CONNECTING NUCLEAR ORGANIZATION AND GENE EXPRESSION: THE ROLE OF THE HISTONE LOCUS BODY IN HISTONE MRNA BIOSYNTHESIS

Deirdre Catherine Tatomer

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 Department of Biology

Chapel Hill
2014

Approved by:

William Marzluff

Robert Duronio

Alan Fanning

A. Gregory Matera

Kevin Slep

ABSTRACT

DEIRDRE CATHERINE TATOMER: Connecting Nuclear Organization and Gene Expression: the Role of the Histone Locus Body in Histone mRNA Biosynthesis
(Under the direction of William F. Marzluff and Robert J. Duronio)

Execution of gene expression involves multiple reactions, many of which are mediated through *cis* elements in DNA or RNA. Organization within the nucleus, both in the arrangement of DNA and the concentration of *trans* factors in discrete nuclear environments called nuclear bodies, facilitates aspects of gene expression such as transcription, RNP metabolism and pre-mRNA processing. These substructures, which are visible under the light microscope, potentially promote the efficiency and fidelity of a reaction. To investigate this proposed function of nuclear bodies (NBs), I studied the relationship between the *Drosophila* replicative histone genes and their associated nuclear body, the histone locus body (HLB). The genes encoding the five histone proteins are tandemly arrayed in 100 copies. The replicative histone genes are highly expressed during S phase of the cell cycle, and encode the only known mRNAs that do not end in a poly (A) tail. We defined a 297 bp sequence that encompassed the *H3-H4* bidirectional promoter and demonstrated that activity from this promoter was necessary and sufficient for HLB maturation as well as for activation of the neighboring *H2a-H2b* gene pair. Thus the HLB plays a role in coordinating gene activation within the histone repeat. I then evaluated the *in vivo* consequence of preventing accumulation of a critical histone pre-mRNA processing *trans* factor, FLASH, in the HLB. When FLASH was present at wild type levels in the nucleus, but not localized to the HLB, longer, unprocessed nascent transcripts accumulated at the histone locus. I suggest that the presence of these pre-mRNAs indicates a slower rate of histone pre-mRNA processing. Finally, I characterized the phenotypes of mutants in a pre-

mRNA processing factor, Symplekin (Sym), involved in histone pre-mRNA processing, cleavage and polyadenylation and cytoplasmic polyadenylation. I established an *in vivo* system for testing candidate separation of function Sym mutations. Sym was first discovered at tight junctions, and I provide evidence that this localization indicates likely participation in cytoplasmic polyadenylation. Together this work provides insight into how specific regulatory complexes containing common factors assemble and the role of the HLB in the metabolism and regulation of histone mRNAs.

ACKNOWLEDGMENTS

I am deeply grateful to all of the people who have supported and encouraged me along my path towards this degree. I would first like to thank my mentors, Bill Marzluff and Bob Duronio. By allowing me to join both labs, they gave me the intellectual and physical space to learn how to define scientific questions and best approach them. I appreciate their patience with my experimental wandering and that they allowed me the time to learn, with their guidance, how to navigate a project to completion. Their infectious love of science and enthusiasm for teaching fosters a wonderful environment that I hope will exist in my future workplaces. Thank you Bob and Bill for an invaluable experience.

I would also like to acknowledge the amazing scientific community that I have had the privilege of participating in during my time at UNC. This includes, but is not limited to, members of the Marzluff and Duronio labs, my committee (Greg Matera, Alan Fanning and Kevin Slep), Jeff Sekelsky and his lab, the Lieb lab, and Tony Perdue, who taught me how to use a microscope to capture the images present in this dissertation. I also thank three wonderful undergraduates, Michelle Lin, Alison Witkowski, and Anna Orlando for their hard work and optimism. They all contributed data to this thesis and enriched my time in the Marzluff and Duronio labs. In general, the time everyone took to brainstorm solutions for hurdles in my projects, dispense advice and discuss science with me is greatly appreciated.

I am indebted to Christina Tzagarakis-Foster and Aleksandra Cvoro. Their training and encouragement was instrumental in my pursuing a PhD. I also thank Dale Leitman for the

opportunity to work with Christina and Aleksandra in his lab at UCSF. I am so lucky to have them as lifelong mentors and friends.

I also thank all of my friends who have entertained, distracted and supported me through graduate school. This group encompasses members of both labs, fellow graduate students and lovely locals that I have gotten to know during my time here in North Carolina. They have brilliantly complemented my academic experience.

Finally, I thank my parents Bill and Mary, my sister Meghan, and my brother Andy. Their genuine excitement about my pursuits was motivating and their ability to help me recharge during visits home was essential and greatly appreciated. Thank you all for your love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aa	amino acid
bcd	*bicoid*
cas	*castor*
CB	cajal body
cDNA	complementary DNA
CFI_m	Cleavage factor I
CF II_m	Cleavage factor II
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulation factor
CTD	C-terminal domain
FLASH	*FLICE-associated huge protein*
grk	*gurken*
HCC	histone cleavage complex
HDE	histone downstream element
HLB	histone locus body
HO-1	*haem oxygenase*
LacI	Lac Repressor
LacO	LacOperator
MPEF	mouse primary embryonic fibroblast
mxc	*multi sex combs*
ncRNA	non coding RNA
NB	nuclear body
NELF	*negative elongation factor*
nos	*nanos*
osk	*oskar*

PAP	poly (A) polymerase
PAS	poly (A) signal
PIPKIα	*type I phosphatidylinositol 4-phosphate 5-kinase*
PML	*promyelocytic leukemia*
poly (A)	polyadenosine
PtdIns4,5P_2	Phosphatidylinositol-4,5-bisphosphate
scaRNA	small CB-specific RNA
SL	stem loop
SL1	promoter selectivity factor
SLBP	*stem loop binding protein*
snRNP	small nuclear ribonucleoprotein particle
SRE	Smaug response element
Star-PAP	*Speckle Targeted PIPKIα Regulated-Poly(A) Polymerase*
sym	*symplekin*
tBHQ	tert-butyl hydroquinone
TBP	*TATA Binding Protein*
TRF2	*TATA-box-binding protein)-related Factor 2*
YPS	*ypselon schachel*

CHAPTER 1: INTRODUCTION

Overview

Every cell in a complex organism contains an identical blueprint encoded in its DNA, and variation in how this information is accessed and employed generates diversity in cellular function. The genome contains both regulatory sequences as well as those that code for genes. Many factors involved in gene expression function by recognizing and interacting with sequences in DNA. This theme of directing activity through motifs or structures in a nucleic acid can be extended to RNA, as many of the events required to form a mature RNA are mediated through signals in the transcript. Therefore, a fundamental question in the field of gene expression is how do *trans* factors find and regulate a target gene at the right time and place.

One approach to understanding this question is to explore the connection between nuclear architecture and gene expression. The information encoded in DNA is not always instantly available. DNA is wrapped around histone proteins, condensed into chromosomes and packaged in the nucleus of the cell. Thus, a sequence embedded in a nucleosome may not be accessible to transcription factors. The location of a gene within the nucleus can also impact expression. Regions of the genome associated with the nuclear periphery are often inactive (Van de Vosse et al., 2011). In fact, a recent study of the hunchback gene visualized movement of this locus to the periphery after transcriptional repression (Kohwi et al., 2013). In accordance with observations that DNA is actively positioned in the nucleus, advances in technology now indicate that arrangement of chromosomes within the nucleus is not random (Dekker et al., 2013; Marshall et al., 1997). Certain regions of the genome, though not close together when mapped on a linear piece of DNA, cluster in three-dimensional space.

Overlaid on the chromatin scaffold are structures visible under the light microscope. First described by Santiago Ramón y Cajal in 1903, these structures, termed “accessory bodies” and subsequently nuclear bodies (NB), contain high concentrations of protein and often nucleic acid, but are not bound by a membrane (Matera et al., 2009). A number of NBs have been identified and characterized, many of which contain factors involved in gene expression (Morimoto and Boerkoel, 2013). The presence of NBs implies that the genome is not in contact with a homogeneous pool of trans factors but instead is surrounded by discrete environments. This varied accessibility to trans factors provides another mode of regulating interplay between a factor and target gene. Understanding how nuclear bodies interact with the genome and nascent RNA transcripts is an open area of investigation.

Here, I use the replicative histone genes as model to explore the relationship between a nuclear body and gene expression. The metazoan replicative histone genes are the only known messenger RNAs that do not end a polyadenosine (poly (A)) tail but instead end in a conserved stem loop structure (Marzluff et al., 2008). Factors involved in the biosynthesis of this special class of mRNA are concentrated in a nuclear body, the histone locus body (HLB) (Liu et al., 2006). This thesis project addresses three questions: Is the formation of the HLB a cause or consequence of histone mRNA expression? Does localizing a *trans* factor to the HLB affect the rate of histone 3’end formation? How does a protein involved in generating both poly (A) and histone mRNA 3’ends, Symplekin, specifically incorporate into the histone processing machinery and localize to the HLB?

In this chapter, I will introduce our current understanding of nuclear body assembly and function. I will also describe the known mechanisms of 3’end formation and how this aspect of mRNA biosynthesis influences gene expression. Finally, I will focus on the *Drosophila* replicative-histone genes and HLB as a model for understanding the connection between nuclear organization and gene expression.

Nuclear Body Assembly

Nuclear bodies are defined by their components, both because they mark the structure and because the reactions that the concentrated factors participate in suggest possible biological function(s) for each nuclear body. Discovery of NB components is an ongoing and valuable endeavor as it provides new tools for studying NBs and expands the mechanistic details of nuclear body function (Fong et al., 2013). Understanding the assembly and maintenance of nuclear bodies lays the foundation for assessing NB function.

Nuclear bodies are built through interactions between components (Matera et al., 2009). Different experimental approaches have been used to determine the nature of assembly. One strategy is to first deplete or genetically remove a factor from a cell or animal and then determine if the remaining known components of the NB co-localize. This approach was used to characterize coilin, a fundamental component of the Cajal body (CB). *Coilin* mutant flies are viable (Liu et al., 2009). *Coil* -/- mice are also fertile, however they were not detected in expected mendelian ratios and displayed reduced fertility and fecundity (Tucker et al., 2001; Walker et al., 2009). Although there were differences in developmental phenotypes, immunostaining in both systems revealed that other CB markers such as SMN and components of snRNPs did not form foci in the absence of Coilin. These studies indicated that coilin was a core component of the *Drosophila* CB.

Another nuclear body, the promyelocytic leukemia (PML) body, was also shown to have a core protein component, PML (Bernardi and Pandolfi, 2007). In mouse primary embryonic fibroblasts (MPEFs), PML colocalizes with Sp100, Daxx, and ISG20 in discrete foci. Removal of PML through gene targeting resulted in diffuse localization of these factors in PML-/- MPEFS, suggesting a role for this protein in assembling or maintaining the NB (Ishov et al., 1999; Zhong et al., 2000). These examples indicate a role for specific proteins in NB assembly.

An alternate tactic for studying nuclear body formation involves artificially tethering a protein or RNA to an ectopic genomic location followed by screening for recruitment of the

remaining components to that location (Kaiser et al., 2008; Shevtsov and Dundr, 2011). In the initial study, CB components such as Coilin and SMN were fused to the Lac Repressor (LacI) and GFP. LacI binds to the Lac Operator (LacO), and expression of the fusion protein in a cell with 256 LacO repeats resulted in a detectable ectopic focus (Kaiser et al., 2008). Antibody staining for endogenous CB components was then used to screen for co-localization with the LacO array to indicate NB formation. These experiments showed the ability of these tethered components to nucleate the CB, but only in the presence of Coilin and SMN in the cell.

A subsequent study expanded these observations to other NBs and emphasized the tethering of RNA (Shevtsov and Dundr, 2011). RNA was localized to the repeat through inclusion of an MS2 loop in the transcript and expression of a GFP-LacI-MS2 fusion protein. These experiments again indicated the ability of individual NB components to nucleate structures such as the HLB, speckle, paraspeckle and nuclear stress body. For example, Shevtsov and Dundr (2011) showed that histone H2b mRNA could recruit HLB components NPAT and FLASH. Overall, these studies showed that protein or RNA components could participate in assembling a NB. While these experiments show what can happen in a cell, they do not necessarily reflect the endogenous biogenesis of a NB.

A subset of NBs, such as the nucleolus, associate with gene clusters. Factors required for rRNA biogenesis are concentrated in the nucleolus and this NB is associated with clustered repeats of the 18S and 28S rRNA genes. In *Drosophila,* P Element mediated insertion of a single copy of an rRNA gene to an ectopic location resulted in accumulation of transcripts as well as detection of “nucleolar like material” at the site of insertion (Karpen et al., 1988). This suggested that a single gene could recruit the factors necessary to activate gene expression. It was subsequently shown in cultured HT1080 cells that UBF, an rRNA transcription factor, accumulated on an ectopic array of Xenopus ribosomal intergenic sequence (Mais et al., 2005). This interaction resulted in recruitment of RNA Pol I and promoter selectivity factor (SL1), but not accumulation of another nucleolus component, fibrillarin, which is involved in processing the

pre-rRNA. Incomplete assembly of the structure in the absence of the coding region of the genes, as well as dispersal of nucleolar compartments after inhibition of transcription and RNA processing suggest a role for transcription or nascent RNA in nucleolus formation and maintenance (Hernandez-Verdun, 2006).

Although the CB is not consistently associated with a single genomic location, it has been detected at the U1 and U2 snRNA genes, U3 snoRNA genes and the replicative-histone genes (Callan and Gall, 1991; Frey and Matera, 1995; Gall et al., 1981; Gao et al., 1997; Smith et al., 1995). Insertion of an ectopic array of repeated U2 snRNA genes resulted in association of the exogenous DNA and a CB (Frey et al., 1999). The association required an intact promoter as well as the coding region of the snRNA (Frey and Matera, 2001). This suggests that a nascent RNA plays a role in CB dynamics.

Transcription was also required for assembly of another nuclear body, the paraspeckle. Induction of a non coding RNA (ncRNA), Men e/b, resulted in the colocalization of paraspeckle components PSP1, P54nrb, PSF and PSP2 with the ectopic RNA (Mao et al., 2011). All of these approaches have contributed to our understanding of NB formation. This collection of experiments also highlights the role of both protein and nucleic acid in assembling a NB. Identifying new components of NBs and understanding how they contribute to forming and/or maintaining a microenvironment is an ongoing pursuit of the NB field. Defining the domain(s) that localizes a factor to a NB is one of the next steps in understanding NB function as it opens up experimental approaches for isolating the contribution of a NB to its related biological process.

Nuclear Body Function

The biology associated with a number of nuclear bodies has been characterized. For example, the CB is the site of snRNP maturation, the nucleolus is the site of rRNA biosynthesis, the HLB mediates histone mRNA biosynthesis and paraspeckles retain adenosine to inosine (A-to-I)

hyperedited mRNAs (Mao et al., 2011). How the environment created by a NB contributes to a biological reaction is still an open question. Concentrating the elements of a common process can increase the probability of interactions between components, and this is a proposed function for nuclear bodies. Increasing the local concentration of factors is also postulated to influence the rates of biological reactions (Machyna et al., 2013; Mao et al., 2011; Matera et al., 2009).

Studies of the CB provide evidence for the hypothesis that NBs increase the rate of a biological reaction. One function of the CB is to mediate small nuclear ribonucleoprotein particle (snRNP) assembly. An efficient splicing reaction requires the preassembled U4/U6.U5 tri-snRNP. Stepwise production of this complex first requires SART3 to mediate formation of the U4/U6 di-snRNP intermediate (Stanek and Neugebauer, 2004). Prp31 is then necessary for the subsequent U5 addition to complete tri-snRNP assembly (Schaffert et al., 2004). Cell biological measurements combined with mathematical modeling show that the rate of di and tri-snRNP assembly increased by an order of magnitude in the CB (Klingauf et al., 2006; Novotny et al., 2011). Experiments in zebrafish indicate that CB function is essential during embryogenesis. Depletion of Coilin by morpholino injection resulted in splicing defects and embryonic lethality. snRNP levels did not change; however, their assembly was affected and this lethal manipulation of CB function could be rescued by injection of pre-assembled human snRNPs (Strzelecka et al., 2010).

It is possible that the lethality in zebrafish is due to a splicing defect; however loss of Coilin has different consequences in different model systems. In contrast to the zebrafish result, half of *Coil* -/- mice died at a late developmental stage of gestation. Those mice that did survive had reduced fertility and fecundity (Tucker et al., 2001; Walker et al., 2009). How disruption of the CB contributed to this phenotype has not been resolved. Mutation of the *coilin* gene in *Drosophila* was not lethal (Liu et al., 2009). An snRNA maturation reaction associated with the CB, modification of snRNAs guided by scaRNAs, was not affected in these mutant animals,

suggesting that concentrating factors in the CB was dispensable for this reaction (Deryusheva and Gall, 2009). Whether snRNA modification happened as efficiently in the absence of the CB is yet to be determined.

The Impact of 3' End Formation on Gene Expression

Production of a mature eukaryotic mRNA requires addition of an RNA 7-methylguanosine cap, removal of introns (if present) by splicing and, with the exception of the histone mRNAs, cleavage and polyadenylation. The 3' UTR of an mRNA often contains information that regulates the stability, localization and translation efficiency of the transcript. Therefore, where the processing machinery defines the end of an mRNA post transcriptionally affects expression of the gene.

Studies in *Drosophila* provide numerous examples of these mechanisms. The first seven nuclear divisions of *Drosophila* development occur in the absence of transcription (Pritchard and Schubiger, 1996); therefore during this time all gene expression is mediated through regulation of transcripts deposited into the egg by the mother. Four genes expressed at this stage of development have been extensively characterized and are critical for establishing the embryonic axis: *oskar (osk)*, *nanos (nos)*, *bicoid (bcd),* and *gurken (grk)*. The transcripts from each of these genes are positioned during oogenesis, and their translation ultimately establishes protein gradients that provide distinct cellular environments in the early syncytial embryo (Kugler and Lasko, 2009), and the 3'UTRs of these mRNAs are essential for this process.

Trafficking an mRNA to a particular region of the cell ensures localized accumulation of the encoded protein. Extensive *in situ* analysis of mRNA in the fly embryo indicated widespread employment of this mechanism (Lecuyer et al., 2007). During *Drosophila* oogenesis, polyploid nurse cells synthesize the mRNA and protein required for early embryogenesis. This material, including *osk, nos, bcd and grk* mRNA, is deposited in the oocyte. Within the developing egg, *osk and nos* mRNA localize to the posterior region, *bcd* mRNA accumulates at the posterior of

the oocyte and *grk* mRNA localizes to the anterodorsal oocyte corner (Kugler and Lasko, 2009). This positioning localizes subsequent protein accumulation to distinct regions of the developing animal.

Detection of a message does not necessarily indicate presence of a protein, as numerous mechanisms of translational control have also been described. Multiple aspects of translational regulation control *osk* expression (Wilhelm and Smibert, 2005). The protein CUP, which binds the initiation factor eIF4E-BP, represses *osk* translation by preventing assembly of the translation initiation complex on the message. The length of the poly (A) tail at the end of an mRNA can also affect translation (Gebauer and Richter, 1995). A critical component of this machinery is CPEB, which binds to the 3' UTR of an mRNA and can recruit the cleavage and polyadenylation machinery, as well as a poly (A) polymerase to extend the poly (A) tail. The *Drosophila* homologue of CPEB is Orb, and this protein is required for *osk* translation. These mechanisms contribute to temporal and localized production of Osk.

Stability of an mRNA can also affect expression. The 3'UTR of the *nos* mRNA contains two Smaug response elements (SREs) that mediate Smaug binding to the message. This protein recruits the CCR4-Not deadenylation complex, which removes the poly (A) tail, initiating mRNA decay (Zaessinger et al., 2006). It was also shown that the piRNA pathway also mediates degradation of *nos* mRNA in the *Drosophila* embryo (Rouget et al., 2010). This aspect of mRNA also ensures precise Nos expression.

Overall, post transcriptional regulation of mRNA levels and function contributes extensively to gene expression. While the examples provided were from *Drosophila,* these mechanisms are widespread among tissues and species. As RNA sequencing experiments indicate an increasing number of mRNAs containing alternative polyadenylation sites, understanding how the cell determines the end of a message is an active area of investigation.

Mechanisms of 3' End Formation

Currently, there are three characterized mechanisms for mRNA 3' end formation: canonical cleavage and polyadenylation, cleavage followed by STAR-PAP polyadenylation in stress conditions, and histone pre-mRNA processing. All three of these processes are directed by *cis* elements in the 3'UTR of an mRNA and ultimately recruit and activate CPSF-73, the endonuclease that cleaves the pre-mRNA to form the end of the mRNA. The *trans* factors that recognize the encoded sequences provide specificity for the reaction, assembling distinct molecular machines that define the 3'end. Here, I will summarize the requirements for and participants in each of these reactions (Figure 1.1).

Cleavage and Polyadenylation

The untemplated addition of polyadenosine to the 3' end of every mature mRNA, with the exception of histone mRNA, requires two reactions: cleavage and polyadenylation (Fig. 1.1A). Hybridization of oligo-dT to the poly (A) tail allowed for purification of mRNA as well as a means for priming a message to synthesize complementary DNA (cDNA) (Aviv and Leder, 1972; Proudfoot, 1976). Evaluating the globin genes indicated that eukaryotic mRNA contained a non coding sequence (3'UTR) and a sequence that was soon after shown to be critical for 3' end formation, AAUAAA (Proudfoot and Brownlee, 1976; Proudfoot and Longley, 1976). This poly (A) signal (PAS) and the G/U rich downstream element, flank the cleavage site, which primarily terminates the encoded message in a CA (Chen et al., 1995; Gil and Proudfoot, 1984; Gil and Proudfoot, 1987; McLauchlan et al., 1985). The G/U rich element was identified through deletion of sequences after the cleavage site, however a motif for this element is not always well defined. The PAS is typically located 10-30 nucleotides upstream of the cleavage site, while the G/U rich element is often located 30 nucleotides downstream of the cleavage site (Mandel et al., 2008). Additionally, other sequence elements, the auxiliary upstream and auxiliary downstream elements, have been shown to increase the efficiency of cleavage and polyadenylation (Hu et

al., 2005). How sequences other than the two critical *cis* elements in the 3'UTR modulate the site of cleavage and polyadenylation is still an open area of investigation. As RNA sequencing data amasses, it is clear that many genes contain multiple cleavage and polyadenylation sites, and usage often changes with cell type (Hu et al., 2005; Lianoglou et al., 2013). How sequences in the pre-mRNA mediate these differences is yet to be determined.

A number of *trans* factors assemble on the 3' end of a pre-mRNA to direct cleavage and polyadenylation. Identification of the poly(A) polymerase (PAP) preceded discovery of the PAS (Winters and Edmonds, 1973; Winters and Edmonds, 1973), however a functional assay for cleavage and polyadenylation proved instrumental in biochemically isolating the complexes involved in 3'end formation (Moore and Sharp, 1985). Incubating *in vitro* transcribed mRNA with HeLa cell nuclear extracts resulted in cleavage and polyadenylation, and this assay provided a means for identifying active components of the machinery from purified cell fractions (Christofori and Keller, 1988; Gilmartin and Nevins, 1989; Takagaki et al., 1988; Takagaki et al., 1989). Along with PAP, four complexes and an additional protein were isolated: Cleavage and polyadenylation specificity factor (CPSF, 5 proteins), Cleavage stimulation factor (CstF, 3 proteins), Cleavage factor I (CF $I_{m,}$ 3 proteins), Cleavage factor II (CF II_m 2 proteins), and the protein Symplekin (Mandel et al., 2008).

Anchored to the pre-mRNA by *cis* elements, this macromolecular machine assembles around key protein-RNA interactions. CPSF binds to the PAS through the largest member, CPSF-160 (Moore et al., 1988; Murthy and Manley, 1995), and CstF component CstF64 binds to the G/U rich element (MacDonald et al., 1994). CPSF-73 and its heterodimeric binding partner CSPF-100 form the endonuclease that cleaves the message (Dominski et al., 2005; Mandel et al., 2006; Ryan et al., 2004). Addition of the tail by PAP does not require cleavage factors *in vitro,* though this protein is often required for the *in vitro* cleavage reaction (Christofori and Keller, 1988; Christofori and Keller, 1989; Gilmartin and Nevins, 1989; Raabe et al., 1991; Raabe et al., 1994; Ryner et al., 1989; Wahle, 1991). PAP association with the cleavage

machinery is mediated through interaction with CPSF160 (Murthy and Manley, 1995) and hFIP (Kaufmann et al., 2004). Contacts between members of CPSF and CstF as well as with CF I_m, CF II_m and Symplekin are required for efficient 3' end formation (Mandel et al., 2008). A recent study identified the proteins associated with the SV40 late and adenovirus L3 pre-mRNAs by mass spectrometry and the number of proteins associated with the 3' end increased from 14 to 85 (Shi et al., 2009). While a subset of these proteins, such as WRD33 (which is a previously undiscovered subunit of CPSF), were potential homologues of known yeast poly(A) factors, the vast majority of assembled proteins may be involved in coordinating cleavage and polyadenylation with other transcriptional events.

Interactions between the cleavage and polyadenylation machinery and factors involved in upstream transcriptional mechanisms such as initiation and elongation suggest coordination of transcription and 3' end formation. CPSF-160, the protein that binds the PAS, interacts with a core promoter element, TFIID (Dantonel et al., 1997). CPSF-160, as well as CstF components CstF-50 and CstF-77, also associate with the C-terminal domain (CTD) of RNA polymerase (McCracken et al., 1997). The CTD of RNA Pol II is subject to various posttranslational modifications and provides a dynamic platform for complex assembly as RNA polymerase moves along a gene (Heidemann et al., 2013). The association of cleavage and polyadenylation components with a gene before the transcription of *cis* regulatory sites potentially provides a regulatory mechanism for defining the 3' end of the message, particularly in the cases where multiple signals are present in the 3' UTR. Overall, the processes of cleavage and polyadenylation require a variety of inputs ensuring expression of the correct transcript.

Cleavage Followed by STAR-PAP Polyadenylation

Not all poly (A) tails are added to mRNA by PAP. Depletion of a second poly (A) polymerase, Speckle Targeted PIPKIα Regulated-Poly (A) Polymerase (Star-PAP) followed by microarray analysis revealed altered levels of over 2,000 mRNA transcripts (Mellman et al., 2008). A

significant portion of the messages encoded genes involved in detoxification and / or stress response. Validation and further analysis of one of the targets, haem oxygenase-1 (HO-1) showed that STAR-PAP directly bound to the pre-mRNA upstream of the PAS and depletion of this enzyme resulted in read-through HO-1 transcripts. The same group subsequently showed that STAR-PAP is also required for cleavage of the HO-1 transcript. Immunoprecipitation of STAR-PAP indicated that a subset of poly (A) proteins, CPSF-160, CPSF-100, CPSF-73, CPSF-30, CstF-64 and Symplekin assembled into this complex (Laishram and Anderson, 2010). These data showed that while the core processing machinery (CPSF and CstF) is required for polyadenylation, the processing machinery for polyadenylated mRNA is not uniform for every transcript.

STAR-PAP was identified by yeast two hybrid through its interaction with the nuclear speckle-targeting region of a type I phosphatidylinositol 4-phosphate 5-kinase (PIPKIa). Notably, both of these proteins co-localize in a nuclear structure, the nuclear speckle. PIPKIa generates Phosphatidylinositol-4,5-bisphosphate (PtdIns4,5P_2) and this signaling molecule stimulates STAR-PAP polymerase activity (Mellman et al., 2008). Treatment of cells with tert-butyl hydroquinone (tBHQ) induces an antioxidant response in cells, including induction of STAR-PAP mediated HO-1 expression (Mellman et al., 2008). Therefore, a component of the stress response involves activating a unique 3' end processing mechanism. While the STAR-PAP reaction is distinct, all factors but the polymerase are involved in canonical cleavage and polyadenylation. Understanding why this complex is required for RNAs that have a consensus PAS and determining if concentrating this enzyme in the nuclear speckle affects the reaction remain intriguing questions.

Histone pre-mRNA Processing

The histone mRNAs are the only known mRNAs that do not end in a poly (A) tail (Fig. 1.1B). They instead end in a conserved stem loop (SL) sequence. Like the poly (A) signal, this

conserved sequence is present at the 3' end of histone transcripts (Marzluff et al., 2008). After identification of the SL and purine rich element as critical for histone mRNA 3' end formation (Birchmeier et al., 1982), it was subsequently shown by injection of exogenous histone pre-mRNA into frog oocytes that the histone genes were generated by endonucleolytic cleavage, and not transcription termination (Krieg and Melton, 1984). In fact, this was the first experiment to identify endonucleolytic cleavage as the means of defining the end of an mRNA and separated 3' end formation from transcription termination. Also, like polyadenylated mRNAs, these experiments show that the histone mRNA cleavage site is also flanked by two *cis* elements that assemble the complex that ultimately results in cleavage of the mRNA.

Additional experiments in the frog oocyte were instrumental in identifying one of the *trans* factors that directly bind the histone pre-mRNA transcript. Cleavage of sea urchin H3 pre-mRNA expressed from the sea urchin gene cluster injected into the Xenopus oocyte, required co-injection of components from a sea urchin nuclear extract. This activity was ultimately identified as a 60nt RNA, which they named U7 snRNA (Galli et al., 1983; Strub et al., 1984; Strub and Birnstiel, 1986). After the site of cleavage is a purine rich region called the histone downstream element (HDE), and it was also shown in this system, and in HeLa cell extracts, that U7 snRNA base pairs with the HDE to direct histone pre-mRNA processing (Mowry and Steitz, 1987; Schaufele et al., 1986). While it was known that the U7snRNA associated with Sm proteins (Strub and Birnstiel, 1986), the full identify of the Sm ring was not known until 2003 (Pillai et al., 2001; Pillai et al., 2003), when the Lsm10 and Lsm11 proteins were identified. The Sm ring of proteins in this snRNP matches those of the splicing snRNPs with two exceptions; Sm D1 and D2 are replaced with Lsm10 and Lsm11. It was also shown that the larger of these two unique components, Lsm11, contained an N-terminal region required for histone pre-mRNA processing and a C-terminal region required for assembly of the U7 snRNP (Pillai et al., 2003). Therefore, the U7 snRNP participates in histone pre-mRNA processing through snRNA base pairing to the pre-mRNA as well as through the N-terminus of Lsm11.

The other *cis* element required for histone pre-mRNA processing is the stem loop. The identity of the protein that binds this structure, Stem Loop Binding Protein (SLBP) was discovered by the yeast-three hybrid system (Martin et al., 1997; Wang et al., 1996). Recently, the molecular details of this interaction have been revealed by crystallography (Tan et al., 2013). SLBP is required for histone pre-mRNA processing (Dominski et al., 1999; Sullivan et al., 2001). In addition to a role in histone pre-mRNA processing, SLBP is also required for export (Sullivan et al., 2009), translation (Sanchez and Marzluff, 2002), and degradation (Mullen and Marzluff, 2008) of histone mRNA. Therefore, SLBP mediates the histone mRNA lifecycle.

Cleavage of a histone pre-mRNA substrate by mammalian nuclear extracts provided a system for biochemical identification of additional components involved in histone pre-mRNA processing (Gick et al., 1986). Heat inactivation of the extract abolished processing activity, however this inactivated extract could complement a nuclear extract depleted of Sm proteins (Gick et al., 1987). This suggested the presence of additional factor(s), termed the heat labile factor (HLF) in the histone pre-mRNA cleavage complex.

