao,J.Y., Dynlacht,B., Imai,T., Hori,T., Harlow,E. (1998). Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry. Genes and Development *12*, 456-461.

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

254. Zheng,L., Dominski,Z., Yang,X.C., Elms,P., Raska,C.S., Borchers,C.H., Marzluff,W.F. (2003). Phosphorylation of stem-loop binding protein (SLBP) on two threonines triggers degradation of SLBP, the sole cell cycle-regulated factor required for regulation of histone mRNA processing, at the end of S phase. Mol.Cell Biol. *23*, 1590-1601.