A critical shared component of all pre-mRNA 3' end formation mechanisms is the endonuclease CPSF-73. Crosslinking experiments with a histone pre-mRNA substrate identified CPSF-73 as the histone cleavage factor (Dominski et al., 2005). At this time, CPSF was the suspected endonuclease for cleavage of polyadenylated mRNA (Ryan et al., 2004) and this was confirmed by crystallography (Mandel et al., 2006).

Just after the discovery of CPSF-73 as the histone pre-mRNA cleavage factor, purification, fractionation and analysis by mass spectrometry of the HLF attributed this activity to members of CPSF and CstF as well as Symplekin (Kolev and Steitz, 2005). Complementation of a heat inactivated extract by *in vitro* transcribed and translated Symplekin indicated a critical role for Symplekin in histone pre-mRNA processing, either as the missing critical component, or as a scaffold that re-assembled the machinery required for histone pre-mRNA 3' end formation.

This study showed that, in addition to CPSF-73, the histone pre-mRNA processing complex and cleavage and polyadenylation machinery share components.

Encoded in the *Drosophila* intergenic region after every HDE are cryptic polyadenylation signals (Lanzotti et al., 2002; Sullivan et al., 2001) (Fig. 1.2). Disruption of histone pre-mRNA processing, either by mutation or depletion of factors such as SLBP or U7snRNP, results in readthrough transcripts that are polyadenylated (Godfrey et al., 2006; Godfrey et al., 2009; Lanzotti et al., 2002; Sullivan et al., 2001) (Figure 1.2). Mis-processed histone mRNA is exported and translated, minimally supporting development of mutant animals to wandering third instar larvae, thus permitting genetic studies of the histone pre-mRNA processing reaction(Burch et al., 2011; Godfrey et al., 2006; Godfrey et al., 2009; Lanzotti et al., 2002; Sullivan et al., 2001). This feature of the fly genome enabled a genome-wide screen to identify histone pre-mRNA processing factors, which included Symplekin, CPSF-73 and CPSF-100 (Wagner et al., 2007).

How general factors are recruited to the histone pre-mRNA 3' end is still an open question. Recent studies indicate that assembly of the cleavage complex requires an interaction between Lsm11 and another histone processing factor called FLASH (Sabath et al., 2013; Yang et al., 2013). FLASH was identified through a yeast-two hybrid screen for binding partners of the N-terminus of Lsm11 (Yang et al., 2009). Complementary to the HLF experiment (Kolev and Steitz, 2005), recombinant fragments of pre-bound FLASH and Lsm11 immunoprecipitated components of CPSF, CstF and Symplekin from mammalian and *Drosophila* nuclear extracts. Two proteins Symplekin and CstF-64 associated with the FLASH-Lsm11 complex under stringent conditions. Mammalian Symplekin and CstF-64 bind directly (Takagaki and Manley, 2000) and this interaction is likely required for histone pre-mRNA processing but dispensable for cleavage and polyadenylation (Ruepp et al., 2011). These results suggest residues in Symplekin and or CstF-64 could participate in bridging the association between histone specific *trans* factors and the CPSF complex containing the endonuclease.

Figure 1.1

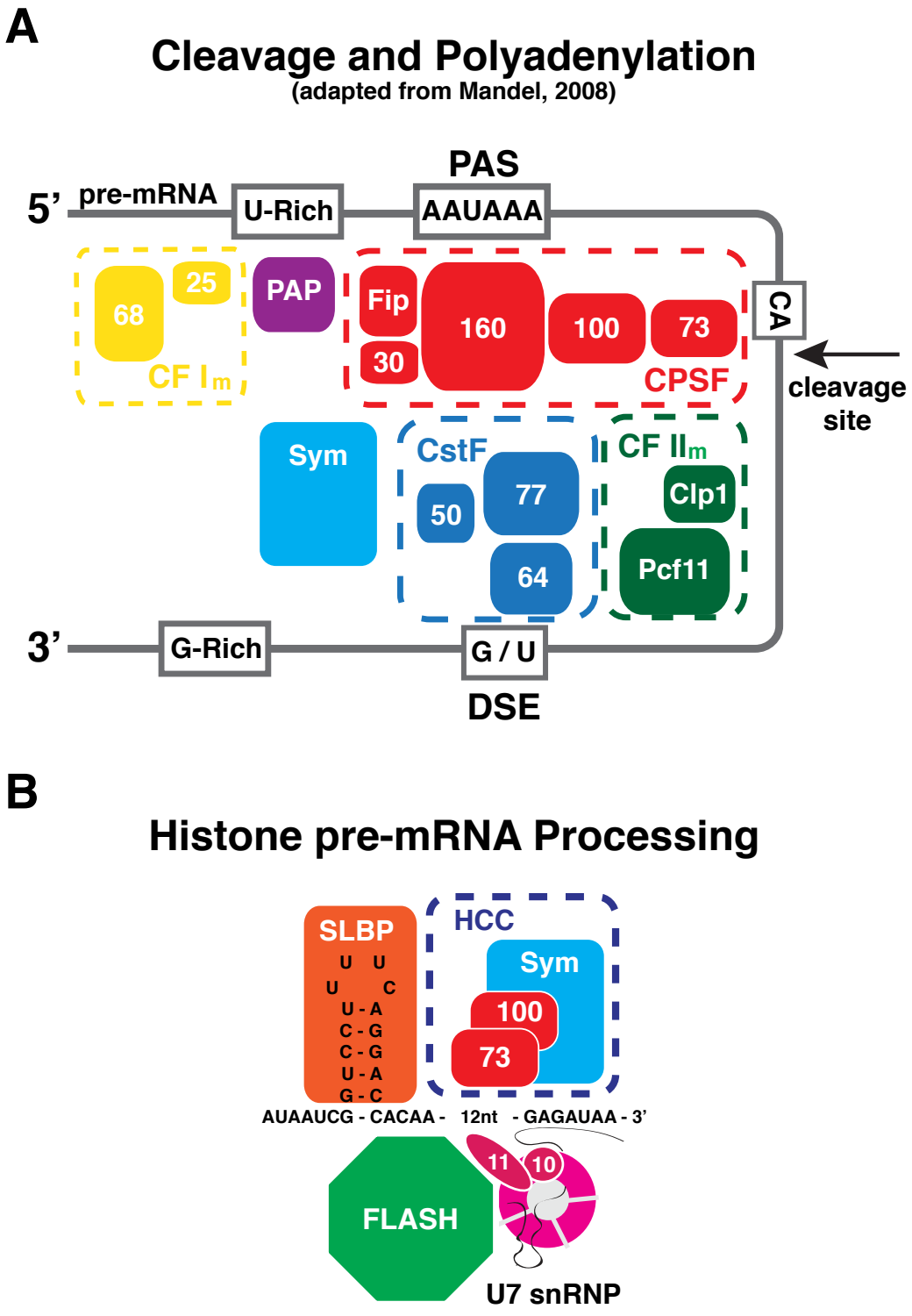
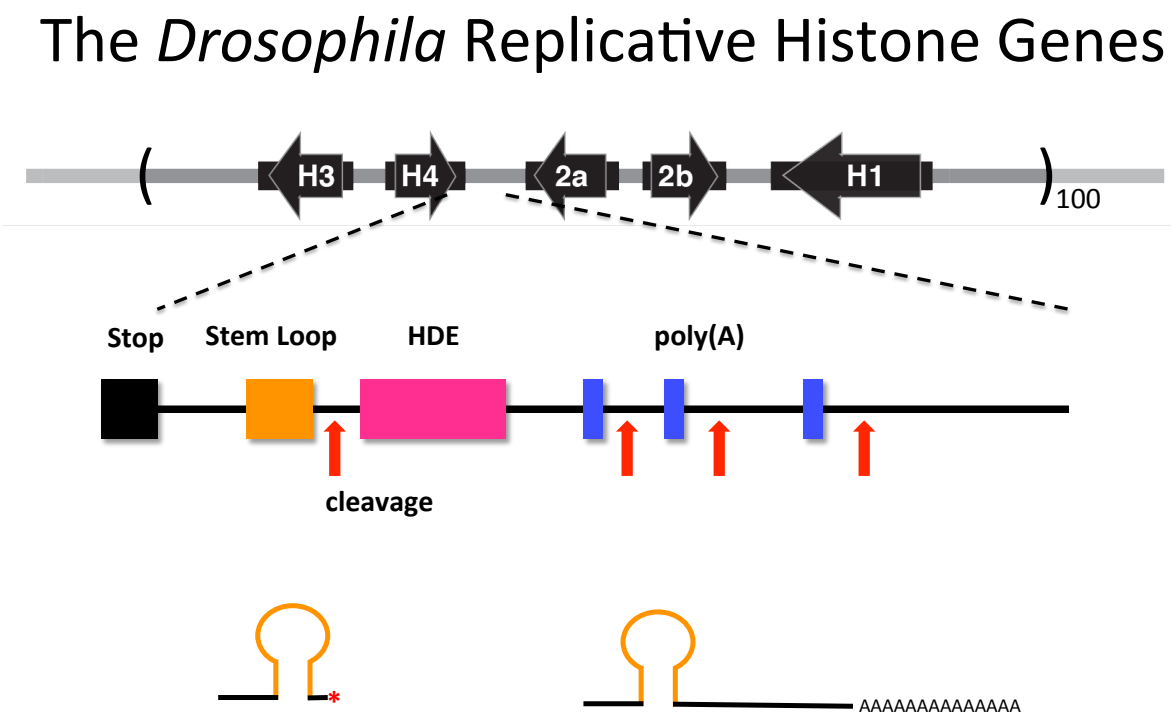

A
Cleavage and Polyadenylation
(adapted from Mandel, 2008)
PAS
5’
pre-mRNA
U-Rich
AAUAAA
68
25
CF Im
PAP
Fip
30
160
100
73
CPSF
CA
cleavage site
Sym
CstF
50
77
64
CF IIm
Clp1
Pcf11
3’
G-Rich
G / U
DSE
B
Histone pre-mRNA Processing
SLBP
U U
U C
U - A
C - G
C - G
U - A
G - C
HCC
Sym
100
73
AUAAUCG - CACAA - 12nt - GAGAUAA - 3’
11
10
FLASH
U7 snRNP

Figure 1.2



Components of the HLB

The HLB was identified and named in 2006 by Joe Gall, based on studies in *Drosophila* (Liu et al., 2006). Factors involved in the biosynthesis of histone mRNA accumulate in the HLB. This was first observed by *in situ* hybridization for U7snRNA in Xenopus oocyte germinal vesicles. U7 snRNA was enriched in the C snurposome, which was later recognized as a body that also contained coilin (Gall, 2000; Wu and Gall, 1993). Coilin positive foci were detected in mammalian cells and co-localized with the U1, U2 and histone gene loci (Frey and Matera, 1995). It was also shown that a subset of CBs in normal human fibroblasts that contain Coilin also contained the protein NPAT, identified as a protein of unknown function that was phosphorylated by cyclin E/cdk2 (Ma et al., 2000; Zhao et al., 1998; Zhao et al., 2000).

Characterization of the CB in *Drosophila* indicated that the CB, marked by dU2, dU85, dSMN and fibrillarin, and the HLB, visualized with components of the U7 snRNP, were distinct nuclear bodies (Liu et al., 2006). Since the initial visualization of the HLB with U7 snRNA, the list of components detected at the histone genes has expanded.

In *Drosophila,* the replicative histone genes are clustered and repeated 100 times on the second chromosome (Lifton et al., 1978) (Fig.1.2). Four HLB components accumulate primarily at the histone genes and also remain associated with the locus in the absence of histone mRNA transcription (outside of S phase). One of these components is the protein encoded by *multi sex combs (mxc).* Mxc is the fly homologue of NPAT (White et al., 2011). In both species, this protein is a Cyclin E-Cdk2 substrate, and Mxc is the major MPM-2 reactive component of the *Drosophila* HLB (White et al., 2007; White et al., 2011; Zhao et al., 2000). MPM-2 is an antibody that recognizes phosphorylated epitopes, and was shown to react with an HLB protein during S phase in a Cyclin E-Cdk2 dependent manner (White et al., 2007). Another factor, Mute, is enriched at the histone genes and is thought to negatively regulate histone mRNA expression (Bulchand et al., 2010). An addition component of the histone pre-mRNA processing machinery, FLASH, is also enriched at the histone locus (Burch et al., 2011; Godfrey et al., 2009; Liu et al., 2006; Yang et al., 2009).

A number of other proteins are both enriched in the HLB, but also detected elsewhere in the nucleoplasm. Coilin, a fundamental component of the CB, is at times, such as in late stage *Drosophila* ovarian nurse cells co-localized with HLB components, but also forms discrete foci that do not contain HLB markers in the nurse cells of earlier stages of oogenesis (Liu et al., 2009). General transcription and elongation factors were detected throughout the nucleoplasm and were also enriched at the histone genes. These include TATA Binding Protein (TBP), (TATA-box-binding protein)-related Factor 2 (TRF2), RNA pol II, and Spt6 (Isogai et al., 2007; White et al., 2011). One of these factors, Spt6, was shown to temporally associate with the histone genes during S phase. This suggests that detection of this elongation factor is a

consequence of high expression of the histone genes. It has not been determined if HLB localization of the other general factors is also confined to S phase.

In addition to the poly (A) factors described above, the genome-wide screen for factors required for processing *in vivo* also identified a new HLB component, MCRS1 (Wagner et al., 2007). Validation of a subset of these newly identified participants in histone 3' end formation involved visualizing localization of tagged, ectopically expressed protein. These results showed enrichment of myc-tagged MCRS1 and Symplekin in the HLB, but not accumulation of CPSF-73 and CPSF-100. Another protein identified in the screen, negative elongation factor (NELF), was independently visualized by ectopic expression of a NELF-A-EYFP fusion protein in HeLa cells and accumulated at the histone locus in mammalian cells (Narita et al., 2007). It is yet to be determined if these multifunctional proteins localize temporally to the histone locus like the general transcription factors, or are constitutive components of the HLB.

Depletion experiments performed in S2 cultured cells identified FLASH and Mxc as critical components of the HLB (White et al., 2011). Treatment of cells with dsRNA targeted to the *mxc* coding region prevented accumulation of FLASH, Mute and Lsm11 in foci and the cells no longer contained MPM-2 reactive foci. FLASH depletion resulted in a similar loss of foci with Mute, Lsm11 and MPM-2 as well as reduced enrichment of Mxc. Depletion of Mute did not affect FLASH, Lsm11 or Mxc accumulation in the HLB, and similar results were observed after reduction of Lsm11. These experiments suggest that the HLB forms in stages with Mute and U7 snRNP requiring FLASH and Mxc for enrichment at the histone locus.

Our understanding of HLB formation, as well as the availability of markers, provides a framework for detailed characterization of HLB components such as Symplekin and FLASH. For example, the C-terminus of FLASH is required for ectopic FLASH accumulation in the HLB of cultured cells (Burch et al., 2011), therefore we can test whether FLASH localization to the HLB is essential for *Drosophila* development. Because the HLB is spatially associated with the histone locus, we can also isolate the effect of concentrating FLASH in the HLB on the reaction

of histone mRNA processing. We hypothesize that the HLB plays a critical, measurable role in histone mRNA biosynthesis.

Dissertation Goals

Using the *Drosophila* replicative-histone genes as a model, my studies investigate how nuclear organization contributes to mRNA 3' end formation.

In Chapter 2, together with Harmony Salzler, I investigate HLB assembly and identify a sequence within the histone locus that is essential for Drosophila HLB assembly. We show that FLASH and Mxc nucleate the HLB and that transcription from the H3-H4 promoter is required for maturation of this nuclear body. In the course of these experiments, we also found that histone pre-mRNA processing factors are recruited to the H3-H4 promoter in the absence of the *cis* elements required for histone pre-mRNA processing. This work implies that the HLB assembles to promote histone gene expression, possibly by coordinating transcription and 3' end formation, rather than as a consequence of histone gene expression.

I next asked if localizing a factor to the HLB affected the rate of histone pre-mRNA processing. In Chapter 3, I experimentally evaluate the consequences of preventing FLASH accumulation in the HLB. In the course of these experiments, I characterized the *in vivo* role of FLASH in histone pre-mRNA processing. This structure function analysis uncovered a role for the HLB in histone 3' end formation.

Finally, in Chapter 4, I characterize *Drosophlila symplekin*. Symplekin participates in all three mechanisms of 3' end formation as well as cytoplasmic polyadenylation. Understanding the mechanism of Symplekin incorporation into these distinct complexes at the molecular level will provide insight into how the cell assembles the machines that carry out mRNA processing events. To understand the role of Symplekin in the animal, I generated an allelic series of mutants that alter the levels of Symplekin protein. The ultimate goal is to use these mutants to identify regions of Symplekin (if any) used solely for histone pre-mRNA processing. This series

of experiments highlight the many roles of Symplekin in gene expression and provide tools for subsequent analysis of this critical component of mRNA regulation.

CHAPTER 2: A SEQUENCE IN THE *DROSOPHILA H3-H4* PROMOTER TRIGGERS HISTONE LOCUS BODY ASSEMBLY AND BIOSYNTHESIS OF REPLICATION-COUPLED HISTONE MRNAS

Preface

This work was previously published as a research article. The study was done in collaboration with a former graduate student Harmony Salzlar, who initiated the project. Dr. Salzlar designed and built all transgenic constructs. With the exception of the salivary glands that I stained in Figure 2.5B and 2.6, she performed all of the cell biology experiments in this tissue. Dr. Salzler quantified co-localization of our panel of HLB markers for each construct. Pamela Malek performed the quantitative RT-PCR analysis and Prem Fort performed the western blot presented in Figure 2.7G. With the help of undergraduate student, Anna Orlando, I performed all of the other RNA analysis and also characterized the early embryonic phenotype in the histone deletion embryos. Stephen McDaniel contributed the embryo image in Figure 2.7F. Dr. Salzler wrote the initial draft of the paper in her dissertation, which was then edited and expanded by Dr. William Marzluff, Dr. Robert Duronio and myself.

Salzler, H.R*., Tatomer, D.C*., Malek, P.Y., McDaniel, S.L., Orlando, A.N., Marzluff, W.F. and Duronio, R.J (2013) A sequence in the Drosophila H3-H4 promoter triggers Histone Locus Body assembly and biosynthesis of replication-coupled histone mRNAs. Dev. Cell 24(6):623-34.
(*) These authors contributed equally

INTRODUCTION

In the last decade, compartmentalization of nuclear processes has emerged as an important organizing principle of the genome. Nuclei contain a host of distinct compartments or “nuclear bodies” such as nucleoli, speckles, paraspeckles, Cajal bodies, PML bodies, and histone locus

bodies (HLBs) where factors involved in processes such as transcription, RNA processing and maturation, and DNA replication and repair are concentrated (Carmo-Fonseca and Rino, 2011; Handwerger and Gall, 2006; Matera et al., 2009; Misteli, 2007) . Despite a role for NBs in a wide range of biological processes, a complete understanding of the relationship between NB formation and the associated biochemical reactions (e.g. transcription and pre-mRNA splicing/processing) is lacking.

NBs are thought to enhance the efficiency of reactions by concentrating reaction components (Matera et al., 2009; Misteli, 2007) . While there is some evidence for this idea (Chen et al., 2010; Strzelecka et al., 2010; Wagner et al., 2007) , the importance of the contribution that NBs make to their associated processes is not always clear. The Cajal body, a NB involved in snRNP biogenesis, provides a good example. Mutation of the gene encoding Coilin, a critical assembly component of Cajal bodies, is lethal in zebrafish due to failure to form sufficient snRNPs (Strzelecka et al., 2010), but not in flies (Liu et al., 2009). Moreover, while *Coilin* mutant mice are not fully viable or fertile (Tucker et al., 2001; Walker et al., 2009), *coilin* mutant flies, which lack detectable Cajal bodies, are fertile and correctly perform several snRNA modifications that are specified by scaRNAs normally localized to Cajal bodies (Deryusheva and Gall, 2009).

Determining how NBs form is critical for understanding how NBs affect their associated biochemical processes. Current evidence suggests that NBs form by a process of "self-organization" in which individual factors encounter other NB components through random molecular collisions, and remain in proximity due to binding affinities (Handwerger and Gall, 2006; Misteli, 2001; Nizami et al., 2010) . The high affinity between factors associated with specific processes results in the formation of microscopically visible structures. Two contrasting models of self-organization have been proposed (Matera et al., 2009; Misteli, 2007) . Tethering experiments that artificially localize individual nuclear body components support a stochastic self-organization model, where nuclear body components can assemble in any order (Kaiser et

al., 2008; Shevtsov and Dundr, 2011). In contrast, genetic evidence supports a hierarchical model, which posits that NBs assemble in a particular order with assembly of some components predicated on prior assembly of other components (Rajendra et al., 2011; White et al., 2011). Recently, a hybrid model has emerged from studies of the HLB and paraspeckles suggesting that the stochastic and hierarchical models of nuclear body formation are not mutually exclusive (Dundr, 2011; Mao et al., 2011; White et al., 2011).

The HLB is an excellent model for investigating both the mechanism and function of NBs. HLBs assemble at replication-coupled histone genes in animal cells and contain factors associated with the transcription and processing of histone mRNA (Bongiorno-Borbone et al., 2008; Bongiorno-Borbone et al., 2010; Frey and Matera, 1995; Ghule et al., 2008; Liu et al., 2006; Nizami et al., 2010; White et al., 2007; White et al., 2011) . HLBs contain factors such as U7 snRNP and FLASH that are necessary for an endonucleolytic cleavage of histone pre-mRNA resulting in a unique 3' stem-loop structure that mediates all aspects of histone mRNA regulation, rather than a poly(A) tail (Marzluff et al., 2008). In fact, HLBs were originally defined through studies of the localization of the U7 snRNP specific proteins, Lsm10 and Lsm11, as well as U7 snRNA (Liu et al., 2006). HLBs also contain the protein NPAT, a substrate of Cyclin E/Cdk2 that is concentrated at the two clusters of human histone genes (Ma et al., 2000; Zhao et al., 2000) . NPAT is essential for entry into S-phase and for expression of histone mRNA, although the precise molecular basis of NPAT action is not understood (Ma et al., 2000; Miele et al., 2005; Wei et al., 2003; Ye et al., 2003; Zhao et al., 2000) . There is no evidence that NPAT directly binds DNA, and more likely it acts as a cofactor for histone gene transcription and possibly coordinates the multiple steps in histone mRNA biosynthesis.

In metazoans, the accumulation of replication-dependent histone mRNAs is confined to S-phase when histone proteins are required for chromatin assembly (Marzluff and Duronio, 2002; Marzluff et al., 2008). The tight regulation of histone accumulation during the cell cycle is essential for genetic stability (Gunjan and Verreault, 2003; Marzluff, 2010; Meeks-Wagner and

Hartwell, 1986) , suggesting that the S phase role of the HLB in histone biosynthesis is likely to impact a wide range of genomic functions. The HLB is also present during G1 and G2 phase, when histone mRNAs are not actively synthesized (White et al., 2011; White et al., 2007), indicating that HLB assembly and/or maintenance is not strictly dependent on active transcription and/or mRNA processing. Histone gene expression is activated in cycle 11 of *Drosophila* embryogenesis, the same time as the HLB forms. A subset of HLB components, FLASH and Mxc, the Drosophila ortholog of NPAT, accumulates at the histone locus before the onset of histone gene expression in the early *Drosophila* embryo, and we term this complex a "proto-HLB" (White et al., 2011).

Because HLBs assemble only at histone genes, we hypothesized that a sequence element(s) within or associated with the histone locus would drive HLB assembly. Here we present the surprising result that despite the fact that all five replication-dependent histone genes are coordinately transcribed and processed, only a single element in the *Drosophila* histone gene locus is capable of nucleating the HLB. We demonstrate that formation of the *Drosophila* HLB depends on a sequence in the 300 nt histone *H3-H4* bidirectional promoter, and that this sequence is essential for expression of other histone genes in the cluster. A proto-HLB assembles on the minimal sequence in the absence of transcription, but transcription driven by this sequence is necessary for formation of a complete HLB. In addition, we show that proto-HLBs form transiently even in the absence of histone genes, indicating that some HLB components have self-organization properties. Together these results support a model whereby transcription-dependent ordered assembly and stochastic self-organization of components both contribute to HLB assembly during development.

MATERIALS AND METHODS

Drosophila Strains

DNA sequences from the histone locus were engineered for insertion into pattB (Gift from the Basler lab) with restriction enzymes Kpn1 and XbaI. Table 2.1 summarizes the boundaries for each insert and also any internal features represented in the different constructs. Histone locus sequences were integrated into either 86Fb (BDSC 23648) or 102D (BDSC 24488) by ϕC31 mediated recombination (BestGene, Inc).

Histone Expression Analysis

Total RNA was extracted from tissues with Trizol (Invitrogen). For the RT-PCR, Total RNA was extracted with Trizol (Invitrogen) from wandering 3rd instar salivary glands. Following DNase treatment (Fermentas), either 250 ng of total RNA was reverse transcribed with random hexamers for qPCR analysis or 1 µg of total RNA was reverse transcribed with a combination of oligodT (250ng) and random hexamers (50ng) to synthesize cDNA with RevertAid reverse transcriptase (Fermentas). cDNA was used for SYBR green (Fermentas) mediated qPCR quantification with an Applied Biosystems 7900HT PCR machine. qPCR results are presented as an average of at least 3 biological replicates; error bars represent SEM.. Histone-Vector hybrid transcripts and genomic DNA control template were amplified and resolved on a 2% agarose gel. Primers are provided in Table 2.2. To analyze histone RNA expression by S1 nuclease protection assay, Total RNA was extracted with Trizol (Invitrogen) from wandering 3rd instar salivary glands, whole 3rd instar larvae, ovaries and 3-6h embryo collections. We used RNA from glands from 20 larvae, or the indicated amount for each assay. The probes were created by digesting plasmids containing the FLAG-tagged histone *H2a* gene with XhoI or the FLAG-tagged *H4* gene with NcoI. After removing the 5' phosphate with Calf intestinal phosphatase, the fragments were labeled at the 5' end with γ-32P-ATP and T4 polynucleotide kinase. The probe was released from the plasmid by digestion with SpeI (H2a) or SacI (H4) and

gel purified. The probe was hybridized to either total salivary gland RNA or control yeast tRNA followed by digestion with S1 nuclease as described and were resolved on a 6% polyacrylamide-7M urea gel and detected by autoradiography (Lanzotti et al., 2002). The ratio of expression of the transgene to endogenous histone mRNA was quantified with ImageQuant.

Immunofluorescence

Polytene squashes were prepared as described (Paro, 2008) . Staining conditions and antibodies used are summarized in Tables 2.3 and 2.4. All samples were blocked with Enhancer Reagent (Invitrogen) at room temperature for 30 min and all samples were stained with 4',6-diamidino-2-phenylindole (DAPI) (1:1000) for 1 min to visualize DNA. Images were obtained on a Zeiss 510 confocal microscope. Ectopic HLBs were quantified by determining the percentage of ectopic foci in nuclei of 15-20 1μM sections of a 150 μM^2 area in the posterior of 7-10 salivary glands. Graphs represent the mean and SEM for each indicated transgene.

Analysis of Histone Deletion Embryos

Post-blastoderm histone deletion embryos were collected from *Df(2R)Ds6*/CyO,twi-GFP parents and identified by lack of GFP expression. Syncytial stage histone deletion embryos were genotyped by the number of FLASH foci per nucleus. Briefly, nuclei of 20 WT embryos ranging from cycles 11-14 were counted to (1) ensure that 3+ foci are never present and (2) determine that WT embryos have at least 10% of nuclei containing 2 foci. Images from 59 embryos of the *Df(2L)Ds6*/CyO, twi-GFP collection were sorted into three classes based on foci present in the nuclei of each image. Images with only 1 or 2 foci in each nucleus were identified as control siblings. This group was further categorized by identifying embryos containing 90% or greater nuclei with a single foci as heterozygous for the histone deletion, the remaining being WT. Histone deletion embryos were identified by the presence of nuclei containing 3+ foci. Significance was assessed by chi-squared analysis. HLB formation was assessed for deletion and sibling control embryos by examining pairs of HLB markers. For each genotype, FLASH foci were identified and then scored for overlap with Mxc or Mute. A nucleus was only considered

positive if all FLASH foci co localized with the other marker. Nuclei contained 1, 2 or 3+ foci and the graph presents the percent of total overlap out of the total number of nuclei counted for each of these classes within a genotype for either FLASH/Mxc or FLASH/Mute. The results are presented as a box (25th -75th quartiles) and whiskers (10-90th percentile) plot. (GraphPad Software, La Jolla California, USA)

Table 2.1. Boundaries for Each Histone Locus Construct

Construct	Boundaries
HL-FL	F: ggtacctaatgcatatgtggcgaggccatgtgttaactgaagaatgtgt R: ccatgtgttaactgaagaatgtgttctagatgtcgaagtttgcttgaagtg
HLT-FL	See HL-FL
H3-H4	F: ggtaccaccaataaaattaatact R: **t**ctagaaaagttataaatagtcggcaac
H2a-H2b	F: ggtacctcatattcgatgattggt R: tctagattacaacaaattgccaagcta
H1	F: ggtaccgccgactatttataacttta R: tctagagttttattgttgctgcgaac
H3-H4^{PS}	See H3-H4
H2a-H2b^{PS}	See H2a-H2b
H3-H4P	H3 5'UTR: *attgtgttttcaaacgtgaagtagtgaacgtgaactttagtgaaacccaaatcgg* H4 5'UTR: *ttcactgttctatactattatacacgcacagcacgaaagtcactaaagaactaatt*
CORE	F: ggtaccgtttcatgtcatgaattac R: tctagaaaagttataaatagtcggcaac
CORE^{TATA}	See CORE
2xH3-H4	F: gtttaaaccaccaataaaattaaatact R: ggtaacaaagttataaatagtcg
Internal Feature	Hybrid Sequence

FLAG H4	cctcctttaccacgaccagtggccttgtcatcgtcatccttgtaatcCAT*ttttcactgttctatac*
FLAG H2a	*taagtgaaataaacgcaaagcaaa*ATGgattacaaggatgacgatgacaaggcctctggacgtgg aaaagg
H3-H4^{PS}	H2a 5'UTR/Flag-H4: *actaagtgaaataaacgcaaagcacc*ATGgattacaaggatgacgatgac H2b 5'UTR/H3: cctttccactagttttcggagcCAT*ggttcacgttacttatattttca*
H2a-H2b^{PS}	H4 5'UTR/Flag-H2a: cttgtcatcgtcatccttgtaatcCAT*ggttcactgttctatactattatacac* H3 5'UTR/H2b: *ttagtgaaacccaaatcggcc*ATGgctcgtaccaagcaaa
H4 TATA D	aagaactaatttcaacgtttctgtgtgcccc**gggccc**taggtaaaacgacaaaaacccgagagagtac
H3 TATA D	ctctctctctttcaccgtccacgattgc**ggccgc**agtaggtagcaaatgctctgatcgttta
HA H4	*agtgaaacccaaatcggag*ATGtacccatacgatgttccagattacgctgctcgtaccaagcaaact

The following features are annotated in the sequences: engineered restriction sites are underlined; tags are grey; UTR sequence is italicized; TATA mutations are bold.

Table 2.2. Primers used in RT-PCR

Primer Name	Sequence
H2a R	5'- gcagctaggtaaactggag-3'
H4 R	5'- gtaggtcacggcatcacg-3'
FLAG F	5'-gattacaaggatgacgatgacaag-3'
Actin F	5'- ggtcacgataccgtgctc-3'
Actin R	5'- aacggctctggcatgtg-3'
H3 5'UTR	5'-caaacgtgaagtagtgaacgtga-3'
H4 5'UTR	5'-tttagtgactttcgtgctgtgc-3'
H3 flanking vector	5'-gcccccaactgagagaact-3'
H4 flanking vector	5'-aaaatgccttggatttcactgg-3'

Table 2.3. Tissue Preparation for Antibody Staining

Tissue	Fix	Permeabilization
polytene squash	3.7% Formaldehyde, 2’	
whole salivary gland (1)	3.7% formaldehyde, 15’	1.0% TritonX, 15’
whole salivary gland (2)	7% formaldehyde 20’	0.2% Tween 15’
embryo	1:1 7% formaldehyde: heptane, 20’	MeOH

Table 2.4. Antibody Concentrations

Primary Antibody	Raised In	Source	Concentration	Incubation
α-FLASH	rabbit	(Yang et al., 2009)	1:2000	4 C, overnight
α-Mxc	guinea pig	(White et al., 2011)	1:2000	4 C, overnight
α-Mute	guinea pig	(Bulchand et al., 2010)	1:2000	4 C, overnight
α-MPM-2	mouse	Millipore	1:2000	4 C, overnight
C1A9 α-HPI	mouse	Developmental Studies Hybridoma Bank	1:1000	4 C, overnight
α-GFP	chicken	Upstate	1:1000	4 C, overnight
α-Lsm11	rabbit	(Liu et al., 2006)	1:1000	room temperature, 2h
Secondary Antibody	**Recognizes**	**Source**	**Concentration**	**Incubation**
Alexa-488	rabbit-IgG	Invitrogen	1:2000	room temperature, 2h
Cy5	guinea pig IgG	Jackson	1:1000	room temperature, 2h
Cy3	mouse IgG	Jackson	1:1000	room temperature, 2h

RESULTS

An HLB Can Assemble at a Single, Active Histone Gene Repeat

The *Drosophila* replication-coupled histone genes are present in a single locus on chromosome 2L as a tandem 5 kB repeat present in about 100 copies. Each repeat unit contains one copy of each of the five histone genes (Fig. 2.1A). The *H2a-H2b* and *H3-H4* gene pairs are divergently transcribed, while the *H1* gene is located about 1.5 kB 3' of the H3 gene, and ends about 300 nts before the 3' end of the *H2b* gene (Fig. 2.1A). To determine whether sequences within a single repeat unit were sufficient to direct the formation of an HLB, we used a construct containing 1.2 copies of the repeat unit such that all contiguous sequences 500 nts long were represented in the construct (*H*istone *L*ocus-*F*ull *L*ength, or *HL-FL*; Fig. 2.1A) and generated transgenes at specific loci in the *Drosophila* genome by ΦC31-mediated integration (Bateman and Wu, 2007; Bischof et al., 2007) . To test whether the chromatin environment around the ectopic histone genes can influence expression, *HL-FL* was inserted into two specific sites: a euchromatic site on chromosome 3 (86Fb) and a heterochromatic site on chromosome 4 (102D).

We visualized HLBs using antibodies to four components of the HLB: Multi sex combs (Mxc), the *Drosophila* orthologue of NPAT that we recently identified (White et al., 2011); FLASH, a histone pre-mRNA processing factor (Burch et al., 2011a; Yang et al., 2009); Mute, an essential protein of unknown function homologous to YY1 associated protein (Bulchand et al., 2010); and Lsm11, a component of U7 snRNP (Azzouz and Schumperli, 2003; Pillai et al., 2003). For some experiments we utilized a *Drosophila* line expressing V5 tagged Lsm11, which rescues an *Lsm11* null mutant, and visualized U7 snRNP in the HLB using an anti-V5 antibody (Godfrey et al., 2009). This panel of reagents includes factors that are first detected in the HLB before (Mxc and FLASH) and after (Mute and Lsm11) the onset of zygotic histone transcription (White et al., 2011; White et al., 2007).

We analyzed chromosome spreads from 3rd instar larval salivary gland cells. In these polyploid cells the genome reaches more than 1000C and individual chromatids line up in register, resulting in polytene chromosomes that provide high resolution for cytological experiments (Fig. 2.1B). Using antibodies to Lsm11, Mute, and FLASH, we observed HLB assembly at the ectopic *HL-FL* locus at 86Fb on chromosome 3, as well as at the endogenous histone locus at 39D-E on chromosome 2 (Fig. 2.1B). In contrast, when the repeat was located at 102D on chromosome 4, HLB assembly was not observed (Fig. 2.1C), although its genomic presence was confirmed by PCR (not shown). We conclude that one copy of the histone repeat is sufficient to assemble an HLB at a euchromatic but not a heterochromatic site.

To assess whether ectopic genes were expressed, identical transgenic lines were generated containing 5' FLAG tags on the *H2a* and *H4* genes (*H*istone *L*ocus *T*agged-*F*ull *L*ength*; HLT-FL*, Fig. 2.1A). We used a FLAG specific qRT-PCR primer to determine the expression of the ectopic *H2a* and *H4* mRNAs relative to actin mRNA, and normalized these results to the *HLT-FL* insertion at 86Fb on chromosome 3. We detected expression of both genes from the ectopic repeat located at 86Fb but not from the transgene inserted at 102D on chromosome 4 (Fig. 2.1D). Thus, sequences present in the histone repeat are sufficient to direct HLB assembly and histone gene expression, and no sequences flanking the histone locus are necessary. However, other factors such as local chromatin structure influence HLB assembly and histone gene expression.

Figure 2.1

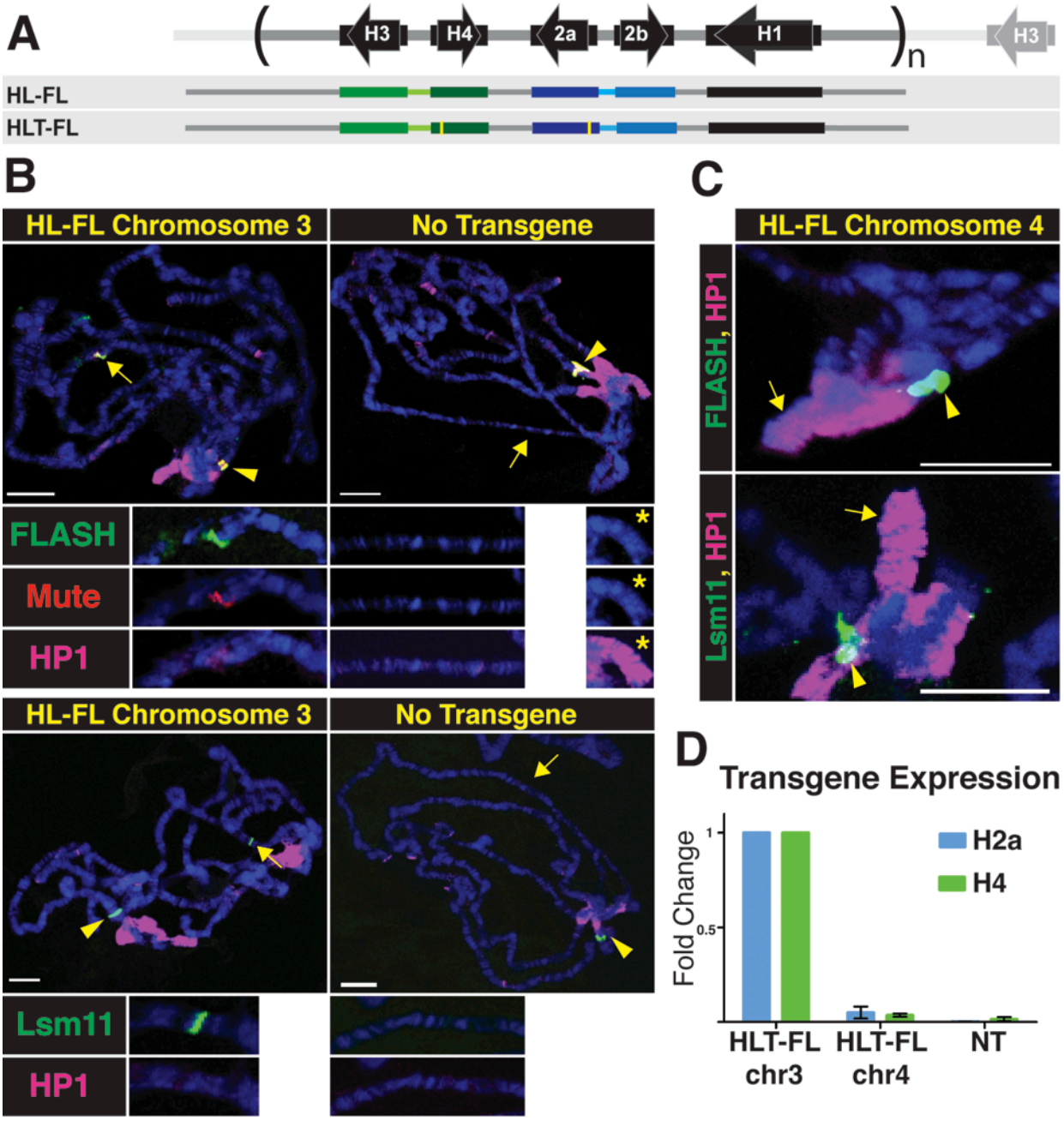

A
H3
H4
2a
2b
H1
n
H3
HL-FL
HLT-FL
B
HL-FL Chromosome 3
No Transgene
FLASH
Mute
HP1
HL-FL Chromosome 3
No Transgene
Lsm11
HP1
C
HL-FL Chromosome 4
FLASH, HP1
Lsm11, HP1
D
Transgene Expression
Fold Change
1
0.5
H2a
H4
HLT-FL chr3
HLT-FL chr4
NT

Figure. 2.1. **An HLB forms at an ectopic locus containing one histone gene repeat unit.**
A. Diagram of the histone repeat (chromosome 2). The 5.1 kB histone repeat unit is indicated by parentheses. A fragment containing 1.2 repeat units (*HL-FL*) was cloned and inserted into the *Drosophila* genome at either site 86Fb on chromosome 3, or at site 102D on chromosome 4. The yellow bars in the HLT-FL construct represent N-terminal FLAG tags in *H2a* and *H4*.
B. Chromosome squashes from salivary glands of third instar larvae containing the *HL-FL* at 86Fb (left; n=15) or no transgene (right; n=7) stained with Mute (red), FLASH (green) and HP1 (pink, top panel) or Lsm11 (green) and HP1 (pink, bottom panel). The insets show a higher magnification of the 86Fb chromosome region except the panel with * which shows chromosome 4 (102D). The arrowhead indicates the endogenous HLB and the arrow indicates chromosomal position 86Fb. Bars = 10 mm.
C. Chromosome 4 from salivary glands of 3rd instar larvae containing the *HL-FL* transgene at position 102D (arrow; n=8) stained with HP1 (pink) and either FLASH (green, top) or Lsm11 (green, bottom). The endogenous histone locus near the chromocenter is indicated by the arrowhead.
D. RT-PCR analysis of *H2a* and *H4* expression from *HLT-FL* located at 86Fb (chr3) and 102D (chr4) compared to no transgene (NT). Histone gene expression was normalized relative to the expression of actin mRNA. Error bars represent SEM.

Histone Gene Expression Correlates with HLB Assembly

The formation of HLBs at ectopic sites that expressed histone genes provided us with an opportunity to define sequences within the histone gene repeat that direct HLB assembly and histone gene expression, and to determine how these two processes are functionally related. We made transgenic flies with constructs inserted at 86Fb that contain only the *H3-H4* gene pair, the *H2a-H2b* gene pair, or the histone *H1* gene plus the long intergenic region between it and the 3' end of the *H3* gene (Fig. 2.2A). We assessed HLB formation by quantifying the presence of ectopic HLBs in intact salivary gland nuclei (Wagner et al., 2007) using multiple pairs of HLB markers (Fig. 2.2B). Ectopic HLBs were defined by co-localization of two or three HLB components in a focus (arrows, Fig. 2.2B), in addition to the endogenous HLB (arrowhead, Fig. 2.2B). For each experiment, ectopic HLBs were quantified using 1mm sections of a 150mm^2 area through the posterior portion of the salivary gland, and the data are presented as percent mean ectopic focus formation of 7-10 individuals (>100 cells for each construct) (Fig. 2C).

HLT-FL supported HLB formation in nearly 100% of nuclei with all marker pairs (Fig. 2.2B,C). The *H3-H4* construct formed ectopic HLBs in 40% to 95% of the cells depending on the marker pair examined (Fig. 2B,C). In contrast, less than 20% of cells containing either the *H2a-H2b* gene pair or the *H1* gene formed an ectopic HLB, similar to non-transgenic controls (Fig. 2.2C). All four markers were present in the HLBs that formed on the *H3-H4* gene pair, which was similar in size to the *H2a-H2b* or the *H1* transgenes. These data indicate that a specific sequence(s) within the *H3-H4* gene pair directs HLB assembly.

To test whether HLB assembly was important for histone gene expression, we developed an S1 nuclease protection assay using a 5' end-labeled probe (P) containing the FLAG-tagged *H2a* or *H4* genes (Fig. 2.2D). These probes detect both the endogenous (E) H2a and H4 histone mRNAs and the longer mRNAs produced by the transgenes (T) (Fig. 2.2D). This assay is quantitative and allowed us to determine the relative level of ectopic versus endogenous histone mRNA accumulation by comparing signal intensities between the S1 nuclease protected fragments in each sample. To validate the assay, we analyzed varying amounts of total RNA isolated from dissected salivary glands, 3-6 hr old embryos (diploid cells), and ovaries and whole 3rd instar larvae (mixed diploid and polyploid cells). The ectopic *H2a* and *H4* genes in *HLT-FL* were expressed at ~7% the level of the endogenous genes in salivary glands, compared to 2.5% in ovaries and 1% in embryos and whole larvae (Fig. 2.2D). While the basis for these differences is not known, they may be due to under-replication of the endogenous histone genes relative to the rest of the salivary gland genome (Hammond and Laird, 1985). The relatively high expression of the ectopic histone mRNA as measured by S1 nuclease protection assay made salivary gland RNA the best source to carry out subsequent experiments.

We determined the expression of the FLAG tagged *H2a* and *H4* genes in the *H3-H4* and *H2a-H2b* transgenic lines relative to the expression of the corresponding gene in the *HLT-FL* full repeat unit. The *H4* gene in the *H3-H4* line was expressed at 35% of the level of the *H4* gene in

HLT-FL (Fig. 2.2E lanes 4-6). In contrast, the *H2a* gene in the *H2a-H2b* gene pair was expressed at <5% of the level in the full-length transgene (Fig. 2.2E lanes 1-3). These S1 nuclease assays were consistent with data obtained by qRT-PCR (Fig. 2.2F). These data correlate well with HLB assembly, which occurred with the *H3-H4* gene pair but not *H2a-H2b* (Fig. 2.2B,C), suggesting that HLB assembly contributes to histone gene expression.

Figure 2.2

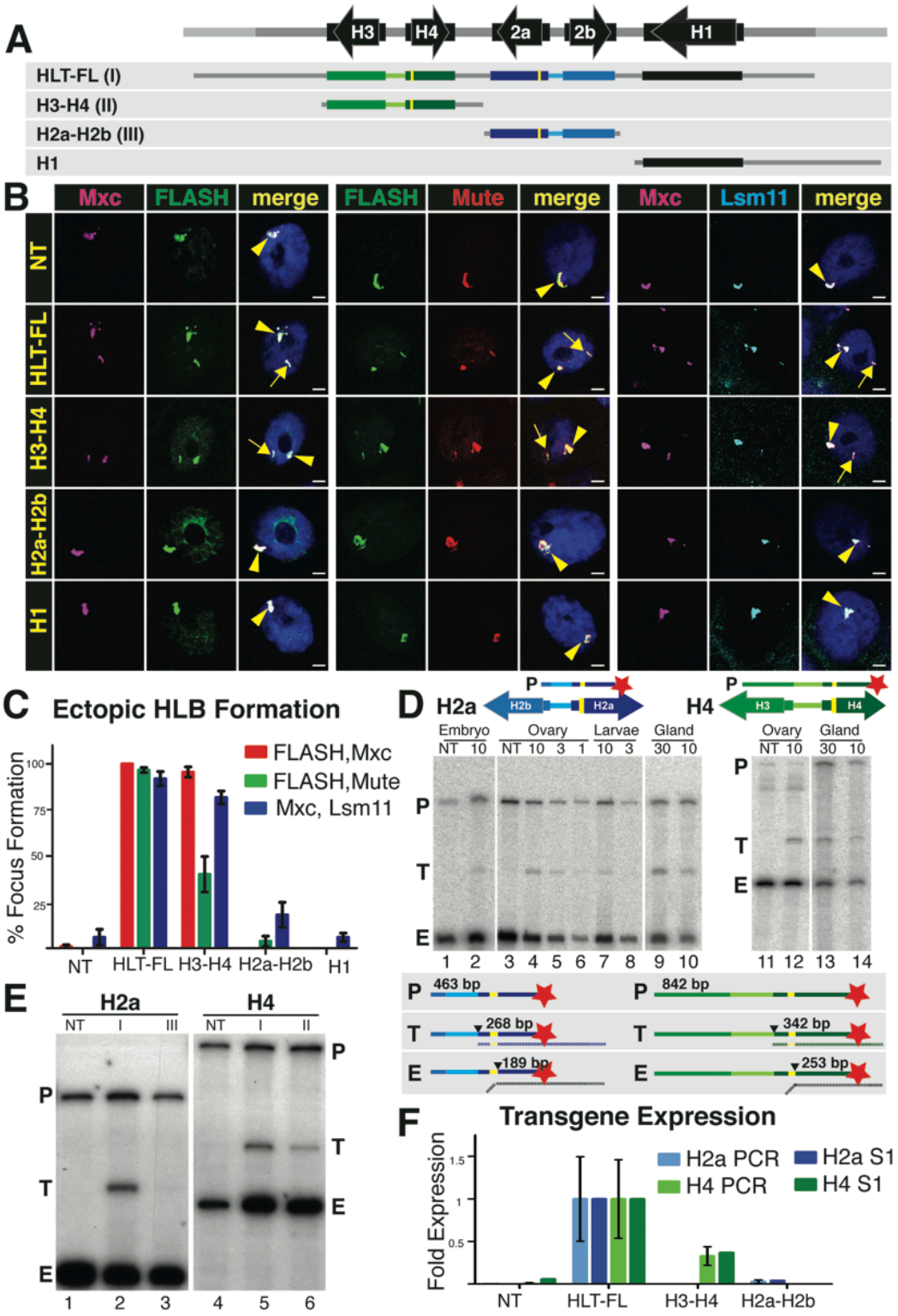

A
H3
H4
2a
2b
H1
HLT-FL (I)
H3-H4 (II)
H2a-H2b (III)
H1
B
Mxc
FLASH
merge
FLASH
Mute
merge
Mxc
Lsm11
merge
NT
HLT-FL
H3-H4
H2a-H2b
H1
C
Ectopic HLB Formation
% Focus Formation
FLASH,Mxc
FLASH,Mute
Mxc, Lsm11
NT
HLT-FL
H3-H4
H2a-H2b
H1
D
H2a
H4
Embryo
Ovary
Larvae
Gland
P
T
E
463 bp
268 bp
189 bp
842 bp
342 bp
253 bp
E
H2a
H4
NT
I
III
II
F
Transgene Expression
Fold Expression
H2a PCR
H4 PCR
H2a S1
H4 S1
NT
HLT-FL
H3-H4
H2a-H2b

Fig 2.2. **The *H3-H4* genes assemble an ectopic HLB.**
A. Diagram of the four constructs inserted into chromosomal location 86Fb. The yellow bars represent N-terminal FLAG tags in *H2a* and *H4.*
B. HLB assembly for each construct (indicated at left) was assessed by confocal microscopy of intact salivary gland nuclei stained with Mxc and FLASH (left), Mute and FLASH (center) or Mxc and Lsm 11 (right). The endogenous HLB contained Mxc (pink), FLASH (green), Mute (red) and Lsm11 (blue) in all samples (arrowhead). Note the assembly of an ectopic HLB with each marker for nuclei containing the *HLT-FL* or the *H3-H4* transgenes (arrow). Scale bar indicates 10 mm.
C. Quantification of ectopic HLB formation. Error bars depict SEM.
D. Expression of histone mRNA from the *HLT-FL* transgene was assessed throughout development by 5' S1 nuclease protection assay using a ^{32}P end-labeled (red star) probe (P) complementary to either the H2a or H4 endogenous and ectopic transcripts. Numbers above the gel indicate the amount of RNA (mg, except glands which were total number of glands) in each reaction. The S1 nuclease assay is diagrammed below the gel. Numbers indicate the length in nt of the probe (P), ectopic (T) and endogenous (E) protected H2a or H4 transcripts. The black triangle indicates nuclease cleavage of the probe at the point where the RNA (vertical dashed line below probe) is not complementary.
E. Expression of histone mRNA was assessed in salivary glands by 5' S1 nuclease protection assay. Roman numerals indicate the transgene inserted in each sample (depicted in A). Note that ectopic histone expression (T) was detected from constructs carrying *HLT-FL* and *H3-H4*.
F. Relative histone mRNA expression was measured for H2a (light blue columns) and H4 (light green columns) by qRT-PCR and quantification of the S1 protection assay (dark columns). Both assays are presented as fold expression compared to *HLT-FL*, which was set at 1.0. Error bars depict SEM.

The H3-H4 Promoter is Necessary and Sufficient for HLB Formation

Both the *H3-H4* and *H2a-H2b* constructs contain a ~300 nt bidirectional promoter, the two coding regions, and the mRNA 3' end processing signals. To determine the sequences responsible for HLB formation, we swapped the intergenic promoter region (i.e. from start codon to start codon) of the *H2a-H2b* genes with the corresponding region of the *H3-H4* genes (Fig. 2.3A) and generated transgenic insertions at 86Fb. We kept the FLAG tag on the N-terminus of the *H2a* and *H4* genes to allow us to assess expression of the ectopic genes. Strikingly, the *H2a-H2b* gene pair containing the *H3-H4* promoter (*H2a-H2b*^{PS}) now formed an HLB, while the *H3-H4* gene with the *H2a-H2b* promoter (*H3-H4*^{PS}) did not (Fig. 2.3B). Moreover, the *H2a-H2b*^{PS} transgene formed an HLB with the same efficiency as the *H3-H4* transgene (Fig. 2.3C). Thus the *H3-H4* intergenic region containing the bidirectional promoter is the critical element for HLB formation.

In addition to transferring the capability for HLB assembly, the promoter swap also resulted in expression of the *H2a* transgene as determined by S1 nuclease protection and qRT-PCR (Fig. 2.3D,E). The *H2a* gene in *H2a-H2b*^{PS} was expressed at levels similar to the *H4* gene in the *H3-H4* transgene (about 15% of the intact repeat; Fig. 2.3D, E), while the *H4* gene in *H3-H4*^{PS} was no longer expressed. Note that the promoter swap results in a smaller protected fragment with the H2a probe (Fig. 2.3D, lane 4) because the chimeric *H2a* gene now contains the *H4* 5'UTR, causing the S1 nuclease to cleave at the end of the FLAG tag rather than at the end of the *H2a* 5' UTR (Fig. 2.3D, diagram). Also note that with the overloading of the endogenous histone mRNA, a low-level of expression (~3%) of the *H2a* gene in the original *H2a-H2B* gene pair is detected by the S1 nuclease assay (Fig. 2.3D, lane 3). Correspondingly, expression of the *H4* gene in *H3-H4*^{PS} was very low, and essentially undetectable with the S1 protection or the qPCR assay (Fig. 2.3D, lane 8; 2.3E). These results demonstrate that there is a sequence in the *H3-H4* promoter that directs HLB assembly and high level histone gene expression.

Figure 2.3

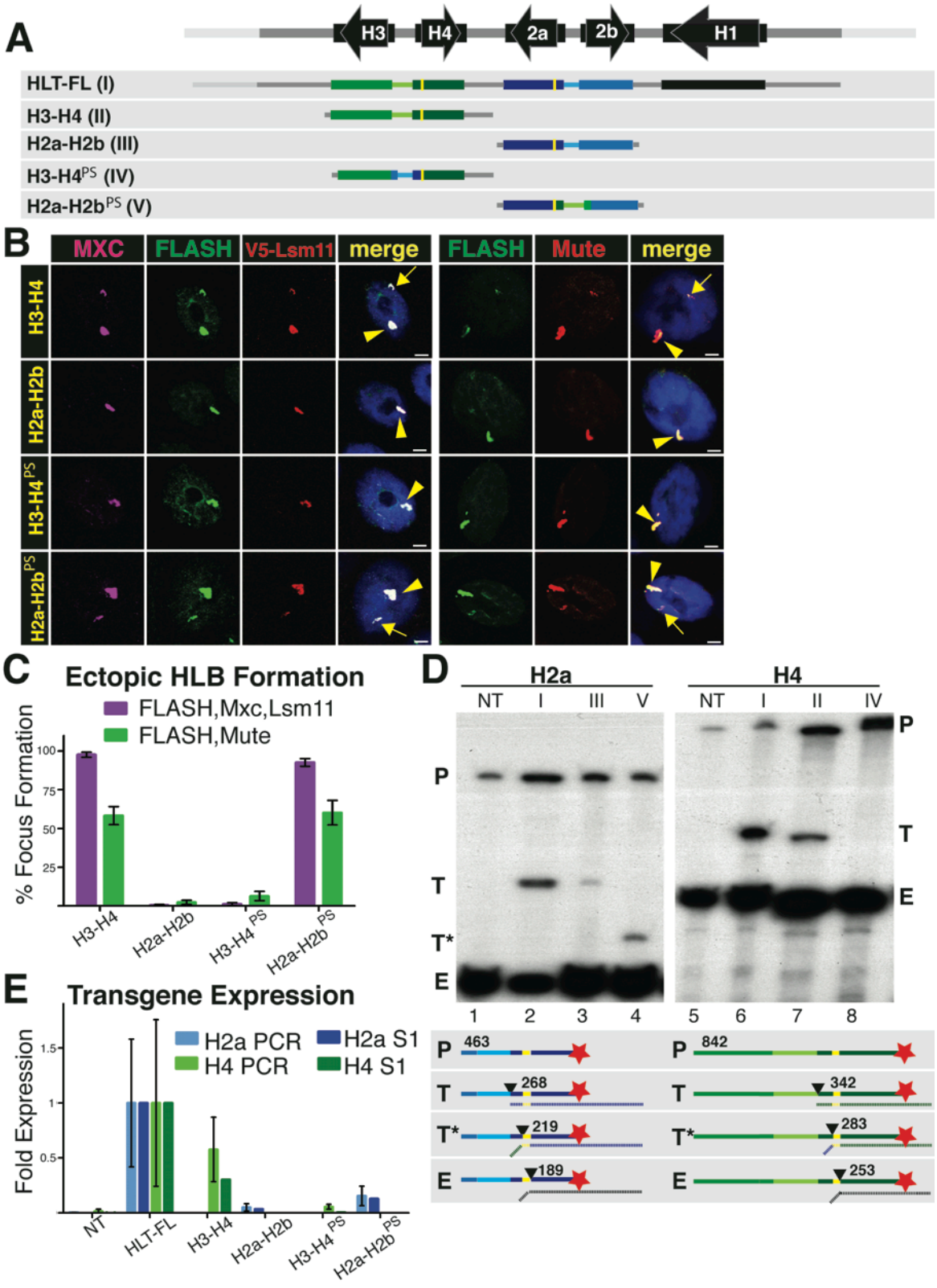

A
H3
H4
2a
2b
H1
HLT-FL (I)
H3-H4 (II)
H2a-H2b (III)
H3-H4PS (IV)
H2a-H2bPS (V)
B
MXC
FLASH
V5-Lsm11
merge
FLASH
Mute
merge
H3-H4
H2a-H2b
H3-H4 PS
H2a-H2b PS
C
Ectopic HLB Formation
FLASH,Mxc,Lsm11
FLASH,Mute
% Focus Formation
100
75
50
25
H3-H4
H2a-H2b
H3-H4 PS
H2a-H2b PS
D
H2a
H4
NT
I
III
V
NT
I
II
IV
P
T
T*
E
1 2 3 4 5 6 7 8
P 463
T 268
T* 219
E 189
P 842
T 342
T* 283
E 253
E
Transgene Expression
H2a PCR
H4 PCR
H2a S1
H4 S1
Fold Expression
1.5
1
0.5
NT
HLT-FL
H3-H4
H2a-H2b
H3-H4 PS
H2a-H2b PS

Figure 2.3. **The *H3-H4* Promoter Assembles an Ectopic HLB.**
A. Diagram of the five constructs inserted into chromosomal location 86Fb. The yellow bars represent N-terminal FLAG tags in *H2a* and *H4.* The promoter swap (PS) includes the bidirectional promoter and 5'UTR from each gene pair.
B and **C.** HLB assembly for the indicated constructs was assessed with the indicated markers and quantified as in Figure 2. V5 antibody was used to detect Lsm11 in a strain where V5-Lsm11 replaces the endogenous protein. Note that the *H3-H4* promoter (*H3-H4, H2a-H2b*^{PS}) assembles an HLB, regardless of the associated transcript.
D and **E.** Histone gene expression was assessed and quantified as in Figure 2. Numbers in the diagram indicate the length in nt of the probe and possible protected fragments. Roman numerals refer to the depicted transgene, and NT is the no transgene control. Note robust histone mRNA expression from *HLT-FL, H3-H4* (T) and *H2a-H2b*^{PS} (T*)

mRNA Processing Signals are Dispensable for HLB Assembly

Our results thus far reveal an HLB assembly element in the *H3-H4* intergenic region and a strong correlation between HLB formation and histone gene expression, but we cannot conclude a causal relationship between these two activities. For example, do the unique 3' processing elements of a histone pre-mRNA contribute to HLB assembly? To determine whether mRNA produced from an intact histone gene influences HLB formation, we introduced just the 300 nt histone *H3-H4* intergenic region (from start codon to start codon) into 86Fb (Fig. 2.4A). This fragment, *H3-H4P*, efficiently recruited Mxc, FLASH, and Mute, as well as U7 snRNP (Lsm11) (Fig. 2.4B,C), and was transcriptionally active. Hybrid transcripts containing either the H3 or H4 5'UTR and respective flanking vector sequences were detected by RT-PCR (Fig. 2.4D). *H3-H4P* generates transcripts containing only the 57 and 59 nt 5' UTRs of the histone H3 and H4 mRNAs followed by flanking vector sequence, but contains no histone ORF, 3' UTR or pre-mRNA processing signals. Hence, the observation that both FLASH and U7 snRNP were recruited by the *H3-H4P* construct rules out the possibility that the histone processing factors are recruited to the HLB by interacting with *cis* elements in the nascent transcript. They must be recruited directly to the HLB. However, as with the intact histone genes, there is still strong correlation between HLB assembly and transcription.

Figure 2.4

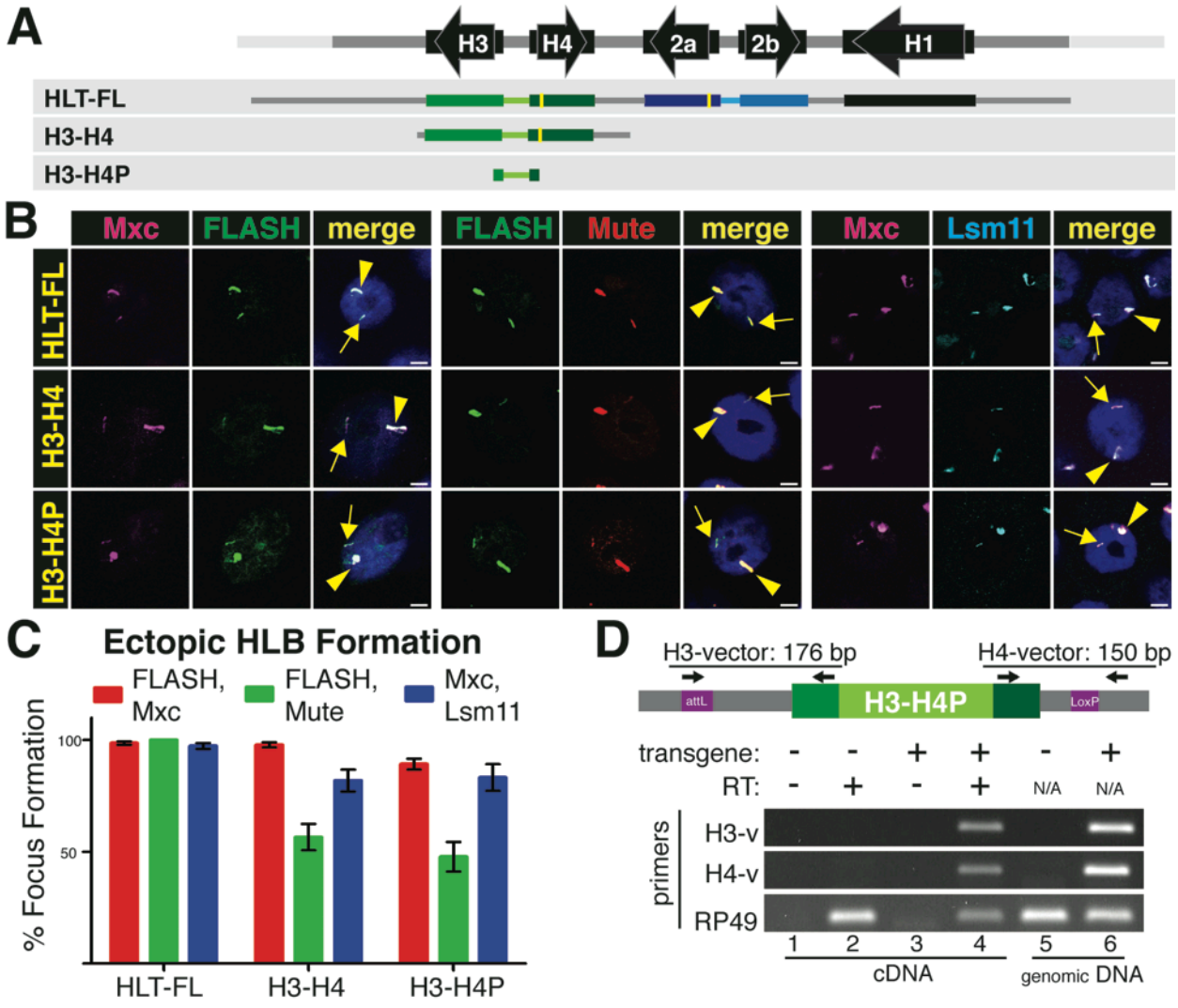


Figure 2.4. **The *H3* and *H4* coding region and 3' processing signals are not required for HLB assembly.**
A. Diagram of the three constructs inserted into chromosomal location 86Fb. The yellow bars represent N-terminal FLAG tags in *H2a* and *H4.*
B and **C.** HLB assembly for the indicated constructs was assessed with the indicated markers and quantified as in Figure 2. Note that all three constructs assemble an HLB
D. Transcription from the *H3-H4P* transgene was assessed by RT-PCR using a primer in either the H3 or H4 5'UTR and corresponding flanking vector sequence as diagramed above the gel. Transcripts were detected in the *H3-H4P* strain (lane 4) and not the NT control. RP49 transcripts were detected in all cDNA preparations. Genomic DNA was analyzed in lanes 5 and 6, confirming the presence of the transgene.

Transcription Stimulates HLB Maturation

In the developing embryo, the transition from proto-HLB to a mature HLB occurs at the onset of zygotic transcription (White et al., 2011). To test whether transcription plays a direct role, we determined the effect of preventing transcription from the *H3-H4* promoter on HLB formation. We also asked if transcription from this promoter was necessary for expression of other histone genes in the cluster. To address these questions we inserted a transgene (*CORE*) at 86Fb containing just the 4 core histone genes (i.e. *H3-H4* and *H2a-H2b* gene pairs) and a nearly identical transgene ($CORE^{DT}$) different only in that both TATA boxes in the *H3-H4* promoter were mutated (Fig. 2.5A). FLAG tags on the *H4* and *H2a* genes and an HA-tag on the *H3* gene allowed us to measure expression from these transgenes. The *CORE* construct assembled an ectopic HLB as efficiently as the full repeat unit (*HLT-FL*) and more efficiently than the *H3-H4* gene pair, and both the *H4* and *H2a* genes were expressed at levels close to that of the full repeat (Fig. 2.5B, C). As expected there was no expression of the *H4* gene or the *H3* gene from the $CORE^{DT}$ construct, as analyzed by S1 nuclease mapping (Fig. 2.5D, lane 4) or RT-PCR (not shown), respectively. Interestingly, Mxc and FLASH were recruited to the $CORE^{DT}$ construct, although the intensity of the signals and the frequency of HLB assembly were substantially lower than with the CORE construct (Fig. 2.5B,C; Fig. 2.6). In addition, Mute was not recruited to this construct (Fig. 2.5B,C). Thus, some HLB components, including Flash a pre-mRNA processing factor, assemble into a proto-HLB containing Mxc and FLASH in the absence of *H3-H4* transcription.

Surprisingly, while H2a was expressed equally well from the *HLT-FL* and *CORE* constructs (Fig. 2.5E, lanes 2 and 3), there was little expression of the histone *H2a* gene from the $CORE^{DT}$ construct (Fig. 2.5E, lane 4). H2a expression from $CORE^{DT}$ was at least 10-fold less than that from the histone *CORE* construct (Fig. 2.5F), and similar to both non-transgenic controls (Fig. 2.5F) and the low level found from the single *H2a-H2b* gene pair (Fig. 2.2E,F). Thus, we conclude that transcription from the histone *H3-H4* promoter is essential for activation

of transcription of the histone *H2a* gene and likely the *H2b* gene as well, and for the stable recruitment of Mute to the HLB. Furthermore, these data suggest that HLB assembly nucleated at the *H3-H4* intergenic region, together with expression from these promoters, is required for complete assembly of the HLB and full expression of all core histone genes in the repeat. The formation of a proto-HLB rather than a complete HLB, evidenced by failure of Mute to accumulate on the ectopic TATA mutant transgene, suggests that Mute recruitment, and hence HLB maturation, depends on transcription initiation from the *H3-H4* promoter, and is not solely directed by a sequence element in the histone locus. We also observed a reduction in, but not the absence of, Mute localization to the *H3-H4* construct (Fig. 2.2C, 2.5C). We hypothesize that the amount of Mute recruitment to the HLB may be related to the amount of transcription of the histone gene cluster, and that expression from two genes recruits insufficient amounts of Mute for us to detect it in all HLBs. Alternatively, Mute recruitment might require sequences from both *H3-H4* and *H2a-H2b* genes. To distinguish between these possibilities, we replaced the *H2a-H2b* gene pair in the CORE construct with a second copy of *H3-H4* (Fig. 2.5A). In order to directly compare the level of *H4* gene expression from this transgene to our other transgenes, only one of the two *H4* genes contained a FLAG sequence. Mute recruitment to the *H3-H4/H3-H4* construct was higher than to the *H3-H4* construct and comparable to that of the *H3-H4/H2a-H2b CORE* transgene (Fig. 2.5B,C). H4 expression from the *H3-H4/H3-H4* construct also increased 1.6 fold compared to the *H3-H4* construct (Fig. 2.5G,H). We conclude that Mute recruitment to the HLB positively correlates with the number of active histone gene promoters, and suggest that transcription from the *H3-H4* promoter at the histone locus is an essential step in the development of a mature and stable HLB.

Figure 2.5

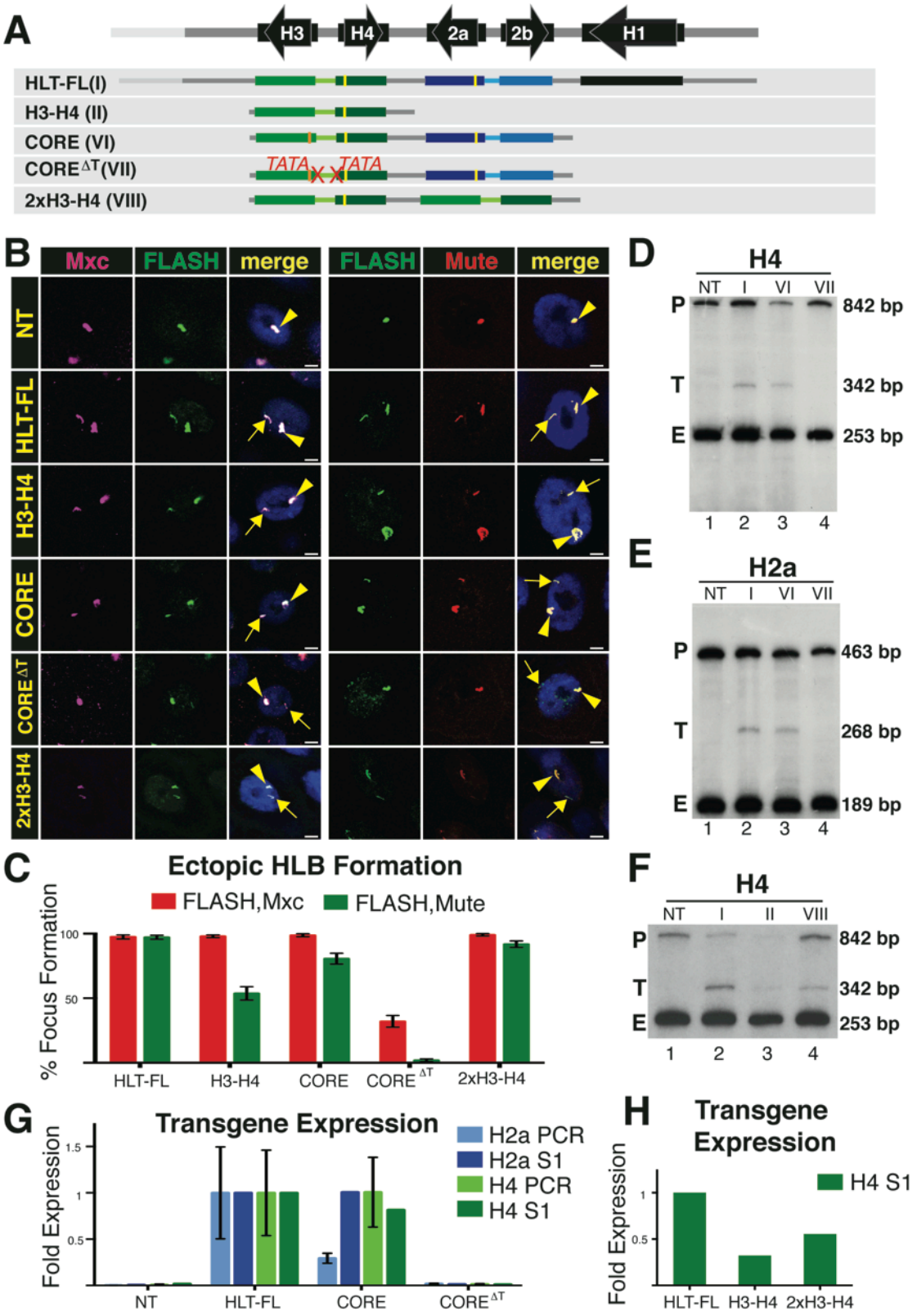

A
H3
H4
2a
2b
H1
HLT-FL(I)
H3-H4 (II)
CORE (VI)
CORE ΔT (VII)
TATA
TATA
2xH3-H4 (VIII)
B
Mxc
FLASH
merge
FLASH
Mute
merge
NT
HLT-FL
H3-H4
CORE
CORE ΔT
2xH3-H4
C
Ectopic HLB Formation
FLASH,Mxc
FLASH,Mute
% Focus Formation
100
50
HLT-FL
H3-H4
CORE
CORE ΔT
2xH3-H4
D
H4
NT I VI VII
P
T
E
842 bp
342 bp
253 bp
1 2 3 4
E
H2a
NT I VI VII
P
T
E
463 bp
268 bp
189 bp
1 2 3 4
F
H4
NT I II VIII
P
T
E
842 bp
342 bp
253 bp
1 2 3 4
G
Transgene Expression
H2a PCR
H2a S1
H4 PCR
H4 S1
Fold Expression
1.5
1
0.5
NT
HLT-FL
CORE
CORE ΔT
H
Transgene Expression
H4 S1
Fold Expression
1
0.5
HLT-FL
H3-H4
2xH3-H4

Figure 2.5. **Transcription is required for HLB assembly**.
A. Diagram of the constructs inserted into chromosomal location 86Fb. HLB assembly was assessed and quantified as in Figure 2. The orange bar indicates the HA tag added to the *H3* gene. The 2867 bp *CORE* construct contains both the *H2a-H2b* and *H3-H4* gene pairs. The *CORE*^{DT} construct contains mutations in both the *H3* and *H4* TATA (DT) boxes. *2xH3-H4* contains a duplication of the *H3-H4* gene pair in which only one of the *H4* transgenes contains a FLAG tag.
B and **C.** HLB assembly for the indicated constructs was assessed with the indicated markers and quantified as in Figure 2. Note that while the *CORE* construct assembles an HLB, mutating the *H3* and *H4* TATA boxes reduces Mxc/FLASH assembly and results in undetectable Mute accumulation. Also note that increasing the number of transcription units increases Mute recruitment (compare *H3-H4* and *2xH3-H4*).
D-H. Histone gene expression was assessed and quantified as in Figure 2. Roman numerals refer to the depicted transgene, and NT indicates a no transgene control. Note the expected lack of transgenic H4 transcription (Fig.2.5 D) and absence of ectopic H2a (Fig.2.5 E) from the *CORE*^{DT} construct. Also note that ectopic H4 mRNA levels increased upon addition of another *H3-H4* gene pair (F).

Figure 2.6

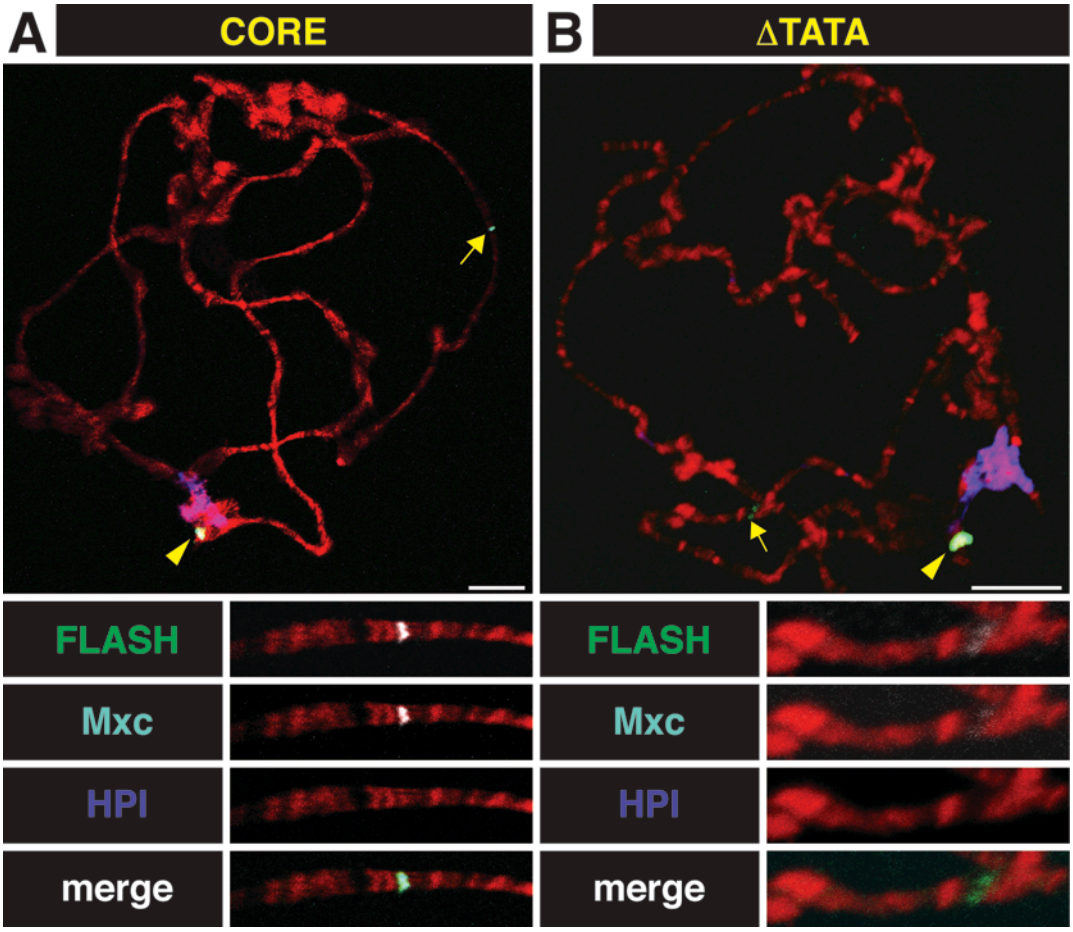


Figure 2.6. ***CORE* and *CORE$^{\Delta T}$* Assemble an Ectopic HLB**
A and **B**. Chromosome squashes from salivary glands of third instar larvae containing the *CORE* (left; n=8) or *CORE$^{\Delta T}$* (right n=4) constructs inserted at 86Fb and stained with FLASH (green), Mxc (cyan) and HP1 (blue). DNA is visualized with DAPI (red). The arrowhead indicates the endogenous HLB and the arrow indicates chromosomal position 86Fb. The insets show a higher magnification of the 86Fb locus. Bars = 20 mm.

Stable Assembly of the HLB During Development Requires the Histone Gene Cluster

Our transgenic experiments in salivary glands have determined which sequences from the histone locus are sufficient to nucleate ectopic HLB formation. To determine whether formation of the HLB requires histone genes, we took advantage of a *Drosophila* mutation, *Df(2L)Ds6*, in which the entire histone gene cluster is deleted (Moore et al., 1983) . Heterozygote *Df(2L)Ds6*/+ females are viable and fertile, and deposit sufficient maternal histone protein into the egg (shown for WT in Fig. 2.7G) such that when mated to *Df(2L)Ds6*/+ males, the resulting homozygous mutant embryos lacking histone genes develop normally through S phase of cycle 14 (~3 hours of development), after which zygotic histone expression is required for normal S phase and cell cycle progression (Gunesdogan et al., 2010) . If the histone locus is essential for all HLB components to assemble into a nuclear body, then we should not observe nuclear focus formation with any HLB marker in *Df(2L)Ds6* homozygous mutant embryos. However, if any HLB components have self-organizational properties, we may detect nuclear foci with particular HLB markers even in the absence of histone genes.

To address these questions, we stained populations of syncytial blastoderm staged embryos collected from *Df(2L)Ds6*/+ heterozygous parents with antibodies against FLASH and Mxc or with antibodies against FLASH and Mute. These embryos were also stained with MPM-2 monoclonal antibodies, which detects a phosphoepitope on Mxc present in cells with active Cyclin E/Cdk2 (Fig. 2.7A-E, Fig. 2.8A,B) (White et al., 2011; White et al., 2007). Interestingly, all embryos at cycles 11 (when mature HLB assembly occurs and histone transcription normally begins) and 12 contained nuclei with foci of co-localizing Mxc and FLASH that were all MPM-2 positive, even though 25% of these embryos lack histone genes (Fig. 2.7A,B; Fig. 2.8A,B).

However, the pattern of HLB marker staining differed between embryos collected from *Df(2L)Ds6*/+ parents versus wild type parents. In wild type embryos, all nuclei contained either 1 or 2 FLASH/Mxc foci or FLASH/Mute foci, which represent paired (1 focus) and unpaired (2 foci) homologous histone loci, respectively (Fung et al., 1998; White et al., 2007) . In *Df(2L)Ds6*/+ collections, ~75% of the embryos contained nuclei with either 1 or 2 Mxc + FLASH foci (Fig. 2.7A) and the remaining ~25% of embryos contained a small fraction (7%) of nuclei with 3 or more Mxc + FLASH foci, along with nuclei containing 1 or 2 foci (Fig. 2.7B,E). Because we never see nuclei with three or more Mxc + FLASH foci in wild type embryos, we conclude from these data that this phenotypic class represents the homozygous histone deletion genotype (Table 2.5; $p < 0.05$ via Chi-squared analysis).

In the embryos lacking histone genes, the FLASH foci were smaller than in controls (Fig. 2.8C). None of these foci stained intensely with Mute, in contrast to the cells in embryos containing histone genes at cycle 11, although weak Mute staining was detected in cells with one or two foci (and hence a greater amount of accumulated FLASH) (Fig. 2.7C-E). These results suggest that the Mxc/FLASH foci in histone deletion embryos are proto-HLBs. The percentage of nuclei with either one or two Mxc/FLASH foci in successive nuclear cycles was statistically similar between wild type and histone deletion embryos (Fig. 2.8D; $p<0.05$). These percentages reflect the degree of homologous chromosome pairing during early embryogenesis (Fung et al., 1998). This result suggests that Mxc/FLASH foci may be assembling on chromatin in the absence of histone genes. In addition, MPM-2 staining (and thus Cyclin E/Cdk2 phosphorylation of Mxc) occurs on these foci independent from histone transcription. Finally, by germ band extended stages (i.e. cycles 15-16) there were no nuclear foci present in the histone deletion embryos as assessed by staining for Mxc (Fig. 2.7F) or FLASH (not shown). These data indicate that Mxc and FLASH can self organize into NBs (proto-HLBs) in the absence of histone genes. Without the scaffold of the *H3-H4* promoter sequence, these proto-HLBs are not stable, mature HLBs, and do not persist through embryogenesis.

Figure 2.7

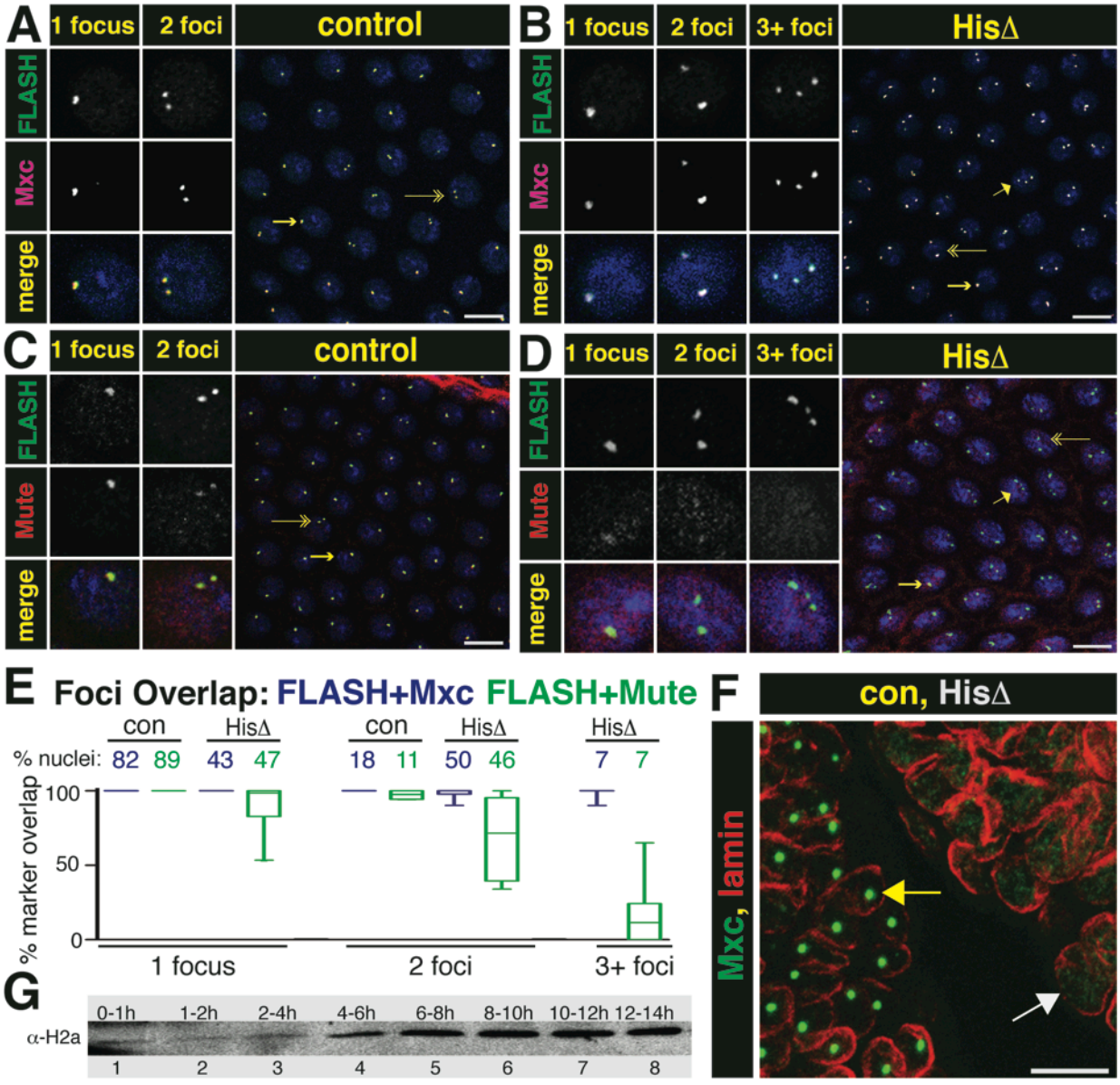


Figure 2.7. **HLB assembly in the absence of the Histone Locus.**
A.-D. Syncytial *Df(2L)Ds6* and control sibling embryos from a 1.5 to 2.5 hr collection were stained with aFLASH and aMxc (panels A,B) or aFLASH and aMute (panels C,D. Arrows indicate the nuclei chosen for split channel magnification to highlight cells with 1 focus (long arrow), 2 foci (double arrow) and 3+ foci (short arrow). Scale bar = 10 mm.
E. Percent overlap was calculated by measuring complete overlap of FLASH foci and either Mxc (blue) or Mute (green) in each nucleus. The percent of total nuclei represented in each class (control or histone deletion for each antibody pair) is given across the top X axis (FLASH+Mxc, con N=943, HisD N=758, FLASH+Mute, con N=1144, HisD N=1397). The box represents the 25th to 75th percentiles, the line in the box represents the median value, and the bars extend to the 10th and 90th percentiles.

F. MXC foci (green) are not present in germ band extended mutant embryos (cell cycle 15 and 16). Mxc HLBs are present in the GFP positive sibling control nuclei, as outlined by lamin (red, left embryo, yellow arrow), whereas no foci are present in the GFP negative histone deletion embryo (right, white arrow).
G. Western blot analysis of histone H2a levels from an equal number of staged wild type embryos.

Figure 2.8

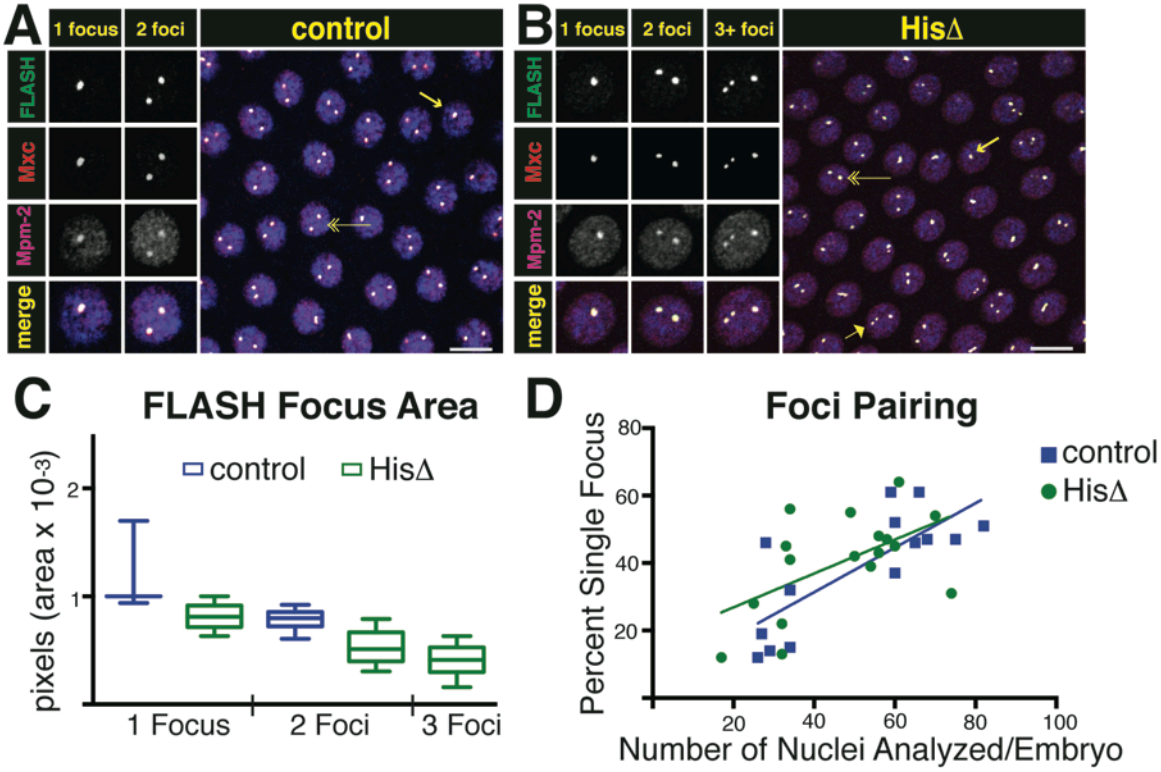


Figure 2.8 HLB assembly in the absence of the Histone Locus
A and **B.** Syncytial *Df(2L)Ds6* and control sibling embryos from a 1.5 to 2.5 hr collection were stained with aFLASH, aMxc and MPM-2 monoclonal antibody. Arrows indicate the nuclei chosen for split channel magnification to highlight cells with 1 focus (long arrow), 2 foci (double arrow) and 3+ foci (short arrow). Scale bar is 10 mm.
C. Focus size for control and HisD embryos was quantified as the area of FLASH staining in single plane confocal images using the Particle Analyzer function of ImageJ (Schneider et al., 2012). Each focus was assigned to a category based on genotype and number of foci in the nucleus (WT 1 focus n=12, Ds6 1 focus n=29, WT 2 foci n=95, Ds6 2 foci n= 137, Ds6 3+ foci n=45). In histone deletion embryos, the foci of the 3 foci class nuclei are significantly smaller ($P<0.0001$ t-test) than those with a single focus. The box represents the 25th to 75th percentiles, the line in the box represents the median value, and the bars extend to the 10th and 90th percentiles.
D. Nuclei containing a single focus were quantified in control and HisD blastoderm syncytial embryos. Embryonic nuclear divisions are restricted in space; therefore the number of nuclei in an image increases with time (X axis). Linear regression was performed for both populations

and both genotypes display a relationship between focus pairing and time ($p<0.05$). The regression coefficients did not differ significantly (t-test $p<0.05$).

Table 2.5. Identification of Histone Deletion Genotypes

	1 Foci	2 Foci	3 Foci
Genotype	+/-	+/+	-/-
Observed	29	13	17
Expected	29.5	14.75	14.75

DISCUSSION

In most organisms the genes encoding the five replication-coupled histone proteins are physically linked, suggesting that there has been selective pressure to maintain this linkage, which is not the case for most sets of genes (e.g. globins, ribosomal proteins) that are coordinately expressed. One reason for tight linkage of the histone genes might be to promote the localization of factors required for histone mRNA biosynthesis, particularly those needed to form the unique histone mRNA 3' end. A second reason might be to help control the coordinated, cell cycle-regulated expression of all the histone genes via a cis-acting element(s) at the locus. Our studies in *Drosophila* provide evidence for a ~300 nt sequence containing only the *H3-H4* bi-directional promoter that mediates the concentration of transcription and processing factors in the HLB and promotes expression of all the replication-coupled histone genes.

How do HLBs Form During Development?

Several lines of evidence suggest that Mxc and FLASH self-organize into a proto-HLB and participate in the earliest steps of HLB assembly. First, an Mxc/FLASH-containing body (that does not contain Mute) forms in cycle 10 of wild-type *Drosophila* embryos, one cycle prior to initiation of histone gene transcription (White et al., 2011). Similarly, in embryos that lack histone

genes we observe nuclear foci containing Mxc and FLASH, but not robust or consistent staining of Mute, in cycles 11 and 12 when wild-type embryos have formed a complete HLB and initiate histone gene expression. In the absence of histone DNA, the initial proto-HLBs are unstable. Furthermore, FLASH and Mxc, but not Mute, were recruited to the ectopic transgene (*CORE^{DT})* in which *H3* and *H4* transcription was ablated. Finally, during mitosis Mxc and FLASH remain associated with the mitotic chromosomes, while Mute and U7 snRNP do not (White et al., 2011), and likely serve to nucleate formation of a mature HLB after each cell division. We suggest that the proto-HLB represents an initial intermediate in HLB assembly that associates tightly with the *H3-H4* promoter.

Although many NBs are not associated with a genomic locus, it has been proposed that a nucleotide sequence (often RNA) "seeds" formation of the structure (Dundr, 2011). Here we show that a ~300 nt sequence containing the bi-directional *H3-H4* promoters is sufficient to recruit multiple HLB components, while the *H2a-H2b* and *H1* promoters, or other regions of the histone repeat, are not. Interestingly, the *H3-H4* promoter fragment alone formed an HLB, indicating that the generation of full-length transcripts is dispensable for HLB assembly. However, preventing *H3-H4* transcription suspends HLB development at the proto-HLB step. The inactivated *H3-H4* promoter stabilizes this intermediate in HLB assembly. We conclude that while the 300nt sequence between the *H3* and *H4* genes provides a scaffold for HLB assembly, activity from the *H3-H4* promoter is necessary to form the mature HLB.

The precise role of transcription from the *H3-H4* promoter in HLB formation is not clear from our studies. Recruitment of the additional factors (Mute and U7 snRNP) to the HLB could be mediated through the assembly of the core transcription machinery at the *H3-H4* promoter, a change in the phosphorylation status of RNA Polymerase II, and/or active transcription opening up the adjacent chromatin. Since many components of the RNA Pol II machinery assemble on promoters and remain "poised" even at times when transcripts are not being actively generated (Nechaev et al., 2010; Zeitlinger et al., 2007) , one possibility is that components of the

transcription machinery bind the *H2a-H2b* promoter and require a signal from the *H3-H4* promoter for initial activation. We also cannot distinguish whether activation of the histone *H2a-H2b* genes either simply requires transcription from the *H3-H4* promoter or formation of the mature HLB. We speculate that the initial transcription from the *H3-H4* promoter stimulates mature HLB formation encompassing the entire histone repeat by facilitating the recruitment of additional Mxc and FLASH to the chromatin.

How might the 300nt between *H3* and *H4* activate transcription of the *H2a-H2b* genes? This sequence has some properties of an enhancer in that it activates genes from a distance. Since this sequence contains the core promoter for the *H3-H4* genes it seems unlikely that it acts like a classical enhancer, by looping the chromatin between the *H3-H4* and *H2a-H2b* promoters. In addition, the function of the putative enhancer would not be affected by mutation of the TATA box. It seems more likely that transcription from the *H3-H4* promoter, which leads to recruitment of the additional HLB factors, results in activation of the *H2a-H2b* genes. This could result either from the HLB factors altering the chromatin structure throughout the histone locus or by them directly recruiting coactivators to the histone genes.

Model for HLB Assembly

In our model of HLB formation (Fig. 2.9), the initial event is a stochastic association of Mxc and FLASH that is triggered by an unknown mechanism at embryonic cycle 10 and that can occur independently of the histone genes. This Mxc/FLASH proto-HLB complex associates with the histone *H3-H4* promoter prior to transcription of the histone locus, and provides a platform for the subsequent recruitment of the remaining HLB components. Recruitment of additional components, such as Mute, requires assembly of the core transcription complex or actual transcription from the promoter. Neither FLASH nor Mxc has obvious sequence-specific DNA binding domains, and current evidence suggest that in mammals the Mxc orthologue NPAT functions as a co-activator binding to some component of a complex present on the promoter,

rather than binding directly to the DNA itself (Miele et al., 2005; Wei et al., 2003; Ye et al., 2003) . There may be another, as yet undefined, component of the HLB that directly binds to the *H3-H4* promoter as well as to Mxc/FLASH to form the initial complex on the histone gene repeat, or else a unique feature of the chromatin in this region recruits Mxc/FLASH. Close inspection of the 300 nt sequence in 12 *Drosophila* species did not reveal any highly conserved elements in the H3/H4 promoter other than the TATAA boxes (Fig. 2.10).

HLBs Compared with Other Nuclear Bodies

The HLB differs from most NBs (e.g. Cajal bodies and PML bodies) because it is constitutively associated with a specific locus and the biosynthesis of a specific class of mRNAs. The nucleolus, which also assembles at a specific, repetitive gene locus, is one nuclear body with many similarities to the HLB. Indeed, a similar approach to ours that utilized ectopic rDNA genes in the *Drosophila* salivary gland has been used to address issues of nucleolar formation (Karpen et al., 1988) . Transcription complexes including PolI, SL1 and Ubf form on arrays of rRNA promoters containing no rDNA coding regions, resulting in assembly of a body that has some features of the nucleolus (Prieto and McStay, 2008) . Maintenance of morphologically complete nucleoli at rDNA genes requires both transcription and processing of rRNA (Hernandez-Verdun, 2006) . Pre-nucleolar bodies form in experimentally induced micro-nuclei containing no rDNA (Hernandez-Verdun et al., 1991) . Thus, there are likely to be similarities between the assembly process of the nucleolus and the HLB (Denissov et al., 2011) . However, the incredible complexity of the nucleolus makes it difficult to comprehensively investigate the relationship between nucleolar structure and its multiple associated functions, some of which are not specifically involved in rRNA production (reviewed in (Boisvert et al., 2007) .

What are the Functions of the HLB?

The HLB components that we have studied here (Mxc, FLASH, Mute and U7 snRNP) are all concentrated exclusively in the HLB and each is essential for proper *Drosophila* development (Bulchand et al., 2010; Godfrey et al., 2006; Godfrey et al., 2009; Saget et al., 1998) (D.C.T, W.F.M. and R.J.D, unpublished). However, we cannot conclude from this observation that the HLB itself is essential. FLASH and U7 snRNP have clearly defined biochemical functions in histone pre-mRNA processing (Burch et al., 2011b; Godfrey et al., 2006; Godfrey et al., 2009; Yang et al., 2009), while NPAT is essential for histone gene expression in mammals and *Drosophila* (White et al., 2011; Ye et al., 2003). The biochemical function of Mute is not known (Bulchand et al., 2010). HLBs may enhance the efficiency or rate of biochemical reactions associated with histone mRNA biosynthesis by increasing the local concentration of low-abundance factors at the histone locus. For instance, we previously found U7 snRNP is expressed but not localized to HLBs in *Drosophila H2aV* mutants, resulting in misprocessing of histone mRNA (Wagner et al., 2007). Concentration of histone biosynthetic factors in the HLB may also provide a mechanism for the coordination of gene expression. It is remarkable that the TATA mutation in the $CORE^{DT}$ construct not only blocked *H3-H4* transcription, but also suppressed transcription from the neighboring *H2a* gene containing an intact promoter, suggesting a role for the HLB in coordinating expression of all the histone genes. Precisely defining the HLB nucleation sequence will allow us to dissect the molecular details of control of coordinate expression of the multiple genes in the histone locus, as well as the specific interactions that direct HLB assembly.

Figure 2.9

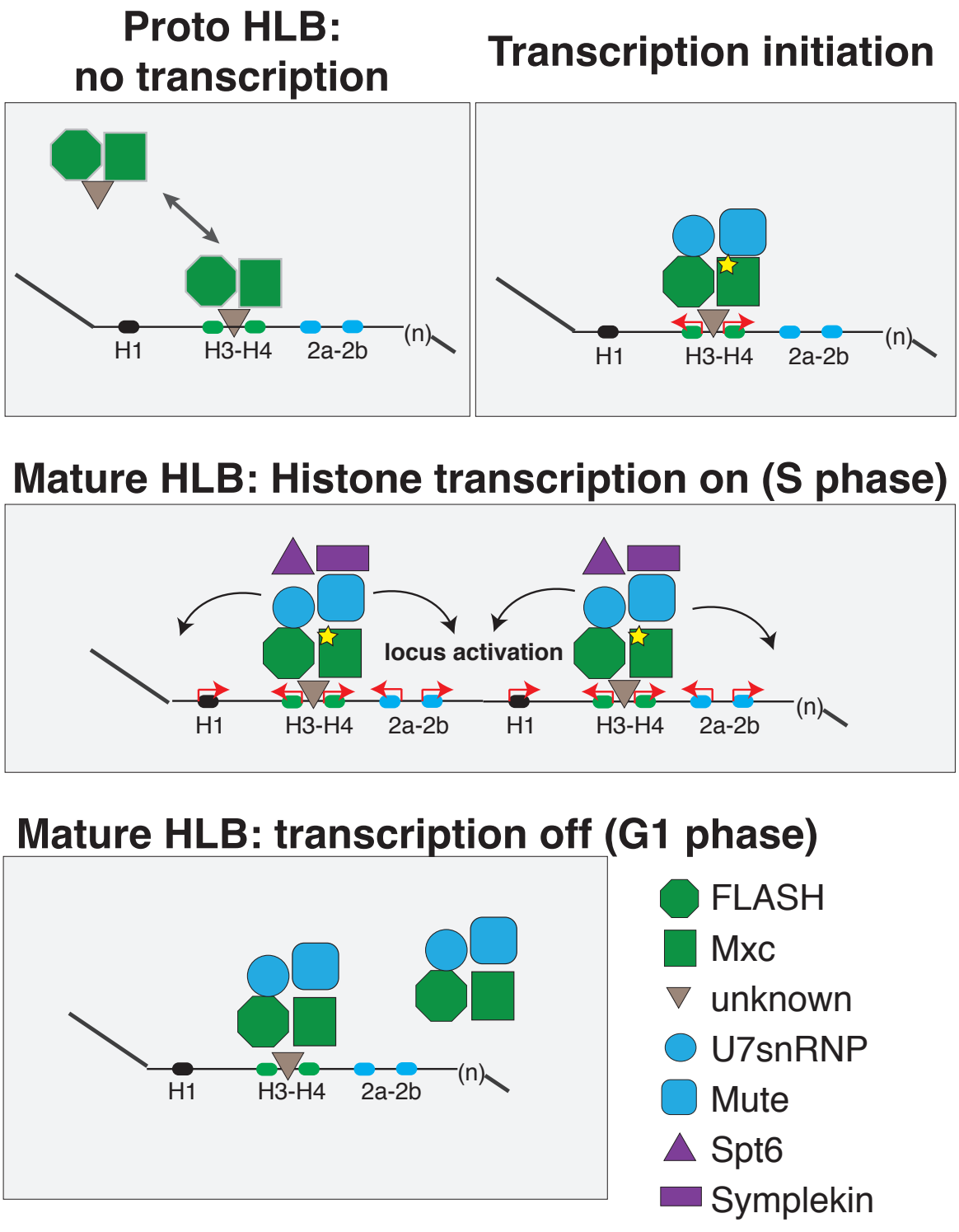
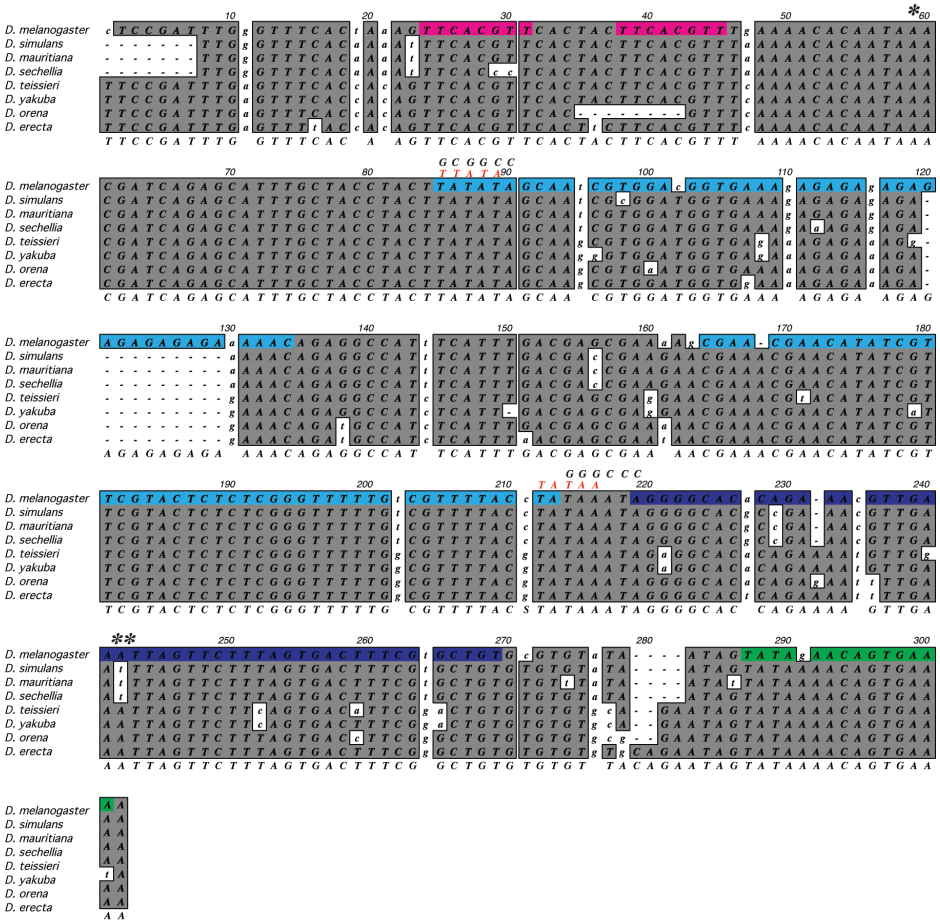

Proto HLB:
no transcription
Transcription initiation
H1
H3-H4
2a-2b
(n)
Mature HLB: Histone transcription on (S phase)
locus activation
Mature HLB: transcription off (G1 phase)
FLASH
Mxc
unknown
U7snRNP
Mute
Spt6
Symplekin

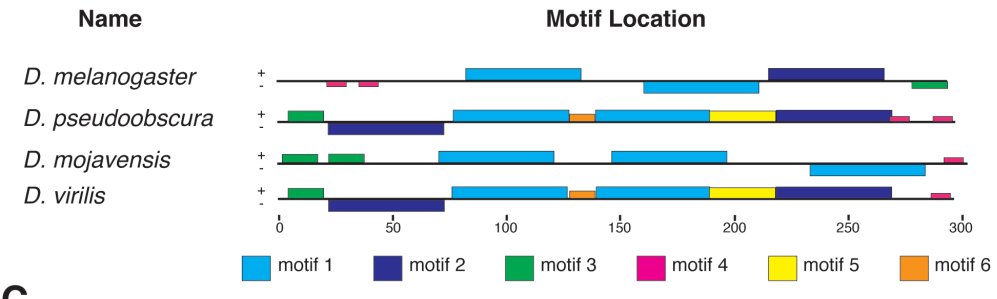
Figure 2.9. Model for HLB assembly and maintenance
FLASH and Mxc form a “proto-HLB” in the absence of transcription. The proto-HLB is detected in syncytial stage embryos (cycle 10), histone deletion mutant embryos, and loci where histone transcription is abolished through mutation of the *H3-H4* TATA boxes. While FLASH and Mxc are the earliest known components that begin to organize the HLB, it is possible that an additional factor(s) facilitates the interaction between Mxc/FLASH and the histone locus (grey triangle). Next, Cyclin E-mediated phosphorylation of Mxc (star) coincides with the onset of zygotic transcription during embryonic cycle 11. Transcription initiation from the *H3-H4* promoter recruits Mute and U7 snRNP (Lsm11) to the HLB, and FLASH and Mxc continue to accumulate at the locus. S phase expression of all replication dependent histone genes transiently recruits to the locus other factors required for mRNA biogenesis, such as Spt6 and Symplekin. Finally, once established, the HLB remains associated with the locus in the absence of histone gene expression.

Figure 2.10

A



B



C

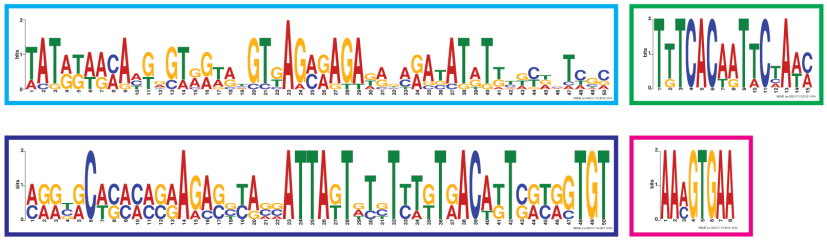


Figure 2.10. Conservation of the *H3-H4* Intergenic Region.
A. The region between the *H4* and *H3* coding sequences was aligned with Clustal W and visualized with MacVector for drosophilid species of the *melanogaster* subgroup (Olson, 1994). The transcription start sites (H4 *, H3 **), TATAA boxes, changes in $CORE^{\Delta T}$, and predicted motifs (colored regions) are indicated.
B. Sequence analysis for species outside of the *melanogaster* subgroup was performed with MEME (Bailey and Elkan, 1994) to identify potential motifs in the *H3-H4* regulatory region. Six different potential motifs were identified and their position within the *H3-H4* intergenic region is indicated by a different color. Motifs present in all four species are highlighted in the alignment (A).
C. The four motifs present in all four species of the *melanogaster* subgroup are displayed as a position weight matrix (E values - motif 1: 1.5e-21, motif 2: 3.8e-14, motif 3: 1.4e+2, motif 4: 1.9e+3).

ACKNOWLEDGEMENTS

We thank Dan McKay and Greg Matera for comments on the manuscript, Jess Nesmith for help analyzing histone deletion embryos, Dr. Joe Gall for Lsm11 antibody, Prem Fort for the histone western blot, and Best Gene for fly transgenesis. This work was supported by NIH grants GM58921 to W.F.M. and GM57859 to R.J.D.

CHAPTER 3: CONCENTRATING PRE-MRNA PROCESSING FACTORS IN THE HISTONE LOCUS BODY FACILITATES EFFICIENT HISTONE GENE EXPRESSION

PREFACE

This work represents a manuscript currently in preparation. My advisors, Dr. William Marzluff and Dr. Robert Duronio and I designed the experiments and analyzed the data. I performed all of the experiments with the following exceptions: Grzegorz Zapotoczny identified the *FLASH*^{LL01602} insertion site shown in Fig.3.2, the processing assay in Fig.3.3F was performed by Ivan Sabath, the GFP-G46 salivary gland staining in Fig.3.6C was performed by Esteban Terzo, the salivary gland staining shown in Fig.3.8A was performed by Dr. Harmony Salzler, and the RNA pol II occupancy shown in Fig.3.9A was mapped and displayed by Dr. Daniel McKay. I wrote the first draft of the manuscript, which was then edited and expanded with Drs. Duronio and Marzluff.

INTRODUCTION

Regulatory information present in DNA and nascent pre-RNA recruits and directs the activity of *trans* acting factors to produce a mature RNA molecule. During this process cells are challenged with the task of assembling diverse molecular machines from a pool of *trans* factors that must navigate the complex 3-dimensional structure of the genome in order to interact with the proper DNA and nascent RNAs in the nucleus. Recent studies mapping long-range chromosome interactions have refined our understanding of how DNA is arranged in the nucleus (Dekker et al., 2013; Marshall et al., 1997), and emphasized that this organization is not random. Integrated together with this chromosome organization are structures termed nuclear

bodies (NBs). Defined microscopically by co-localization of components, each NB contains a suite of factors associated with particular nuclear processes including gene expression, DNA replication or DNA repair (Gall, 2000; Matera et al., 2009; Morimoto and Boerkoel, 2013).

NBs are postulated to provide discrete microenvironments that increase rates of nuclear processes by concentrating reactants and substrates (Dundr, 2011; Mao et al., 2011; Matera et al., 2009). However, the evidence supporting this hypothesis is indirect, relying on modeling as well as eliminating a NB altogether rather than preventing the concentration of specific components. For example, the rate of association for the U4/U6 di-snRNP and U4/U6 U5 tri-snRNP is an order of magnitude higher in the Cajal body than the nucleoplasm (Klingauf et al., 2006; Novotny et al., 2011), and a reduced rate of tri-snRNP assembly was postulated to cause the splicing defects observed in *coilin* mutant zebrafish that lack Cajal bodies (CBs) (Strzelecka et al., 2010). In contrast, eliminating the *Drosophila* CB through mutation of the gene *coilin* or mis-localization of small CB-specific RNAs (scaRNAs), which guide snRNA 2′-O-methylation and pseudouridylation, does not prevent snRNA modification or effect development (Deryusheva and Gall, 2009). In these *Drosophila* experiments it was not determined if the rates of snRNA modification reactions were altered. Therefore it remains unclear whether concentrating specific factors in a NB impacts the rate of a nuclear process.

Efforts to understand NBs have primarily focused on identifying components and determining mechanisms of NB assembly by depleting or ectopic tethering of specific NB components (Dundr et al., 2004; Liu et al., 2009; Mao et al., 2011; Shevtsov and Dundr, 2011; Strzelecka et al., 2010; White et al., 2011). However, these approaches do not specifically address the contribution of the body itself to a particular nuclear process. To determine whether the body itself contributes to a process, it is necessary to have sufficient mechanistic understanding of how individual components are targeted to a NB in order to mis-localize a component without changing its overall level within the cell or disassembling the body. In addition, the ability to accurately measure *in vivo* a specific process directly associated with a

NB is critical. These criteria are satisfied by the histone locus body (HLB), a NB that assembles at histone gene clusters and is involved in histone mRNA synthesis (Liu et al., 2006). Here, we test whether locally concentrating *trans* factors in the HLB affects the rate of histone pre-RNA processing.

Histone genes encode the only known mRNAs that end in a stem loop structure rather than a poly (A) tail (Marzluff et al., 2008). Formation of this unique mRNA 3' end requires U7 snRNP, FLASH (FLICE-associated huge protein), and a complex of polyadenylation factors termed the Histone Cleavage Complex (HCC) (Fig. 3.1A). Two of these components specifically accumulate in the HLB: U7 snRNP, which directly binds histone pre-mRNA at the histone downstream element (HDE) 3' of the stem loop (Mowry and Steitz, 1987), and FLASH, a protein found to be localized at the histone locus in mammalian cells and to be important for histone gene expression (Barcaroli et al., 2006; Bongiorno-Borbone et al., 2008). We subsequently determined that FLASH binds Lsm11, a U7 snRNP specific protein, and is essential for histone pre-mRNA processing in both mammals and flies (Burch et al., 2011; Yang et al., 2009).

The localization of U7 snRNP and FLASH to the HLB raises the possibility that concentrating these components in the HLB promotes their recruitment to histone pre-mRNA, thereby facilitating co-transcriptional cleavage. Here we report that when FLASH or U7 snRNP are produced at normal levels within cells but not concentrated in the HLB, histone pre-mRNA processing is slowed, resulting in transcriptional read-through and the accumulation of unprocessed histone transcripts at the histone locus, as well as small amounts of mis-processed, polyadenylated mRNA. In addition, mutations in FLASH that reduce its pre-mRNA processing activity cause more severe mutant phenotypes *in vivo* when the mutant FLASH protein is not localized to the HLB. These data indicate that localizing factors to the HLB increases the rate of histone mRNA 3' end formation, and demonstrate that concentrating factors within a NB influences gene expression.

MATERIALS AND METHODS

Drosophila Strains

PBac{PB}FLASH^{LL01602} was a gift from Greg Matera who obtained it from the *Drosophila* Genomics Resources Center. The site of insertion was determined by PCR and sequencing (Fig. 3.2, Table 3.3). FLASH constructs previously generated in pIZ/V5 (Life Technologies, (Burch et al., 2011)) were inserted into pATTB (Bischof et al., 2007) with KpnI and XhoI for fC31-mediated transgenesis (Best Gene) and integration into the attP40 landing site. Transgenic GFP-Mxc^{G46} was expressed with the ubiquitin promoter of pUGW in the *mxc*^{G48} null background (White et al., 2011) after integration at *VK00033*. Additional strains and features of FLASH transgenes are summarized in Tables 3.1, 3.2.

Table 3.1 Drosophila Strains

Strain	Source	Reference
PBac LL01602	DGRC: 140418	
Df (2R) 8057	Bloomington: 7871	
PBac c02047 (*Lsm11)*	Exelixis Harvard: c02047	(Godfrey et al., 2009)
His2aV^{810}	Bloomington: 9264	(van Daal and Elgin, 1992)
Mxc^{G46}	Bloomington: 32114	(White et al., 2011)
Mxc^{G48}	Bloomington: 7141	(White et al., 2011)
MCRS1 rnai	Vienna: v108017	
MBD-R2 rnai	Vienna: v110429	
ptc-Gal4	Bloomington: 2017	

Table 3.2 Construct Features

FLASH Transgene Feature	Sequence
attB-KpnI-promoter	5′-GATCTCTAGAGGTACCcgagattattgtttgtattgtgatgt tttttttaacgaatttaa-3′
V5-XhoI-AttB	5′-*GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACG* **TAA**CTCGAGCCGCGGCCGCAGAT-3′
5'UTR-EcorI-coding	5′-GAAGAAACGTAAGCGATTGGTGAATTCCAAA**ATG**-3′
EcorI-FL	5′-GAATTCCAAA**ATG**GAAACGCCTGCATATGCCAC-3′
EcorI- aa 65	5′-GAATTCCAAA**ATG**GACAGATCCCTTGAACTGGAC-3′
EcorI- aa 78	5′-GAATTCCAAA**ATG**GACGACTTTCAGAAGGCCGA-3′
aa 844-SacII-*V5*	5′-CTGCTGCgGACAAACCAATTCCGCGGTTCGAA *GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACG-3′*
aa 733-SacII-*V5*	5′-CTCACCCAGACTCCAAAACAGGCTCCGCGGTTCGGA *GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACG-3′*

The following features are annotated: engineered restriction sites are underlined; start and stop codons are bold; the FLASH promoter sequence is grey; the V5 tag is italicized.

Table 3.3 PCR Primers

Primer Name	Sequence
FLASH sequencing F	5′-gaacggtacccgagattattgtttgtattgtgatgt-3′
FLASH sequencing R	5′-gtcttccaaatccatcaagtcctc-3′
5' Pbac	5′-ccgataaaacacatgcgtca-3′
3' Pbac	5′-cgcgataaatctttctctctcg-3′

Sequences for primers used to identify the PBac LL01602. Template was prepared by incubating a squashed fly in squash buffer (10mM Tris pH8.2, 25mM NaCl, 1mM EDTA and 200ug/mL Proteinase K (NEB)) at 37 C for 30. After 10min at 85 C, 1 uL/reaction was used to amplify each product with TAQ DNA polymerase (Thermo). PCR products purified (Thermo) and subjected to Sanger Sequencing.

Immunofluorescence and Fluorescent in situ Hybridization

Salivary glands were dissected in PBS + 0.1% Triton X-100 and ovaries were dissected in Grace's medium (Gibco). *FISH* was performed as described in (White et al., 2007). Staining conditions, antibodies and probe sequences are summarized in Tables 3.4-6. Images were obtained on a Zeiss 510 Confocal Microscope and prepared with Lsm Software (Zeiss) and Photoshop (Adobe)

Table 3.4. Tissue Preparation for Antibody Staining for Immunofluorescence

Tissue	**Fix**	**Permeablization**
salivary gland	7% Formaldehyde, 20'	0.2% Tween, 15'
salivary gland (Fig. 3.6C)	3.7% Formaldehyde, 20'	
salivary gland (Fig. 3.8A)	3.7% Formaldehyde, 15'	1% Triton X-100, 15'
Ovary	7% Formaldehyde, 20'	0.2% Tween, 15'

Table 3.5. Antibody Concentrations

Primary Antibody	Raised In	Source	Concentration	Incubation
FLASH	Rabbit	(Yang et al., 2009)	1:2000	4 C, overnight
V5	Mouse	Invitrogen	1:1000	4 C, overnight
Mxc	Guinea Pig	(White et al., 2011)	1:2000	4 C, overnight
Mute	Guinea Pig	(Bulchand et al., 2010)	1:5000	4 C, overnight
Lsm10	Rabbit	(Liu et al., 2006)	1:2000	4 C, overnight
Lsm11	Rabbit	(Liu et al., 2006)	1:2000	4 C, overnight
HP1	Mouse	Developmental Studies Hybridoma Bank	1:1000	4 C, overnight
H2aV	Rabbit	(Leach et al., 2000)	1:1000	4 C, overnight
GFP	chicken	Millipore	1:1000	4 C, overnight
β-tubulin	rabbit	Abcam	1:5000	room temperature, 2h
Secondary Antibody	Recognizes	Source	Concentration	Incubation

Alexa-488	Rabbit	Invitrogen	1:2000	room temperature, 2h
Alexa-488	Mouse IgG2a	Invitrogen	1:1000	room temperature, 2h
Cy3	guinea pig	Jackson	1:1000	room temperature, 2h
Alexa-555	Mouse IgG1	Invitrogen	1:1000	room temperature, 2h
Cy5	rabbit	Abcam	1:2000	room temperature, 2h
Cy5	rabbit	Jackson	1:1000	room temperature, 2h
ECL donkey HRP	rabbit	GE Healthcare	1:10,000	room temperature, 1h
ECL donkey HRP	mouse	GE Healthcare	1:10,000	room temperature, 1h

Table 3.6. Flourescent *in situ* Hybridization Probe Sequences

Probe	Sequence
H3-coding probe 5'	5'-TAATTTAATAAATGTTGAGC-3'
H3-coding probe 3'	5'-GTTTGCTTGGTACGAGCCAT-3'
H3-ds probe 5'	5'-CGGTACTGGGTCTTAAATCA-3'
H3-ds probe 3'	5'-AGTGCTCTCCTCCTCGATTC-5'

Fluorescent *in situ* probes were synthesized by *in vitro* transcription (Ambion) using the DIG RNA labeling mix (Roche). The H3 coding probe was transcribed with Sp6 and the H3-ds probe was transcribed with T7 RNA polymerase. RNA synthesis was verified by agarose gel electrophoresis.

Ovaries were dissected and fixed in 4% Formaldehyde in PBS for 20 minutes and post fixed in 4% formaldehyde in PBS + 0.1% tween (PBT). Embryos were washed three times in PBT for two minutes and then incubated in 3ug/mL proteinase K in PBT for 10 minutes at room temperature followed by 30 minutes on ice. Digestion was quenched by incubation in 2ug/mL glycine/PBT for two minutes twice. After two PBT rinses, ovaries were post fixed for 20' in 4% formaldehyde. Ovaries were washed 5 times in PBT. This portion of the protocol was modified from(Lecuyer et al., 2008). Samples were then processed as described in (http://midline.bio.unc.edu/Files/InSituHybridiaztion Fluo.pdf) beginning with day 1 step 5.

Histone mRNA Analysis

5 mg of total cellular RNA extracted with Trizol Reagent (Invitrogen) was used for each S1 nuclease protection reaction. The probe was created by 5' end labeling BspEII cut H2a DNA with a-32P-dCTP and Klenow (New England Biolabs). The probe was gel purified and

hybridized to either total larval RNA or control yeast tRNA followed by digestion with S1 nuclease (Lanzotti et al., 2002). Protected fragments were resolved on a 6% polyacrylamide-7M urea gel and visualized by autoradiography.

Immunoblotting

Ovaries of the indicated genotypes were dissected in Grace's media (Gibco) and lysed in RIPA buffer (150mM NaCl, 1% Triton X-100, 50mM Tris pH 7.5, 0.1% SDS, 0.5% Na-deoxycholate and protease inhibitors (Thermo Scientific)). Samples were passed through a 25 gauge needle 100 times prior to resolving through a 10% gel by SDS-PAGE and detection with ECL Prime (Amersham) using antibodies summarized in supplemental methods.

In vitro processing assay

Extracts and *in vitro* processing assays were prepared and performed as described in (Sabath et al., 2013)

RNA Pol II Visualization

High-throughput sequencing reads were processed as previously described (McKay and Lieb, 2013). The following exceptions were made to unambiguously map reads to the histone locus. The exact number of histone gene repeats in *Drosophila* genome is unknown, and the reference genome sequence is thus incomplete. To circumvent this problem, a custom reference genome was created by removing all canonical histone gene repeat sequences, and by adding back a single 5kb histone gene repeat unit. An unlimited number of reads were then mapped to this custom genome with bowtie (Langmead et al., 2009) using the options '--nomaqround' and '--best'. Coverage values were then calculated for each base in the genome, and the data were visualized using IGV (Robinson et al., 2011). Datasets modENCODE_5122 and modENCODE_5569 were downloaded from the modENCODE ftp site (Kharchenko et al., 2011).

RESULTS

The Essential Function of *Drosophila* FLASH is Histone mRNA 3' End Formation

To determine whether concentrating factors in the HLB is necessary for histone pre-mRNA processing, we combined an *in vivo* structure function analysis of FLASH with careful monitoring of the integrity of the HLB. We established a transgenic complementation system to identify the protein domains necessary for FLASH function during *Drosophila* development (Fig. 3.1B). A piggyback transposon insertion in the 5'UTR of the *FLASH* gene (*FLASH*^{PBac}, Fig. 3.2) results in a large reduction of FLASH protein, as analyzed by western blotting (Fig. 3.8D). This hypomorphic allele in *trans* to a deficiency of *FLASH (FLASH*^{Df}*)* results in nearly complete lethality, and the small number (7% of expected) of adult flies that eclose are female sterile but male fertile (Fig. 3.1D, Table 3.7). Expression of V5-tagged full-length FLASH (FLASH^{FL}; Fig. 3.1B) from a single copy transgene using the endogenous *FLASH* promoter rescued both lethality and sterility caused by *FLASH*$^{PBac/Df}$ (Fig. 3.1D, Table 3.7). This system provides a means for testing the ability of transgenes expressing different alleles of FLASH to rescue FLASH mutant phenotypes.

Because FLASH was originally implicated in Fas-mediated apoptosis as an activator of caspase-8 in mammalian cells (Imai et al., 1999), we asked if the developmental phenotypes of the *Drosophila FLASH*$^{PBac/Df}$ mutant were due to histone pre-mRNA processing defects or loss of another FLASH function. The 844 amino acid (aa) FLASH protein has three known biochemical functions. First, the N-terminal 154 aa of FLASH is essential for histone pre-mRNA processing *in vitro*, and contains a region that binds the N-terminus of the U7 snRNP protein Lsm11 (Burch et al., 2011). Second, the Lsm11/FLASH N-terminal interaction creates a distinct interface that recruits the HCC (Sabath et al., 2013; Yang et al., 2013) a complex of proteins that contains the endonuclease CPSF-73, which cleaves histone pre-mRNA (Dominski et al., 2005). Third, the C-terminal 111 aa of FLASH is necessary for localization to the HLB (Burch et al., 2011). To test if these three activities are sufficient for FLASH function *in vivo*, we expressed a transgene with a

547 aa internal deletion of the 844 aa FLASH protein, fusing the HCC and Lsm11 binding domains to the HLB localization domain ($FLASH^{mini}$, Fig. 3.1B). This transgene fully rescued the lethality and sterility of the $FLASH^{PBac/Df}$ mutant (Fig. 3.1D, Table 3.7), suggesting that the essential function of FLASH in *Drosophila* is histone mRNA 3' end formation.

To understand the molecular basis for the FLASH mutant phenotype, we examined the histone mRNA species produced in various *FLASH* mutant genotypes (Fig. 3.1E). When histone pre-mRNA processing is compromised in *Drosophila* (e.g. by mutation of *Slbp* or U7 snRNP components), cryptic polyadenylation signals located downstream of the normal processing site are utilized, resulting in the cytoplasmic accumulation of longer, polyadenylated histone mRNA (Godfrey et al., 2006; Lanzotti et al., 2002; Sullivan et al., 2001). To detect these RNA species, we used an S1 protection assay, which provides a quantitative measurement of the amounts of properly processed histone mRNA (Fig. 3.1E, "W") compared to misprocessed, polyadenylated histone mRNA (Fig. 3.1E, "M"). The S1 assay also detects "read through" transcripts that represent a mixture of unprocessed, nascent pre-mRNA and poly A^+ RNAs that extend beyond the complementary region of the probe (Fig. 3.1E, "R") (Sullivan et al., 2009) .

Using the S1 assay, we found that the $FLASH^{PBac/Df}$ mutant expresses large amounts of misprocessed, polyadenylated histone mRNA (Fig. 3.1D, lane 2), agreeing with a previous observation using northern blotting (Rajendra et al., 2010). We additionally detected small amounts of properly processed and read through RNA (Fig. 3.1E, lane 2). Expression of $FLASH^{FL}$ protein completely rescued this mutant RNA phenotype (Fig. 3.1E, lane 3) and only low amounts of both misprocessed and read through histone mRNA were observed in $FLASH^{mini}$ larvae, indicating that $FLASH^{mini}$ is almost fully proficient in processing *in vivo* (Fig. 3.1E, lane 4). We next made point mutations in FLASH to test whether disrupting histone pre-mRNA processing results in lethality, as suggested by our previous analyses of *Slbp* and U7 snRNP mutants (Godfrey et al., 2006; Sullivan et al., 2001). We created a mutation in FLASH ($FLASH^{NL125}$) that abolishes Lsm11 binding by changing two conserved NL residues (aa 125-

126; Fig. 3.3A) to alanines (Burch et al., 2011). FLASH^{NL125} did not rescue *FLASH*$^{PBac/Df}$ lethality or the mis-processed histone mRNA phenotype (Fig. 3.1D, E, lane 5, Table 3.7). We previously showed that a two-amino acid mutation of Lsm11 that prevents FLASH binding also results in the accumulation of misprocessed histone mRNA and lethality (Burch et al., 2011). Together these data indicate that the central portion of FLASH does not perform any vital developmental functions, and that the interaction between FLASH and Lsm11 is essential for proper histone mRNA 3' end formation and normal fly development.

Figure 3.1

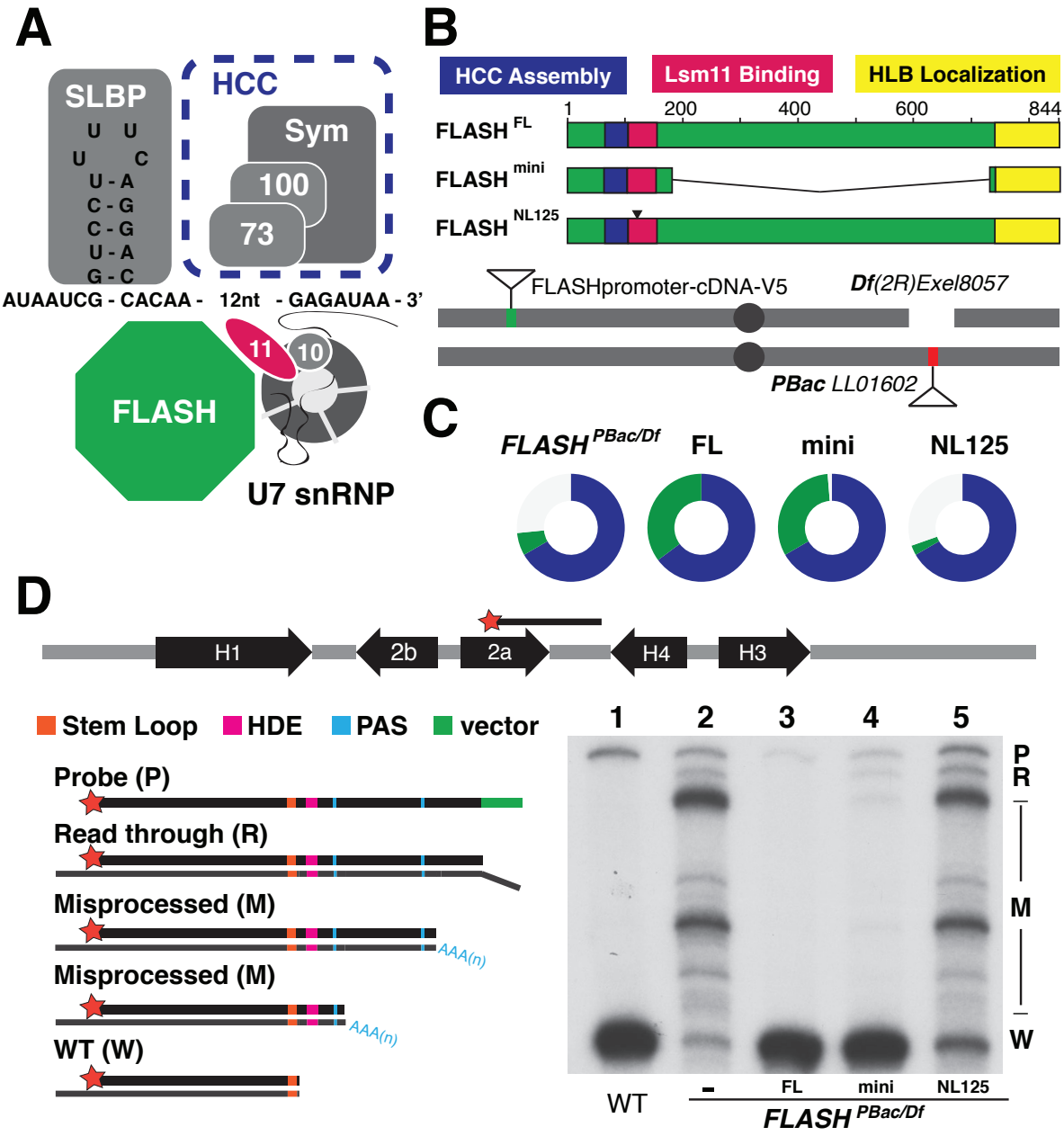


Figure 3.1. Histone pre-mRNA processing is the essential *in vivo* FLASH function.
A. Diagram of the histone pre-mRNA processing machinery. FLASH directly binds Lsm11 (11) of the U7snRNP, which interacts with the HDE (GAGAUAA) in histone pre-mRNA. FLASH and U7snRNP recruit a histone cleavage complex (HCC) containing the Symplekin scaffold (Sym) and the CPSF100/CPSF73 endonuclease (100/73). 10: Lsm10. SLBP: Stem Loop Binding Protein.
B. Functional domains of FLASH and schematic of the transgene rescue system. Transgenes were generated at chromosomal location 25C6 by phiC31-mediated integration and recombined with a deletion of the *FLASH* locus (*Df(2R)Exel8057*). All *FLASH*$^{PBac/Df}$ experimental flies

contained a maternal ***Df*** chromosome in *trans* to a paternal ***PBac*** *LL01602* chromosome. NL125 is a 2 amino acid mutation in the Lsm11 binding region (see Fig. 2A).
C. Visual representation of transgenic rescue of *FLASH*$^{PBac/Df}$. Circles indicate the proportion of *FLASH*$^{PBac/Df}$ mutant (green) and control heterozygous sibling (blue) adult flies obtained in each experiment. The expected fraction of control siblings is 2/3. Thus, 1/3 green indicates full viability and the absence of green indicates lethality.
D. S1 nuclease protection assay. The diagram at top indicates the arrangement of a single *Drosophila* histone gene repeat. Total RNA extracted from wandering 3rd instar larvae was hybridized with a 5'-labeled DNA probe complementary to *H2a* mRNA that extends to the *H4* HDE and contains a small amount of vector sequence (green), and then incubated with S1 nuclease, which digests single stranded nucleic acid. H2a pre-mRNA is normally cleaved between the stem loop (orange) and HDE (pink) resulting in a 340bp fragment that is protected from S1 digestion (W). The *H2a* gene contains multiple downstream polyadenylation signals (PAS, blue) that are variably utilized when normal processing is disrupted, resulting in a heterogeneous collection of longer protected fragments (M). Any transcripts extending beyond the complementary region result in a single protected "read through" fragment (R) that is distinct from undigested probe (P). Protected probe fragments were resolved on an acrylamide urea gel and visualized by autoradiography.

Figure 3.2

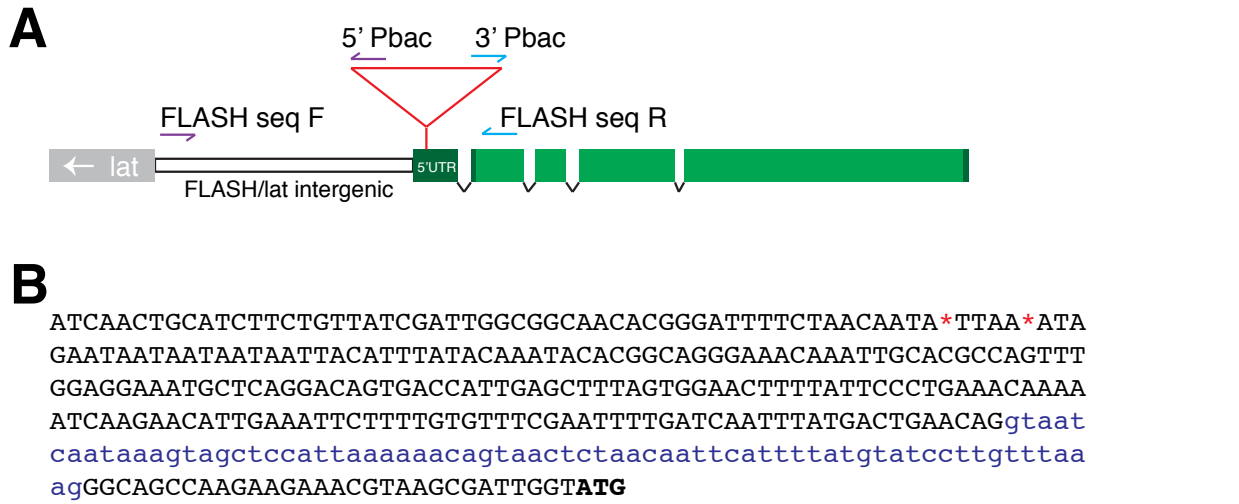


Figure 3.2. Location of *FLASH*^{LL01602} PBac insertion.
A. Diagram of the FLASH locus (exons green) indicating the location of PCR primers used to identify the exact location of PBac LL01602.
B. Sequence of the *FLASH* 5' UTR indicating the molecularly confirmed insertion site (red stars). An intron present in the 5'UTR is indicated in lower case blue letters.

Table 3.7

Genotype	Control Cyo Sibling	Mutant genotype	Total	Mendelian? (Chi-Square p = 0.05)	Significantly different from mutant? (t-test)	Female Fertile	Male Fertile
* $FLASH^{Df}$/cyo, twi-GFP (maternal) x $FLASH^{PBac}$/cyo, twi-GFP (paternal)	610	62	672	N		N	Y
$FLASH^{PBac}$/cyo, twi-GFP (maternal) x $FLASH^{Df}$/cyo, twi-GFP (paternal)	509	213	722	N	Y p < 0.001	N	Y
$FLASH^{PBac}$/cyo, twi-GFP (maternal) x $FLASH^{Df}$/cyo, twi-GFP (paternal)	450	84	534	N	Y p < 0.05	N	Y
* FL	606	329	935	Y	n/a	Y	Y
* mini	443	213	656	Y	n/a	Y	Y
* NL125	607	27	634	N	Y p < 0.05	N	n/d
* 78-844	608	0	608	N	Y p < 0.001	n/a	n/a
* 65-844	596	294	890	Y	n/a	Y	Y
* LDIY 71	387	164	551	Y	n/a	Y	Y
* LDIY 45	381	206	587	Y	n/a	Y	Y
* LDIY 45,71	621	7	628	N	Y p < 0.001	N	n/d
* 1-733	338	189	527	Y	n/a	Y	Y
# $FLASH^{Df}$,LDIY 71, 1-733/cyo, tb-RFP x $FLASH^{PBac}$/cyo, twi-GFP	267	59	326	N	N	N	n/d
mxc^{G46}/fm7c, twi-GFP x mxc^{G46}/Y	276	274	550	Y	n/a	N	Y

Table 3.7. Quantification of transgenic rescue of the *$FLASH^{PBac/Df}$* mutant.
Total flies scored for each genetic rescue experiment presented as circle charts in Figure 1-3. For each experiment a chi square analysis was performed under the null hypothesis that the phenotypic classes of adult flies were present in expected Mendelian ratios (2/3 cyo sibling, 1/3 mutant genotype). For the subset of transgenes that failed to rescue *$FLASH^{PBac/Df}$*, a T test was used to determine whether the numbers of flies in the experiment class was statistically different than when no transgene was present in the *$FLASH^{PBac/Df}$* background. Assessment of fertility for each genotype is also noted. Symbols: * parental genotypes for each transgenic rescue cross. # $FLASH^{LDIY71,1-733}$/cyo, twi-GFP was lethal. Therefore the parental line was carried over the indicated balancer chromosome and only one parental genotype was recovered in the control cyo sibling class.

The FLASH NH_2-terminus is Necessary for Histone mRNA 3' End Formation *in vivo*

Our previous *in vitro* experiments indicated that a domain within the first 77 amino acids of FLASH that is distinct from the Lsm11 binding domain is essential for histone pre-mRNA processing *in vitro* (Sabath et al., 2013; Yang et al., 2013). In both mammalian and *Drosophila* nuclear extracts, this second domain is necessary for recruitment of the HCC to the FLASH/Lsm11 complex. Deletion of residues 1-77 in *Drosophila* FLASH or mutation of a highly conserved LDIY motif (aa 71-74; Fig. 3.3A), to AAAA greatly reduced the ability of FLASH to process synthetic histone pre-mRNA substrates and to recruit the HCC in nuclear extracts

(Sabath et al., 2013). Additionally, amino acids 65-77 were required for histone 3' end formation in cultured cells (Burch et al., 2011).

To test the requirement for this region of FLASH during fly development, we created a series of FLASH mutants and assayed their ability to rescue *FLASH*$^{PBac/Df}$ mutant phenotypes (Fig. 3.3B). Consistent with the *in vitro* and cell culture data, expression of an N-terminal truncation containing amino acids 71-74 (FLASH$^{65-844}$) fully rescued *FLASH*$^{PBac/Df}$ lethality whereas a mutant lacking these amino acids (FLASH$^{78-844}$) did not (Fig. 3.3C, Table 3.7). To our surprise, the FLASH^{LDIY71} mutation in the context of full-length FLASH rescued *FLASH*$^{PBac/Df}$ lethality. Consistent with this finding, FLASH^{LDIY71} mutants contained large amounts of histone mRNA with a normal 3' end, although we could detect small amounts of misprocessed and read through histone mRNA species (Fig. 3.3C; 3.3D, lane 6; Fig. 3.4A and Table 3.7). We conclude that the FLASH^{LDIY71} protein retains most of its histone pre-mRNA processing activity *in vivo*, in contrast to its activity *in vitro*. In addition, the difference in phenotypes between the FLASH$^{78-844}$ truncation mutant and the FLASH^{LDIY71} mutant suggest a role for additional N-terminal FLASH residues in histone pre-mRNA processing *in vivo*.

We noted a second LDIY motif located about 30 amino acids before the HCC recruitment site that is conserved in vertebrates (Fig. 3.3A). This motif is located at amino acids 45-48 in *Drosophila* FLASH, and is not essential since the FLASH$^{65-844}$ transgene completely rescues viability and histone pre-mRNA processing (Fig. 3.3C; 3.3D, lane 5). We mutated these four amino acids to alanine and generated transgenes expressing this mutation alone (FLASH^{LDIY45}) or a protein with combined mutation of the two LDIY motifs (FLASH^{LDIY45,71}). The FLASH^{LDIY45} protein rescued *FLASH*$^{PBac/Df}$ lethality and completely restored histone pre-mRNA processing (Fig. 3.3D lane 5, 3.3E, Table 3.7). However, FLASH^{LDIY45,71} did not function in histone pre-mRNA processing *in vivo* and failed to rescue *FLASH*$^{PBac/Df}$ lethality (Fig. 3.3D lanes 4 and 8, E, Table 3.7), a phenotype that was indistinguishable from the FLASH$^{78-844}$ mutant.

To confirm that FLASH^{LDIY45,71} severely impaired the biochemical role of FLASH in histone pre-mRNA processing, we performed an *in vitro* processing assay using nuclear extracts from cultured S2 cells in which FLASH was depleted by dsRNA. These extracts are deficient in processing synthetic histone pre-mRNA substrates, but this deficiency can be biochemically complemented by addition of *E. coli*-derived proteins containing the N-terminal 178 residues of FLASH (Fig. 3.3F, compare lanes 2,3 to 4,5) (Sabath et al., 2013). Consistent with our previous study, addition of a FLASH^{LDIY71} N-terminal fragment increased processing, but less than addition of a wild type control (Fig. 3.3F, compare lanes 6,7 to 8,9). This slight stimulation of processing was abolished when LDIY45 was also mutated (Fig. 3.3F,lanes 10 and 11). These data indicate that both the LDIY71 motif and the LDIY45 motif contribute to processing.

Figure 3.3

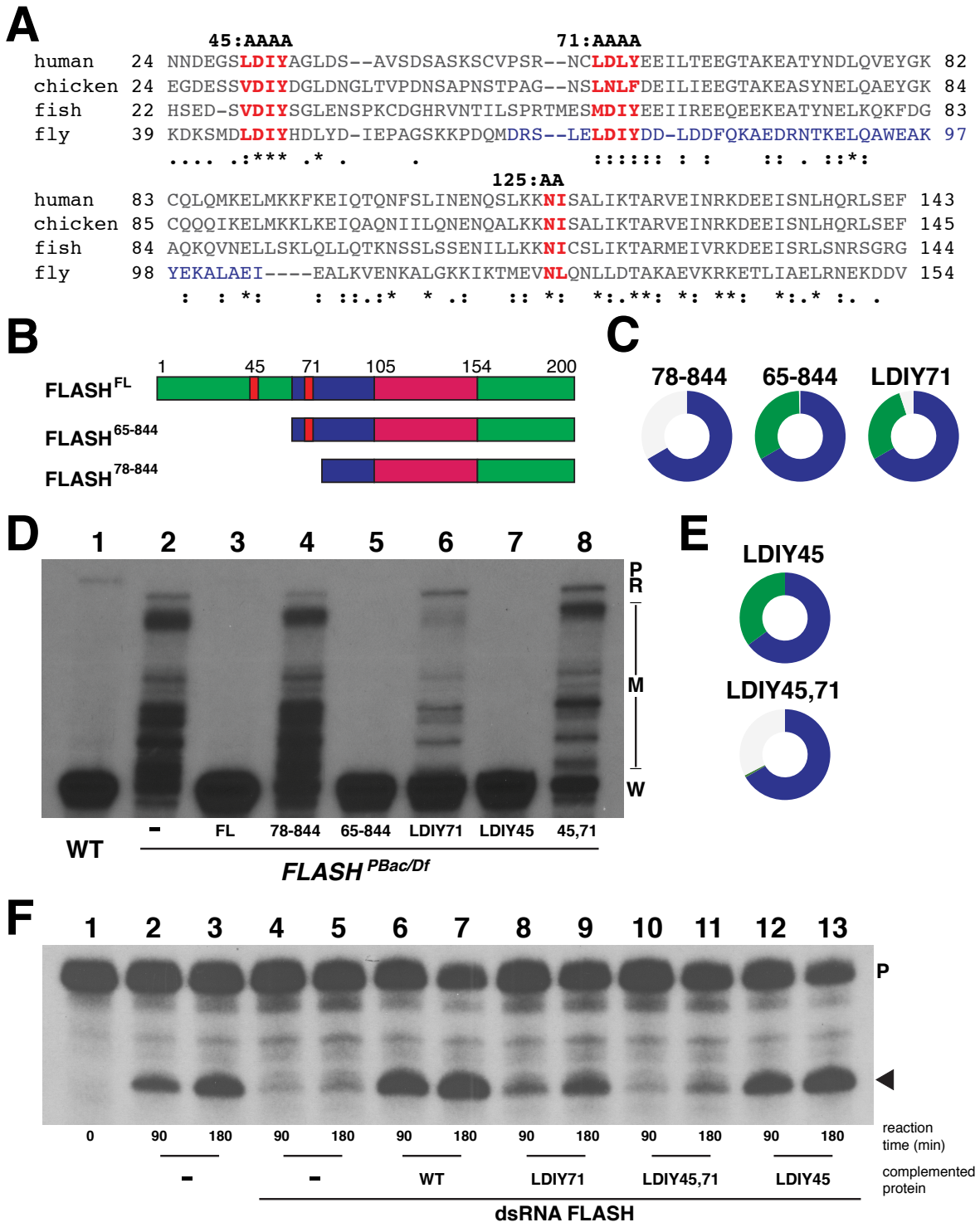

A
45:AAAA
71:AAAA
human 24 NNDEGSLDIYAGLDS--AVSDSASKSCVPSR--NCLDLYEEILTEEGTAKEATYNDLQVEYGK 82
chicken 24 EGDESSVDIYDGLDNGLTVPDNSAPNSTPAG--NSLNLFDEILIEEGTAKEASYNELQAEYGK 84
fish 22 HSED-SVDIYSGLENSPKCDGHRVNTILSPRTMESMDIYEEIIREEQEEKEATYNELKQKFDG 83
fly 39 KDKSMDLDIYHDLYD-IEPAGSKKPDQMDRS--LELDIYDD-LDDFQKAEDRNTKELQAWEAK 97
125:AA
human 83 CQLQMKELMKKFKEIQTQNFSLINENQSLKKNISALIKTARVEINRKDEEISNLHQRLSEF 143
chicken 85 CQQQIKELMKKLKEIQAQNIILQNENQALKKNISALIKTARVEINRKDEEISNLHQRLSEF 145
fish 84 AQKQVNELLSKLQLLQTKNSSLSSENILLKKNICSLIKTARMEIVRKDEEISRLSNRSGRG 144
fly 98 YEKALAEI----EALKVENKALGKKIKTMEVNLQNLLDTAKAEVKRKETLIAELRNEKDDV 154
B
FLASH FL
FLASH 65-844
FLASH 78-844
1 45 71 105 154 200
C
78-844
65-844
LDIY71
D
1 2 3 4 5 6 7 8
P
R
M
W
WT
- FL 78-844 65-844 LDIY71 LDIY45 45,71
FLASH PBac/Df
E
LDIY45
LDIY45,71
F
1 2 3 4 5 6 7 8 9 10 11 12 13
P
0 90 180 90 180 90 180 90 180 90 180 90 180
reaction time (min)
- - WT LDIY71 LDIY45,71 LDIY45
complemented protein
dsRNA FLASH

Figure 3.3 Characterization of FLASH N-terminus function in histone pre-mRNA processing.
A. Alignment of FLASH N-terminal sequences indicating the *Drosophila* HCC (blue text), the conserved LDIY motifs (red; mutated to AAAA in FLASH^{LDIY45} and FLASH^{LDIY71}) and the NL residues (red) of the Lsm11 binding region mutated to AA in FLASH^{NL125}.
B. Diagram of the N-terminal FLASH deletion mutants.
C. Visualization of rescue of *FLASH*$^{PBac/Df}$ with transgenic FLASH N-terminal mutants.
D. Larval H2a mRNA species for the indicated genotypes assessed by S1 nuclease protection. A shorter, phospho-exposure is presented in Fig. S2.
E. Visualization of rescue of *FLASH*$^{PBac/Df}$ with transgenic FLASH LDIY mutants.
F. Biochemical activity of FLASH LDIY mutations was assessed with an *in vitro* histone processing assay. FLASH depleted nuclear extracts were complemented with the indicated recombinant N-terminal FLASH protein (aa 1-178). ^{32}P-labeled synthetic *H3* pre-mRNA was incubated with the extract for 90 or 180 min and the resulting undigested probe (P) and cleaved probe (arrowhead) were resolved on an acrylamide urea gel. See also Fig. S2.

Figure 3.4

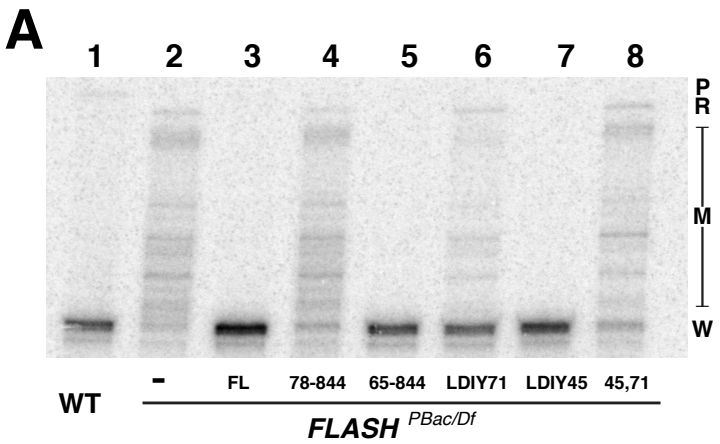


Figure 3.4. Characterization of FLASH N-terminus contribution to histone pre-mRNA processing.
A. Phosphorimager scan of the S1 protection assay shown in Figure 2. The mis-processed histone H2a RNA species are better resolved in this exposure.

Concentrating FLASH in the HLB Facilitates Histone mRNA 3' End Formation

The ability of the FLASH^{LDIY71} mutant to provide full function *in vivo* compared with its reduced activity *in vitro* indicates that some aspect of FLASH^{LDIY71} is compensated for in an intact cell. We reasoned that the HLB may provide a critical contribution to histone 3' end formation *in vivo* that cannot be recapitulated using biochemical assays of nuclear extracts. We therefore tested

the consequences of preventing the accumulation of FLASH in the HLB, to determine if a high local concentration of processing factors contributes to their function.

A mutant of FLASH lacking the C-terminal 111 amino acids ($FLASH^{1-733}$) does not localize to the HLB in cultured S2 cells (Burch et al., 2011). $FLASH^{1-733}$ also fails to localize to the HLB *in vivo* (Fig. 3.5A, 3.5C), yet it fully rescued $FLASH^{PBac/Df}$ lethality (Fig. 3.5B). Although $FLASH^{1-733}$ flies developed to adulthood, we detected both mis-processed and read through transcripts in these animals, indicating that concentration of FLASH in the HLB is required for full FLASH function *in vivo* (Fig. 3.5D lane 4). However, unlike situations in which FLASH is severely depleted (e.g. $FLASH^{PBac/Df}$) or when a FLASH protein that is biochemically deficient in processing (e.g. $FLASH^{LDIY45,71}$) was expressed in the mutant background, the read-through transcripts in the $FLASH^{1-733}$ mutant were present in similar amounts as the misprocessed RNAs (Fig. 3.5D, compare R to M in lanes 2 and 4). The profile of RNA species in $FLASH^{1-733}$ animals is very similar to that in $FLASH^{LDIY71}$ animals, which also have similar amounts of read-through transcripts and misprocessed RNAs (Fig. 3.5D, lanes 3 and 4). Thus, these two FLASH mutants with very different biochemical defects resulted in a similar alteration in histone mRNA biosynthesis.

Since both localization and processing activity might independently contribute to efficient pre-mRNA processing, mislocalizing the partially functional $FLASH^{LDIY71}$ protein could exacerbate the mutant phenotype. We generated a transgene expressing a FLASH protein containing both the LDIY71 mutation and the C-terminal truncation ($FLASH^{LDIY71,1-733}$). $FLASH^{LDIY71,1-733}$ did not localize to the HLB (Fig. 3.5C), and it also failed to rescue $FLASH^{Pac/Df}$ lethality (Fig. 3.5B) and the histone pre-mRNA processing defect (Fig. 3.5D, lane 5). Thus, the $FLASH^{LDIY71}$ protein is only effective in promoting histone pre-mRNA processing *in vivo* if it is present in the HLB (Fig 3.5C,D). These findings provide direct evidence that concentrating factors in the HLB is important for histone mRNA 3' end formation.

This observation suggests that the efficiency of 3' end formation is sensitive to the amount of locally available FLASH activity. To test this, we analyzed histone mRNA at the onset of the *FLASH*$^{Df/Df}$ mutant phenotype. Maternally deposited wild type FLASH protein accumulates at the histone locus before activation of zygotic transcription (White et al., 2011). Therefore, zygotic *FLASH* mutant phenotypes will not be apparent until multiple rounds of cell division deplete maternal FLASH protein. This "maternal run out" situation provides an opportunity to examine embryos with reduced concentrations of FLASH in the HLB, but before FLASH is completely depleted. S1 nuclease protection assays of RNA from stage 17 (14-18h) *FLASH* mutant embryos showed that these embryos had both misprocessed and read-through histone mRNAs in similar amounts, the same molecular phenotype as the FLASH$^{1-733}$ mutant (Figure 3.5D, E). Thus the initial phenotype as maternal FLASH is depleted in the *FLASH* mutant is similar to the effect of reducing FLASH activity and likely results from lowering the local concentration of FLASH in the HLB, supporting our conclusion that concentrating factors in the HLB is important for histone mRNA 3' end formation.

Figure 3.5

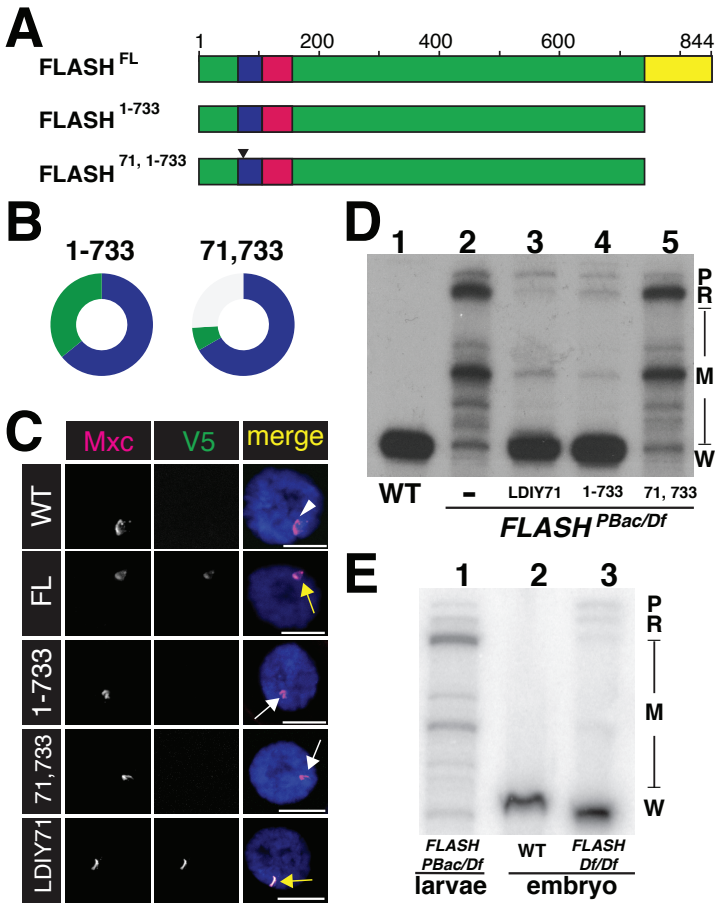


Figure 3.5 Concentrating FLASH in the HLB promotes histone 3' end formation
A. Diagram of C-terminal FLASH deletion mutants.
B. Visualization of rescue of *FLASH*$^{PBac/Df}$ with transgenic FLASH C-terminal mutants.
C. Localization of FLASH transgenic proteins (V5) in larval polyploid salivary gland nuclei stained for an HLB marker (Mxc; white arrowhead). Wild type transgenic FLASH protein localizes to the HLB (yellow arrow) but the C-terminal truncation mutants do not (white arrows). Scale bars = 10 μm.
D,E. H2a mRNA species for the indicated genotypes were assessed by S1 nuclease protection in larvae **(D)** and embryo **(E)**.

High Concentrations of FLASH are Necessary for U7 snRNP Accumulation in the HLB

The HLB forms in a stepwise manner, and FLASH is one of the first components recruited during this process (Salzler et al., 2013; White et al., 2011). To determine whether some of the effects of mutating FLASH might be mediated through changes in HLB composition, we visualized HLBs in different FLASH mutant animals with antibodies against a panel of HLB

markers (Fig. 3.6, Fig. 3.7) (White et al., 2011). We analyzed ovaries of the small number of *FLASH*$^{PBac/Df}$ mutant females that survive to adulthood. Although these females are sterile and we do not observe fully developed eggs, oogenesis consistently proceeds to stage 8 allowing us to examine stage 5 polyploid nurse cells, which provide excellent cytology for HLBs (Liu et al., 2006). We did not detect FLASH in the HLBs of *FLASH*$^{PBac/Df}$ mutant ovaries, although other components of the HLB were present (Fig. 3.6A). We also failed to detect the U7 snRNP-specific protein Lsm10 in the HLBs of *FLASH*$^{PBac/Df}$ mutants (Fig. 3.6A), consistent with a previous observation (Rajendra et al., 2010). Expressing FLASH^{FL} protein in the *FLASH*$^{PBac/Df}$ mutant restored Lsm10 localization to the HLB (Fig. 3.6B). In contrast, expressing the FLASH$^{1-733}$ mutant protein did not restore HLB accumulation of Lsm10 (Fig. 3.6B). These data suggest that high concentrations of FLASH in the HLB are needed for normal U7 snRNP accumulation in the HLB.

To further examine whether concentrating FLASH in the HLB is required to recruit U7 snRNP to the HLB, we analyzed a mutant allele of the *multi sex combs* (*mxc*) gene. *mxc* encodes a 1,837 aa protein similar to human NPAT that is concentrated in the HLB and likely acts as a scaffold for HLB assembly (White et al., 2011). *mxc*^{G46} encodes a C-terminal truncation of Mxc protein lacking aa 1643-1837. Because our anti-Mxc antibody (raised against the last 200 aa of Mxc) does not detect Mxc^{G46} protein (Fig. 3.6C), we stained salivary glands of transgenic GFP-Mxc^{G46} rescued animals with anti-GFP antibodies to confirm that Mxc^{G46} protein supports HLB assembly (Fig. 3.6C). FLASH does not localize to the HLB in *mxc*^{G46} mutants (Fig. 3.6C) (Rajendra et al., 2010), indicating that the C-terminal 194 aa of Mxc are necessary for FLASH recruitment to the HLB. Consistent with our analysis of the mis-localized FLASH$^{1-733}$ protein, Lsm10 also does not accumulate in the HLBs of the *mxc*^{G46} mutant (Fig. 3.6C). These results support our conclusion that concentrating FLASH in the HLB is necessary to concentrate U7 snRNP in the HLB.

Although localization of U7 snRNP to the HLB is dependent on FLASH, its accumulation does not rely on the direct interaction between FLASH and Lsm11 required for histone mRNA processing: Expression of the FLASH^{NL125} protein, which does not bind to Lsm11 or support histone pre-mRNA processing *in vivo* (Fig. 3.1), rescued U7 snRNP localization to the HLB (Fig. 3.6B). Previously we showed that a mutation in Lsm11, which blocked processing by interfering with FLASH binding, was also recruited to the HLB (Burch et al., 2011). Taken together, these data suggest that FLASH functions within the HLB in at least two distinct ways: (1) by helping recruit and concentrate U7 snRNP in the HLB and (2) by interacting directly with the Lsm11 protein of U7 snRNP during histone pre-mRNA processing.

Figure 3.6

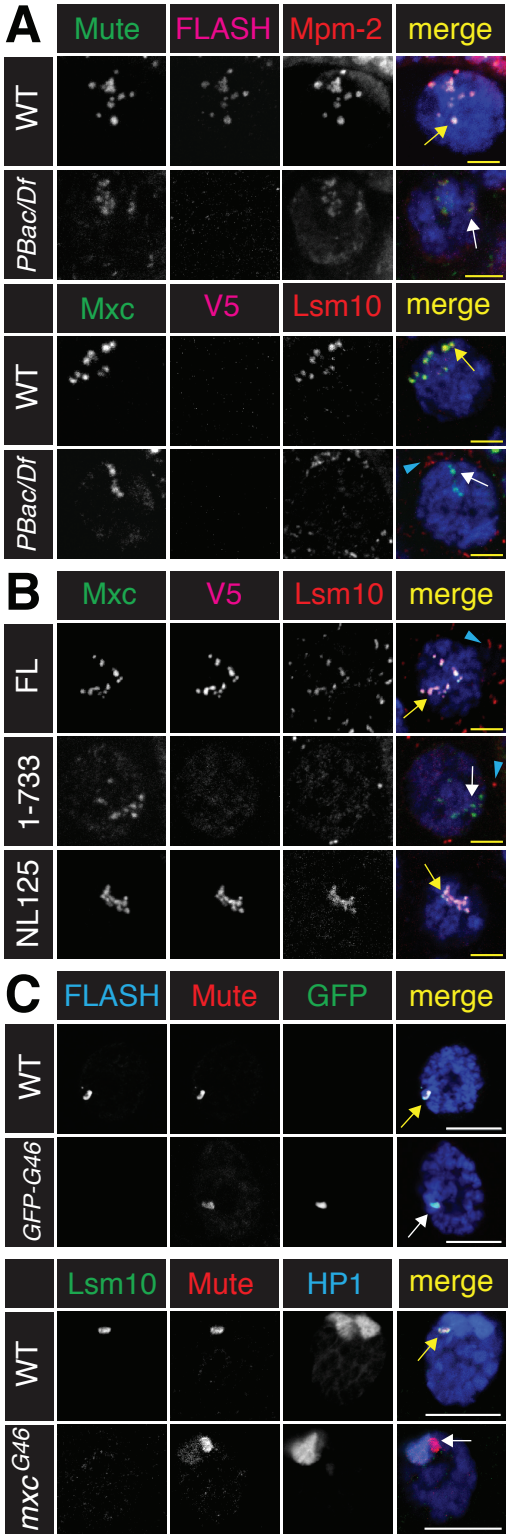

A
Mute
FLASH
Mpm-2
merge
WT
PBac/Df
Mxc
V5
Lsm10
merge
WT
PBac/Df
B
Mxc
V5
Lsm10
merge
FL
1-733
NL125
C
FLASH
Mute
GFP
merge
WT
GFP-G46
Lsm10
Mute
HP1
merge
WT
mxc G46

Figure 3.6. FLASH contribution to HLB assembly.
A, B. HLB assembly was assessed in nurse cells of the indicated genotypes at stage 5 or 6 of oogenesis with antibodies against FLASH, Lsm10, Mute, a constitutive HLB factor of unknown function (Bulchand et al., 2010), Mxc, the putative HLB scaffold (the human NPAT orthologue), as well as with MPM-2 monoclonal antibodies, which detects Cyclin E/Cdk2-dependent phosphorylation of Mxc in S-phase nuclei (White et al., 2011). Note that at these stages of oogenesis the chromatids of the polyploid nurse cells are dispersed, resulting in many HLB foci in close proximity. Yellow arrows indicate proper localization and white arrows indicate mis-localization. Blue arrowheads indicate signal from the Lsm10 antibody that is not present in the HLB; the source of this signal is unknown. Scale bars = 5 μm.
C. HLB assembly in polyploid salivary gland nuclei of mxc^{G46} mutants or mxc^{G48} null mutants expressing transgenic Mxc^{G46} protein (GFP-G46), detected with anti-GFP antibodies. Note that salivary gland chromosomes are polytene, resulting in a single, large HLB. HP1 staining visualizes the chromocenter located near the histone locus. Scale bars = 15 μm.

Figure 3.7

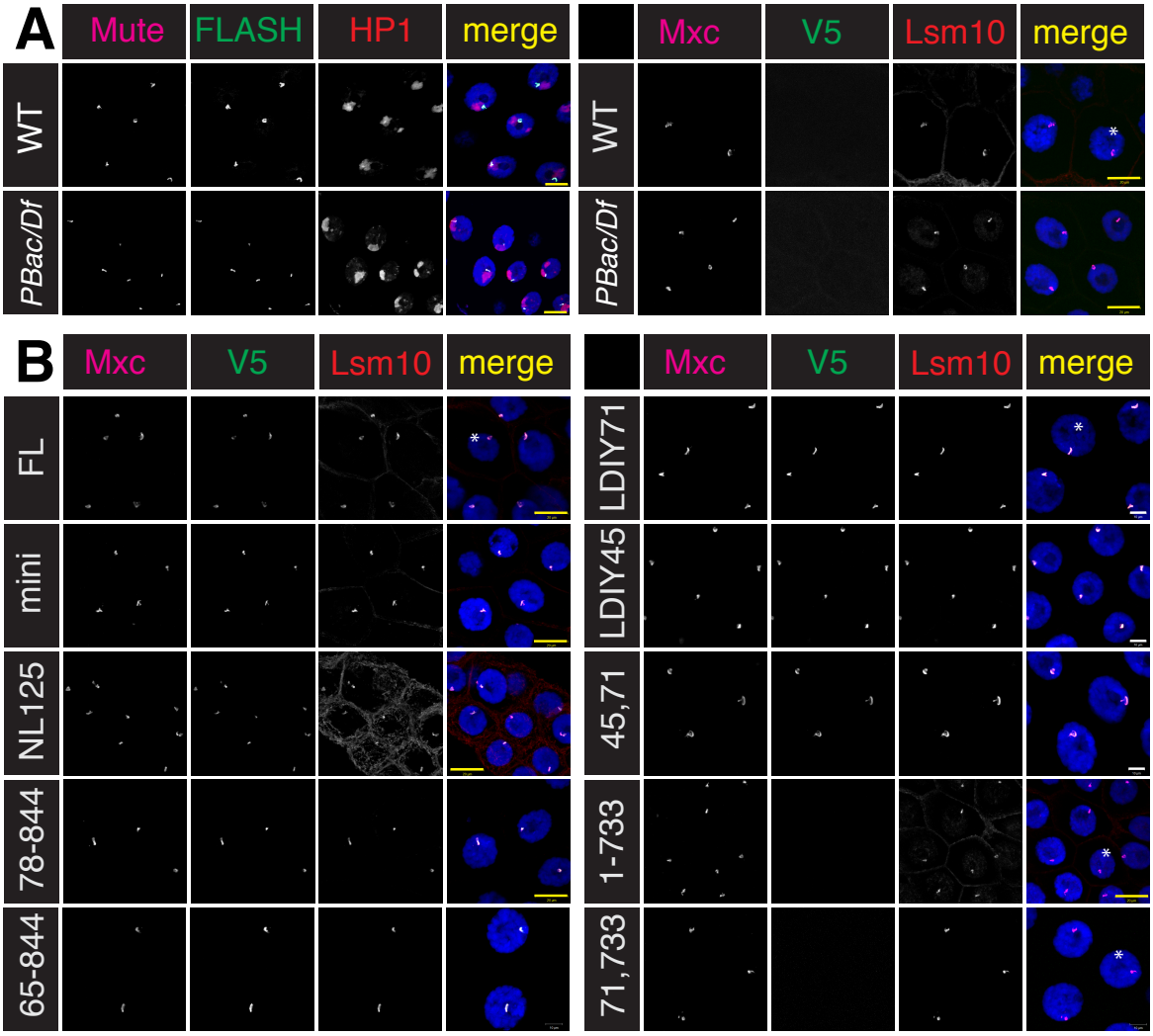


Figure 3.7. Localization of transgenic FLASH proteins.
A. Wild type and *FLASH*$^{PBac/Df}$ wandering 3rd instar larval salivary glands stained with a panel of HLB markers. Unlike in *FLASH*$^{PBac/Df}$ mutant ovaries, endogenous FLASH can be detected in the HLB, likely due to the presence of a stable, maternal pool of FLASH in this tissue. As a consequence, Lsm10 is also present in the FLASH positive HLBs.
B. Localization of FLASH transgenic proteins (V5) in larval salivary glands. Only proteins lacking the C-terminal 111 aa of FLASH fail to localize to the HLB. Asterisks indicate cells that were cropped and presented in Figure 3C. Yellow scale bars = 20 μm. White scale bars = 10 μm.

FLASH is not Sufficient for U7 snRNP Accumulation in the HLB

We previously identified the variant histone H2aV in a genome-wide RNAi screen for genes required for histone pre-mRNA processing (Wagner et al., 2007). We found that U7 snRNP was not localized to the HLB in H2aV depleted S2 cells. These cells accumulate misprocessed histone mRNA, although nuclear extracts prepared from these cells were as active as wild type in histone pre-mRNA processing (Wagner et al., 2007). We tested whether loss of H2aV affects the composition of the HLB. Because *H2aV* is an essential gene, for this analysis we examined salivary glands from third instar $H2aV^{810}$ null mutant larvae. FLASH and two other HLB markers, Mxc and Mute, co-localize in a prominent focus in $H2aV^{810}$ mutant salivary glands, indicating that H2aV is not required for HLB assembly. Lsm11 was absent from these HLBs as shown previously (Fig. 3.8A (Wagner et al., 2007)). We observed an identical result after salivary gland-specific RNAi-depletion of MCRS1 or MBD-R2 (Fig. 3.8A), two proteins that are part of multiple histone acetyl transferase complexes (Mendjan et al., 2006; Prestel et al., 2010; Raja et al., 2010) that we also identified in the genome-wide RNAi screen for histone pre-mRNA processing (Wagner et al., 2007). Thus, depletion of H2aV, MCRS1, or MBD-R2 results in failure to accumulate U7 snRNP in the HLB. Whether this is a direct or indirect effect on U7 snRNP or the HLB is not known. These results demonstrate that the presence of FLASH in the HLB, while necessary, is not sufficient for U7 snRNP localization.

Figure 3.8

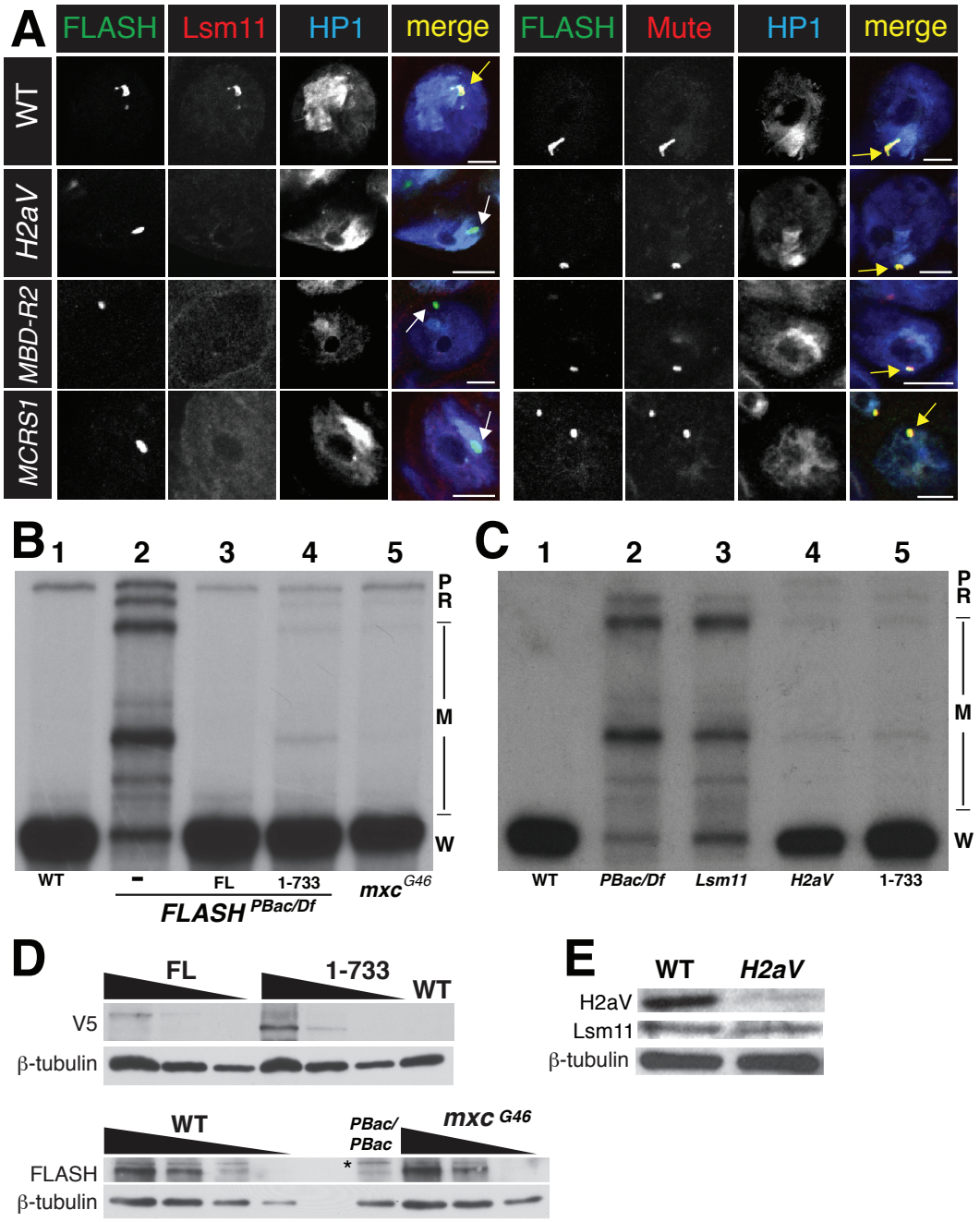


Figure 3.8. HLB concentration of FLASH and/or U7snRNP promotes efficient histone pre-mRNA processing.

A. Larval salivary glands expressing dsRNA with ptc-Gal4 and targeting MCRS1 or MBD-R2 or from *H2aV*^{810} mutants were stained for HLB markers and HP1. Arrows as in Figs. 2 and 3. Scale bars = 10 μm.

B,C. S1 nuclease protection analysis of H2a RNA from ovaries (B) or wandering 3rd instar larvae (C) of the indicated genotypes.

D. Anti-V5 or anti-FLASH western analysis of $FLASH^{FL}$ and $FLASH^{1-733}$ rescue ovaries (top) or wild type, $FLASH^{PBac/PBac}$ and mxc^{G46} ovaries (bottom) using 3 fold serial dilutions. Asterisk: cross-reacting protein. β-tubulin: loading control.
E. Anti-H2aV or anti-11 western analysis of $H2aV^{810}$.

The HLB Promotes Rapid co-transcriptional Histone pre-mRNA Processing

The histone pre-mRNA processing phenotype we observed in the $FLASH^{1-733}$ mutant, abundant WT mRNA and small but equivalent accumulation of read through and misprocessed RNA, could be due to mis-localized FLASH, U7 snRNP or both. We asked if mis-localizing only U7 snRNP *($H2aV^{810}$* mutant) or mis-localizing both FLASH and U7 snRNP (mxc^{G46} mutant) resulted in a similar histone RNA phenotype. In $FLASH^{1-733}$, mxc^{G46} and $H2aV^{810}$ mutants we detected roughly similar amounts of both misprocessed H2a RNA and read-through transcripts, although the majority of the H2a mRNA was properly processed (Fig. 3.8B,C). These results are similar to the $FLASH^{LDIY71}$ mutant (Fig. 3.3), suggesting that the molecular defect is due to reduced FLASH activity in the HLB. Importantly, the total accumulation of transgenic $FLASH^{1-733}$ protein was similar to that of $FLASH^{FL}$, and the amount of endogenous FLASH detected in wild type and mxc^{G46} mutant ovaries was also similar (Fig. 3.8D). The amount of Lsm11 protein was also similar between $H2aV^{810}$ mutant and WT larvae (Fig. 3.8E). These results suggest that the local concentration in the HLB and not the overall availability of FLASH or U7 snRNP ensure efficient histone pre-mRNA processing.

The presence of read through and misprocessed histone RNAs in genotypes that reduce the concentration of processing factors in the HLB indicates that RNA pol II must have transcribed well past the normal cleavage site. Visualizing RNA polymerase II occupancy at the histone locus in WT embryos and cultured Kc cells by mapping CHIP-Seq results from the modENCODE project reveals that RNA polymerase II occupancy extends just past the HDE of all five histone genes and that there is little or no RNA pol II present in the intergenic regions (Fig. 3.9A). This profile indicates that normally there is rapid cleavage followed by efficient transcription termination.

A reduced rate of processing leading to a failure of transcription termination would result in aberrantly long, nascent transcripts that could accumulate at the histone locus. To test whether such transcripts were present when FLASH and U7 snRNP were not concentrated in the HLB, we hybridized *FLASH*$^{PBac/Df}$, *FLASH*$^{1-733}$, and *mxc*^{G46} ovaries *in situ* with a fluorescent probe (H3-ds) derived from the region downstream of the H3 gene (Fig. 3.9A). The H3-ds probe does not detect any transcripts in wild type cells because rapid H3 pre-mRNA processing coupled to transcription termination prevents RNA pol II from entering this region (Fig. 3.9B). We previously showed that the H3-ds probe detects both nascent transcripts at the histone locus and mis-processed, polyadenylated H3 mRNA in the cytoplasm of *Slbp* and U7 snRNP mutants (Godfrey et al., 2006; Lanzotti et al., 2002). Consistent with the S1 nuclease protection data (Fig. 3.8C), *FLASH*$^{PBac/Df}$ mutant ovaries contained abundant misprocessed, polyadenylated histone mRNA in the nurse cell cytoplasm, as detected using either the H3-ds probe or an H3 coding probe (Fig. 3.9B). This phenotype was rescued by FLASH^{FL} (Fig. 3.9B). *FLASH*$^{1-733}$ and *mxc*^{G46} nurse cells contained predominantly cytoplasmic WT H3 mRNA that was detected with a coding probe, but not the H3-ds probe (Fig. 3.9B). With the H3-ds probe we detected foci of aberrantly long, nascent transcripts in *FLASH*$^{1-733}$ and *mxc*^{G46} nurse cells, but no cytoplasmic misprocessed mRNA (Fig. 3.9B). Since the steady-state level of normally processed histone mRNAs is similar in mis-localization mutants and WT animals, many of these longer, nascent transcripts may ultimately be cleaved at the normal site. Thus, the increase in steady state amount of unprocessed, nascent histone H3 pre-mRNA indicates that when FLASH or U7 snRNP is not concentrated in the HLB, the rate of histone pre-mRNA processing slows.

Figure 3.9

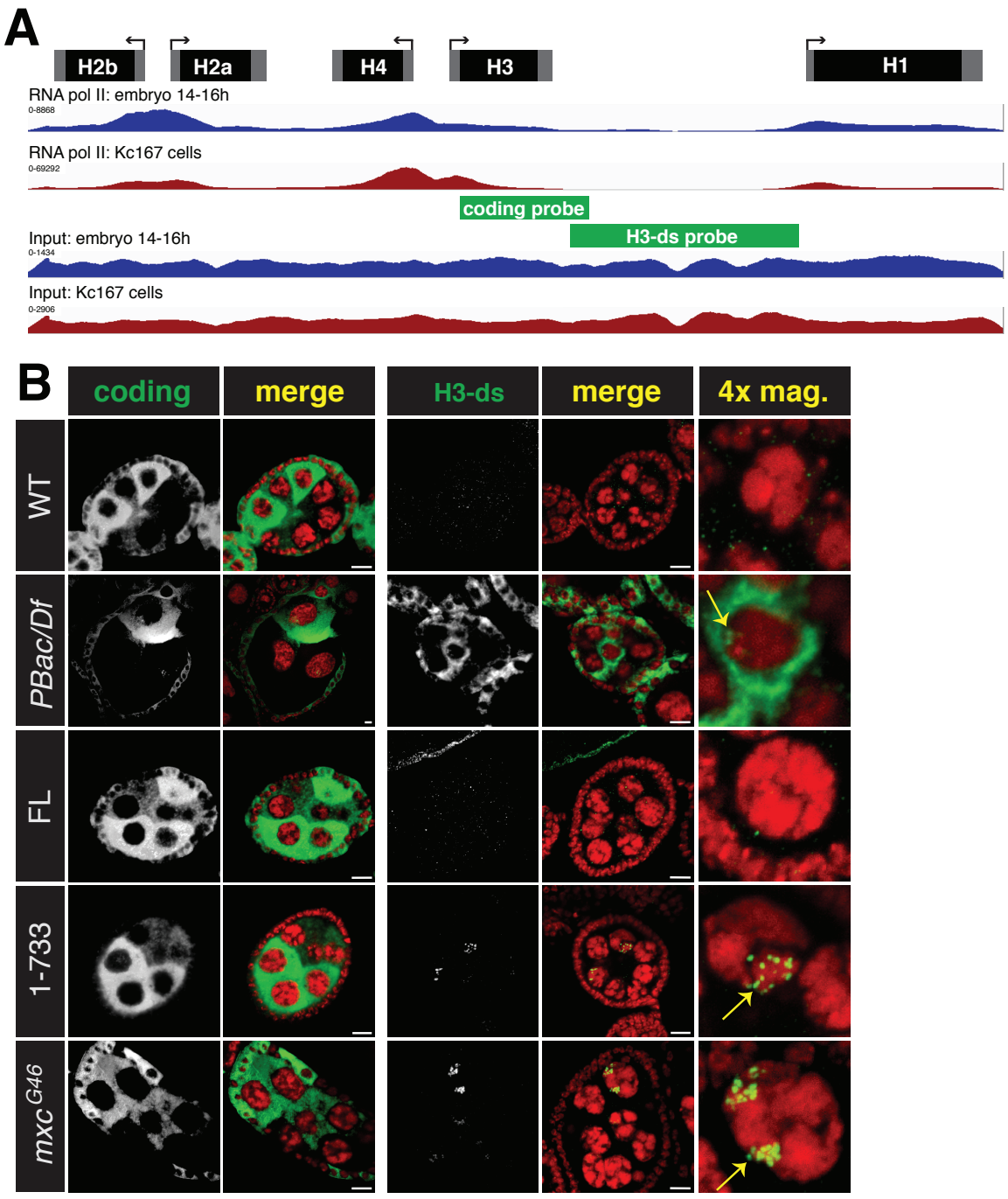


Figure 3.9. The HLB increases the rate of histone pre-mRNA processing.
A. RNA pol II occupancy data for 14-16h embryos and Kc167 cells (Kharchenko et al., 2011) visualized on a single histone repeat. The location of *in situ* hybridization probes used in panel B is indicated, as are the coding (black) and UTRs (gray) of the histone genes.
B. Fluorescent *in situ* hybridization of ovaries of the indicated genotypes using an H3 coding probe or a probe (H3-ds) that only detects mis-processed or read through transcripts. Note that

histone H3 mRNA accumulates only during S-phase and that the nurse cell endocycles are asynchronous. The cytoplasmic H3-ds signal in *FLASH*$^{PBac/Df}$ represents mis-processed H3 transcripts that were polyadenylated and exported. The small amounts of misprocessed RNA in FLASH$^{1-733}$ and *mxc*^{G46} (see Fig. 5B) are below our threshold of detection. The nuclear foci in FLASH$^{1-733}$ and *mxc*^{G46} mutants detected with the H3-ds probe (yellow arrows) are nascent transcripts at the histone locus. Scale bar = 10 μm.

DISCUSSION

The HLB Contributes to Multiple Steps in Histone mRNA Biosynthesis

We previously showed that a sequence in, and activity from, the H3-H4 promoter promotes stepwise HLB assembly and coordinate transcriptional activation of the four core histone genes (Salzler et al., 2013). Here we show that the HLB also functions to ensure proper histone mRNA 3' end formation by concentrating pre-mRNA processing factors at the histone locus. FLASH and U7 snRNP accumulation in the HLB does not require any *cis* elements in histone pre-mRNA, and they persist in the HLB in the absence of histone gene transcription (Salzler et al., 2013; White et al., 2011). Our data indicate that the HLB mediates rapid histone pre-mRNA cleavage and subsequent transcription termination, providing a clear example of how a nuclear body contributes to efficient gene expression.

Concentrating Factors within the HLB Ensures Efficient Histone mRNA Synthesis.

In this and our previous studies, we show that mis-processed, polyadenylated histone mRNAs accumulate i) when genes encoding components of the histone pre-mRNA processing machinery (e.g. FLASH, Slbp or U7 snRNP) are mutated (Godfrey et al., 2006; Godfrey et al., 2009; Lanzotti et al., 2002; Sullivan et al., 2001) ii) after engineering mutations in FLASH that interfere with Lsm11 binding or impair recruitment of the HCC, and iii) by preventing concentration in the HLB of a FLASH protein that is biochemically impaired for processing. Whether the cryptic poly A signals encoded after the HDE in each *Drosophila* histone gene provide an important function is not known. Nevertheless, our data suggest that one role of the

HLB is to promote a high concentration of pre-mRNA processing factors that facilitate stem loop- and HDE-mediated 3' end formation rather than polyadenylation.

When FLASH and/or U7 snRNP are present at normal levels but not concentrated in the HLB, or when the biochemically attenuated FLASH^{LDIY71} is localized in the HLB, we observe by *in situ* hybridization the accumulation of longer, nascent transcripts that extend past the normal processing site. The appearance of these longer, nascent transcripts indicates that normal processing is altered. We propose that the rate of cleavage has slowed resulting in failure of polymerase dissociation and continued transcription. Thus, by increasing the local concentration of processing activity, the HLB affects the rate of histone pre-mRNA processing. Interestingly, S1 analysis indicates that most of these RNAs are processed at the normal site, and only a small amount is polyadenylated. This normal processing requires FLASH, since severely depleting FLASH by mutation results in nearly 100% conversion to polyadenylated histone mRNA.

What is the basis for these RNA phenotypes? Mechanistic differences between processing 3' ends of polyadenylated mRNAs and histone mRNAs may provide an answer. The interaction of polyadenylation factors with the CTD of RNA pol II tightly couples polyadenylation with transcription (Adamson et al., 2005; Hirose and Manley, 1998; Proudfoot, 2004). In contrast, an interaction between RNA pol II and the histone-processing complex is not necessary for histone mRNA 3' end formation *in vitro* (Adamson and Price, 2003). This result suggests that processing at the normal cleavage site can occur after the polymerase has transcribed well past the HDE, making histone pre-mRNA processing more flexible than processing of polyadenylated mRNAs (Figure 3.10).

A second possible way to accumulate properly processed histone mRNA from longer transcripts is to "reprocess" a transcript that has already been either cleaved or polyadenylated (Pandey et al., 1994). In fact, in *C. elegans* the proposed normal pathway of histone mRNA processing is polyadenylation followed by cleavage after the stem loop by an siRNA-like

mechanism (Avgousti et al., 2012; Mangone et al., 2010). Thus, when normal processing is slowed, the histone pre-mRNA processing machinery may be able to act on different RNA substrates generated by continued transcription past the HDE, resulting in cleavage at the normal location (Figure 3.10).

Our analysis of publically available data of RNA pol II occupancy on *Drosophila* histone genes suggests that transcription terminates close to the HDE. Similar results were found in mammalian cells suggesting that this is a general property of histone genes (Anamika et al., 2012; Cheng et al., 2012). In *Drosophila,* strong RNA polymerase II transcriptional pause sites just past the HDE were detected biochemically (Adamson and Price, 2003). Polymerase pausing may promote pre-mRNA processing followed by rapid, subsequent transcription termination. We propose that the HLB contributes to the coupling of histone 3' end formation and transcription termination by accelerating histone pre-mRNA cleavage.

Multiple Domains of FLASH are Required for Rapid Histone pre-mRNA Processing.

Our studies indicate that FLASH contributes to normal histone mRNA 3' end formation in several distinct ways. Interaction of FLASH with another HLB component, Lsm11, and subsequent recruitment of the HCC is required for pre-mRNA processing (Sabath et al., 2013). As shown here, concentrating FLASH in the HLB is also required for recruitment of U7 snRNP to the HLB, an activity distinct from interaction of FLASH with Lsm11 required for pre-mRNA processing. Biochemical studies with both *Drosophila* and HeLa cell extracts suggest that the active form of U7 snRNP is assembled with FLASH and the HCC prior to interacting with histone pre-mRNA (Sabath et al., 2013; Yang et al., 2013). Thus, the interaction of FLASH and Lsm11 to form an active U7 cleavage complex may be regulated within the HLB independently of recruiting these factors to the HLB.

Cellular Microenvironments that Enhance Biological Processes

Our results support the idea that NBs facilitate reactions by concentrating factors in a particular sub-domain of the nucleus (Mao et al., 2011; Matera et al., 2009). The exchange of NB components with the adjacent environment is slower than expected by diffusion, promoting concentration in NBs (Dundr et al., 2004). Such concentration of components creates discrete micro-domains with different physical chemical properties than the surrounding nucleoplasm that could facilitate biological processes (Brangwynne et al., 2011). In the case of the HLB, our data provides evidence for two roles: enhancing the rate of pre-mRNA processing and promoting coupling of steps in histone mRNA synthesis. RNA binding proteins with low complexity domains assemble into bodies *in vivo* and can form a gel in solution, which may mimic some properties of nuclear bodies (Han et al., 2012; Kato et al., 2012; Lee et al., 2013). FLASH and Mxc are large proteins that contain low complexity domains that may promote the formation of the HLB through association with other proteins.

The cytoplasm also contains distinct non-membrane bound microenvironments similar to NBs. While the majority of examples involve RNA regulation (e.g. processing bodies, stress granules, P bodies (Decker et al., 2007; Brangwynne et al., 2009; Wippich et al., 2013) other structures such as the purinosome, an environment containing a concentration of the six enzymes involved in de novo purine synthesis (An et al., 2008), lipid rafts and the centrosome (Decker et al., 2011) each contain high concentrations of factors involved in a distinct process. Increased understanding of how NBs assemble and contribute to biological reactions will likely translate to understanding other macromolecular complexes and microenvironments that function in other cellular compartments.

Figure 3.10

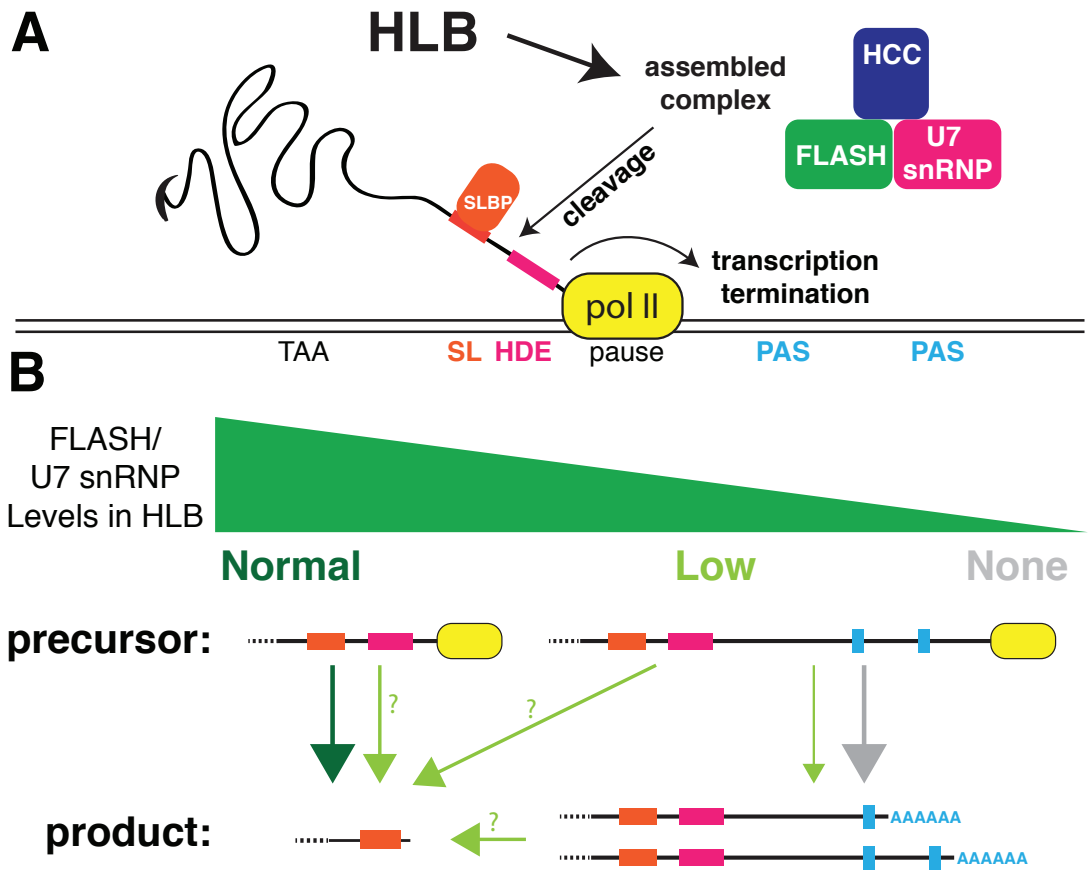


Figure 3.10 Model of HLB participation in histone pre-mRNA processing
A. SLBP (orange) interacts directly with the stem loop (SL) in the pre-mRNA downstream of the stop codon (TAA). U7 snRNP (pink), FLASH (green) and the HCC (blue) assemble into a complex that associates with the pre-mRNA at the histone downstream element (HDE) while the RNA pol II (yellow) is paused. The assembled complex triggers cleavage by CPSF73 followed by transcription termination, preventing RNA pol II encountering downstream polyadenylation signals (PAS).
B. The level of FLASH and U7 snRNP concentrated in the HLB determines both the precursor pre-mRNA substrate and the final histone mRNA products. High levels (dark green arrow) drive cleavage of histone pre-mRNA at the normal site. Low levels (light green arrows) slow normal processing resulting in accumulation of a longer pre-RNA, a small amount of which becomes polyadenylated (thin, green arrow) and most of which is ultimately processed at the normal site. With low levels there are three possible precursor RNAs that could give rise to properly processed histone mRNA (question marks) as described in the text. Absence or severe reduction results in production of a long precursor pre-mRNA that is processed by the polyadenylation machinery (gray arrow).

ACKNOWLEDGEMENTS

We thank B. Burch and S. Lyons for help with construct generation, Joe Gall for the Lsm10 antibody, S. Nowotarski and K. Boltz for critically reading the manuscript, and Greg Matera for helpful discussions. This work was supported by NIH grant GM58921 to W.F.M. and R.J.D.

CHAPTER 4: CHARACTERIZATION OF *DROSOPHILA* SYMPLEKIN TO ESTABLISH AN *IN VIVO* SYSTEM FOR TESTING HISTONE CLEAVAGE COMPLEX ASSEMBLY

PREFACE

This work represents an ongoing project of many years. My advisors, Dr. Robert Duronio and Dr. William Marzluff and I designed the experiments. An undergraduate, Alison Witowski, helped with the excision breakpoint mapping and cell cycle anaylsis of Symplekin localization, shown in Table 4.1 and Figure 4.3B. Lindsay Rizzardi assisted with the P-element excision screen, generating alleles Sym^{441} and Sym^{619}. Lindsay also stained ovaries with the YPS antibody and provided the image in Figure 4.4D. I performed all of the other experiments and wrote this section of the dissertation.

INTRODUCTION

The 3' end of a eukaryotic mRNA is generated by endonucleolytic cleavage followed by, for all but the metazoan replicative-histone mRNAs, addition of a polyadenosine (poly(A)) tail. The site of cleavage is dictated by *cis* elements in the 3'UTR of an mRNA. For polyadenylated mRNA, the signals are the poly (A) signal and G/U rich downstream element (Proudfoot, 2011). These *cis* elements assemble the cleavage and polyadenylation machinery, which minimally consists of four protein complexes, CPSF, CstF, CF I_m and CF II_m, as well as the protein Symplekin and poly(A) polymerase (Mandel et al., 2008). Cleavage of the histone pre-mRNA is also mediated by two *cis* elements, the stem loop and histone downstream element (Marzluff et al., 2008). Cleavage of histone pre-mRNA requires

histone specific *trans* factors, such as stem loop binding protein (SLBP), U7 snRNP and FLASH. It also shares components with the cleavage and polyadenylation machinery, notably Symplekin and the endonuclease CPSF-73 (Dominski et al., 2005; Marzluff et al., 2008).

How different complexes with shared components assemble on distinct sets of pre-mRNAs is an open question. A potential bridge between transcript specific *trans* factors and CPSF-73 is Symplekin. Symplekin is a conserved protein that contains a HEAT domain at the NH_2-terminus (Kennedy et al., 2009; Xiang et al., 2010). *In silico* predictions indicate that a short disordered region separates the HEAT domain from predicted armadillo repeats that extend to the COOH-terminus. Proteins with these domains are frequently classified as scaffolds (Kennedy et al., 2009); therefore Symplekin conceivably could play a critical role in the assembly of distinct complexes containing the endonuclease CPSF73. Symplekin is also required for cytoplasmic polyadenylation together with CPSF and cytoplasmic poly(A) polymerase (Barnard et al., 2004), as well as for cleavage/polyadenylation directed by STAR-PAP (Laishram and Anderson, 2010; Mellman et al., 2008). Thus symplekin may act as a scaffold for multiple complexes involved in mRNA metabolism.

Histone pre-mRNA cleavage is directed by SLBP binding to the stem loop and the U7 snRNA base pairing with the HDE (Martin et al., 1997; Mowry and Steitz, 1987; Schaufele et al., 1986; Tan et al., 2013; Wang et al., 1996). Within the Sm Ring of the U7 snRNP are two unique proteins, Lsm10 and Lsm11 (Pillai et al., 2001; Pillai et al., 2003). The N-terminus of Lsm11 binds to FLASH and this interaction is critical for histone pre-mRNA 3' end formation (Burch et al., 2011; Yang et al., 2009). Although a direct interaction between Symplekin and U7 snRNP or FLASH has not been identified, immunoprecipitation experiments under stringent conditions indicate that Symplekin forms a tight complex with recombinant FLASH and Lsm11 (Sabath et al., 2013; Yang et al., 2013).

Functional experiments demonstrate the participation of Symplekin in histone pre-mRNA processing. Early characterization of histone pre-mRNA processing indicated that a component of the cleavage machinery was heat sensitive (Gick et al., 1987). Subsequently, almost 20 years later, Kolev and Steitz showed that complementation of a heat inactivated extract by *in vitro* transcribed and translated Symplekin restored *in vitro* cleavage of a histone pre-mRNA substrate (Kolev and Steitz, 2005). Symplekin was identified as a factor required for histone pre-mRNA processing in a genome-wide RNAi screen (Wagner et al., 2007). Only two other polyadenylation factors, CPSF-73 and CPSF-100 were found in this screen, and the FIP subunit of CPSF scored weakly. In *Drosophila*, Symplekin, CPSF-73 and CPSF-100 form a stable complex termed the core cleavage complex, which is required for both histone pre-mRNA processing and cleavage and polyadenylation (Sullivan et al., 2009).

Association of Symplekin with components of the cleavage and polyadenylation machinery was first shown by far western blot with CstF member CstF-64 (Takagaki and Manley, 2000). This direct interaction has been mapped to the hinge domain of CstF-64 and residues 391 to 465 of Symplekin (Ruepp et al., 2011; Takagaki and Manley, 2000). Immunoprecipitation experiments have also shown interaction between Symplekin, CstF and CPSF indicating association of these factors to form a single large complex (Hofmann et al., 2002; Shi et al., 2009; Takagaki and Manley, 2000). Whether Symplekin directly contacts other members of the cleavage and polyadenylation machinery is not known. Interaction between the yeast homologues of Symplekin and CPSF-73, Pta1 and Ysh1 respectively, was mapped to the C-terminus of Pta1, and it is possible that this binding has been conserved in metazoans (Ghazy et al., 2009).

Direct interactions between Symplekin and other proteins involved in gene expression have been reported. The HEAT domain of Symplekin binds to the CTD phosphatase, SSU-72 (Xiang et al., 2010). Symplekin also binds to CPEB, a critical

mediator of cytoplasmic polyadenylation (Barnard et al., 2004). Additionally, Symplekin binds to a Y box transcription factor ZONAB, which is a dual localization protein that co-localizes with Symplekin at the tight junctions of polarized epithelial cells (Kavanagh et al., 2006). Curiously, Symplekin was initially discovered at the tight junction and its role in that region of the cell remains unresolved (Keon et al., 1996).

Symplekin localization is also indicative of function. While the role of Symplekin at the tight junction is unknown, accumulation of Symplekin in the nucleus is consistent with the requirement for this protein in all known pre-mRNA 3' end processing reactions. The histone genes have an affiliated nuclear body, the histone locus body (HLB). Co-localization of Symplekin with CPSF-100 in foci has been observed in Xenopus germinal vesicles (Hofmann et al., 2002), and myc-tagged Symplekin was detected in the HLB of *Drosophila* S2 cells (Wagner et al., 2007). Characterizing the dynamics of Symplekin localization to the HLB will provide insight into how the histone pre-mRNA processing complex assembles – is this general RNA processing factor always present at the histone locus, or only recruited during S phase when histone mRNA is synthesized? Whether the cell maintains distinct, pre-assembled RNA processing complexes or regulates recruitment and formation of active machinery when transcribing the RNA is an open question.

To better understand Symplekin function, I have generated and characterized an allelic series of Symplekin mutants. I have determined the requirement for Symplekin during development as well as in both poly(A) and histone pre-mRNA cleavage reactions *in vivo*. This system allows us to ask if Symplekin contains discrete regions for conferring activity. We can test, by complementation, potential separation of function mutants in Symplekin and predict that mutants that interfere with histone pre-mRNA processing would not alter cleavage and polyadenylation. Finally, I have characterized Symplekin localization during development and determined that HLB localization in cycling cells is restricted to S phase. I also show that Symplekin localization to the septate junction (*Drosophila* tight junction)

requires an RNA binding protein, YPS. This suggests that Symplekin may localize to the cell periphery as part of its role in cytoplasmic polyadenylation.

MATERIALS AND METHODS

DNA PCR Analysis

The EY20504 and NP2964 insertion sites were determined my PCR. All primer sequences are documented in Table 4.2. DNA was prepared by single fly squash prep (10mM Tris pH 8.2, 25 mM NaCl, 1mM EDTA, 200 ug/mL Proteinase K) or by the protocol provided by the Berkeley Drosophila Genome Project. 1uL of either prep was used for each PCR. pOUT and either CG2097 DNA 1 F or CG2097 DNA 1 R were used for each PCR reaction to determine the insertion site of each transposon, as well as to screen for excision of the EY20504. PCR was also used to identify the breakpoints of the four excision alleles. CG2097 DNA 1 F and R were used to identify the lesion in Sym^{58} as well as the precise excision Sym^{53}. Primers pairs throughout the EY element are listed and were used to determine the Sym^{441} and Sym^{619} sequence. 152 breakpoint primer F and 152 breakpoint primer R flank repaired DNA in Sym^{152}. Mapping Sym^{152} required isolation of homozygous embryos, as the endogenous locus on the balancer interfered with the PCR. Primer sequences are summarized in table 4.1.

Table 4.1. Primer Sequences

Name	Sequence
POUT	CCGCGGCCGCGGACCACCTTATGTTATTTC
CG2097 DNA F	GCTAAAACACCCAAGAAATG
CG2097 DNA R	TGATGGAAATCCTTACCAAA
EY1F	AGGTCGTCGAACAAAAGGTG
EY1R	CTTCGAACATCCCCACAAGT
EY2F	GGAACCACTCACCGTTGTCT
EY2R	CGGGATATGGAGCAGTTGTT
EY3F	CATTCAGCAGGGTCGTCTTT
EY3R	CATTTGCTCAAGAACGGTGA

EY4F	GAGAGGAGTTTTGGCACAGC
EY4R	CCACAGAAATATCGCCGTCT
EY5F	GGGGACCACTTGTCTCGTAA
EY5R	TATCGTCCTGTCTTGCCACA
EY6F	CGCCAGGTAGCTCGTATCTC
EY6R	CGGAGCTAATTCCGTATCCA
EY7F	GATTCGGGTGGTTCAGTGTT
EY7R	CAGCGAAAGGTGATGTCTGA
EY9F	TTCGAGCTGGGATTATTTGG
EY9R	CGGGACAGTGGAAATTGACT
152 BREAKPOINT PRIMER F	GATGGTCCGGAAATATGCAG
152 BREAKPOINT PRIMER R	GGATGCGTTAAGCCACAGAT
RP49 EXON 3 F	CCCAAGGGTATCGACAACAG
RP49 EXON 3 R	GTTCGATCCGTAACCGATGT
RP49 MISPROCESSING - 2 F	TTCTTGTAACGTGGTCGGAAT
RP49 MISPROCESSING -2 R	ATGGTGCTGCTATCCCAATC
H2A PRIMERS WERE KINDLY PROVIDED BY DR. BRANDON BURCH	
5'UTR SYMP DSRNA F	GGTAATACGACTCACTATAGGGACGGTAACGAGCAAGGAAAA
5'UTR SYMP DSRNA R	GGTAATACGACTCACTATAGGGACGGATGGCAAGGTGTTATC
3'UTR SYMP DS RNA F	GGTAATACGACTCACTATAGGGTGCAGTTTAATTGTTTTGTTTGC
3'UTR SYMP DS RNA R	GGTAATACGACTCACTATAGGGGTATGTGGGTGCGGTCAAG

Drosophila Genetics

EY20504 third chromosome excisions were recovered over Tm3 as white male progeny from *P[EY20504]/ry*^{506} *Sb P[ry* Δ*2-3]* fathers. These single males were crossed to *w;Df(3R)7283 /Tm3 Ser e p[twi-GFP]* females. Males that failed to complement the lethal were next crossed to *y*^{1}*, w*; P[lacW]cas*^{jC2}*/Tm3 Sb and w*^{1118}*;P[EP}Madm*^{EP3137}*/TM6B, Tb*^{1} to ensure that the excision did not disrupt neighboring, essential genes. Excision events were also scored by the same PCR used to identify the transposon insertion site. Of the 1,050 single males tested, we focused on four excision events. Three were internal deletions of EY20504 and the other was a deletion that extended to the *cas* promoter. The sequence of

the breakpoints is presented in Table 4.1. *U7*^{14} was previously characterized (Godfrey et al., 2006), as was the Lsm11c02047, P[V5-Lsm11+] line used for staining (Godfrey et al., 2009). *YPS* was obtained from the Whilhelm Lab. WT indicates *w*^{1118}.

Western Blot

16-20h embryos homozygous for WT, *Sym*^{152}, *Sym*^{EY}, *or Sym*^{58} chromosomes were isolated by lack of GFP expression with the Copas Embryo Sorter. 100 embryos per genotype were boiled in SDS loading buffer followed by shearing 25 times with a 27.5 gauge needle. 7.5 embryos per genotype were resolved on a 10% gel by SDS-PAGE. After transfer to a PVDF membrane, the blot was probed for Symplekin protein with an anti-rabbit Symplekin antibody (Sullivan et al., 2009). After secondary detection and exposure of the ECL signal to film, the blot was probed for tubulin with an anti-β–tubulin antibody (Abcam). After depletion by dsRNA, cultured S2 cells were incubated rotating at 4C for 30' in RIPA buffer (150mM NaCl, 1% Triton X-100,50mM Tris pH 7.5, 0.1% SDS, 0.5% Na-deoxycholate). Equivalent amounts of protein, based on a Bradford Assay (BioRad) were subject to the previously described procedure.

Quantitative RT-PCR

Total RNA was extracted from animals with Trizol Reagent (Invitrogen). 2ug of total RNA was treated with DNase (Fermentas). Samples were divided in half and a master mix of all of the reagents for reverse transcriptase, save the RT enzyme, was added to each sample (cDNA Synthesis Kit, Fermentas). Reverse transcriptase was added to one sample per genotype. After cDNA synthesis, 1:5 dilutions of sample were used for each 20uL q-PCR reaction. Indicated primers were used with SYBR green (Applied Biosystems) for each template and reactions were run on an Applied Biosystems 7900HT real-time PCR machine. The sequences are included in Table 4.1

S1 protection Assay

5 ug of RNA for each genotype was used for the S1 nuclease protection assay. The probe used and method performed have been previously described (Lanzotti et al., 2002).

Immunofluorescence

Lsm11^{c02047}, P[V5-Lsm11^{+}] homozygous embryos were dechorionated, fixed in a 1:1 mixture of 7% formaldehyde/heptane for 25 min, and incubated with primary and secondary antibodies overnight at 4°C and for 1 h at 25°C, respectively. Ovaries were dissected in Grace's medium (GIBCO), fixed in 7% formaldehyde for 25 min and permeabilized in a mixture of PBS and 0.2% Tween for 15 min. Samples were blocked for 30 min with Enhancer (Invitrogen) and incubated with primary and secondary antibodies overnight at 4 C and for 1 h at 25 C, respectively. Primary antibodies used were monoclonal mouse anti-V5 (1:1000; Invitrogen), monoclonal mouse anti-Discs Large (1:1000, Developmental Studies Hybridoma Bank), polyclonal rabbit anti-Symplekin (1:1000), polyclonal rabbit anti-YPS (1:1000), polyclonal rabbit anti-CPSF73 (1:1000), polyclonal guinea pig anti-Mute (1:5000), and monoclonal mouse anti-MPM2 (Millipore). Secondary antibodies used were goat anti-mouse IgG-Cy3 (Jackson Immuno Research Laboratories), goat anti-guinea pig IgG-Cy5 (Jackson Immuno Research Laboratories) and goat anti-rabbit Alexafluor 488 (Invitrogen). DNA was detected by staining embryos with 4,6-diamidino-2-phenylindole (DAPI) (1:1000 of 1 mg/mL stock, Dako North America) for 1 min. Images were obtained on a Zeiss 510 Confocal Microscope. All images were obtained with the 40x objective lens. Images were prepared with Lsm Software (Zeiss) and Photoshop (Adobe)

GST Pull Down

5 μg of GST protein or GST-CstF-64 were incubated at 4^{o}C with pre-equilibrated glutathione-sepharose resin (GE lifesciences) in 100 μl of TEN100 buffer (20 mM Tris [pH 7.5], 0.1 mM EDTA, 100 mM NaCl). *In vitro* transcription and translation of Sym 383-465

from the PCDNA3 vector was carried out using Promega's TNT coupled rabbit reticulocyte kit. Unbound protein was removed by washing 2X with 250 μL TEN100. 10 μL of in vitro translated protein was added to beads along with 10 uL of 10X TEN100 buffer, 14 μL of GDB buffer (10% Glycerol, 10 mM DTT, 0.05 mg/mL BSA) and 76 μL of dH_2O. Proteins were allowed to bind for 2 hours at 4°C while rotating. Glutathione beads were washed 4 times with 1 mL of TEN100 buffer supplemented with 0.1% NP40. 25 μL of 2X SDS loading dye (4% SDS, 10% β-mercaptoethanol, 0.125 M Tris [pH 6.8], 20% glycerol, 0.2% Bromophenol Blue) was added to beads and boiled for 10 minutes. The supernatant was loaded onto an SDS-PAGE gel. Gels were stained with commassie blue to confirm pulldown of recombinant GST protein. Gels were dried and visualized by autoradiography.

Symplekin Δ 64 Construct

Symplekin cDNA was cut with XhoI and AvrII and blunt ends were generated with Klenow fragment. Recovered ligations were amplified and sequenced to ensure that the Symplekin coding region remained in frame.

RNAi Knockdown

Templates for the 3'UTR and 5'UTR dsRNA were PCR amplified from genomic DNA and subcloned into the TOPO-TA cloning vector (Invitrogen). The primers contained a T7 promoter (Table 4.2). dsRNAs were made by T7 transcription followed by digestion of the template by DNASE. RNAi was performed by adding 100ug of dsRNA to S2 cells (2×10^7) on day 1, 200 ug on day 2, and 400 ug on day 3. After two days of recovery, protein was extracted as described above.

RESULTS

In *symplekin* Mutants the Developmental Phenotype Correlates with the Level of Symplekin

To examine the contribution of Symplekin to a developing animal, I generated and characterized an allelic series of mutations in the *Drosophila symplekin* gene. We first obtained two mutant stocks: one from the Berkely Drosophila Genome project that contained a P-element insertion in the 5' UTR of the *symplekin* gene ($Sym^{EY20504}$) and another P-element insertion line, annotated to be in the second exon of *symplekin*, from the *Drosophila* Genetic Resource Center (Sym^{NP2964}). Sequencing a PCR product flanking the transposon insertion sites confirmed the $Sym^{EY20504}$ annotation. Based on experimental validation of the *symplekin* 5'UTR (Nechaev et al., 2010), we determined that the EY insertion occured in the 5'UTR, 63 nt upstream of the translation start site. Our PCR data indicated that the Sym^{NP2964} insertion was also in the 5' UTR 51 nt upstream of the transcriptional start site. (Figure 4.1A, Table 4.1). Analysis of each of these transgenes in *trans* with a genomic deletion of the *symplekin* locus ($Sym^{Df(3R)Exel7283}$) indicated that Sym^{NP2964} was a viable and fertile allele (data not shown) and $Sym^{EY20504}$ was a hypomorphic allele (Figure 4.2B). $Sym^{EY20504}$ progeny of $Sym^{DF(3R)Exel7283}$ mothers did not live past the wandering 3rd instar larval stage while ~32% of the expected adult flies were detected when the $Sym^{Df(3R)7283}$ chromosome was paternal. This difference in maternal contribution between $Sym^{DF(3R)7283}$ and $Sym^{EY20504}$ indicated that the chromosome with a transposon insertion in the 5'UTR expresses some Symplekin protein. We therefore decided to generate additional *symplekin* mutants.

Mobilization of a transposable element produces double strand DNA breaks, so we designed a screen to recover imprecise repair events after excising the $Sym^{EY20504}$ P-element. (see Experimental Procedures for details). Of the 1,050 flies screened, we focused on four recovered alleles for our analysis of Symplekin function (Table 4.1). Three of the recovered alleles contained internal deletions of the EY element. Sym^{441} and Sym^{619} still

contained a large portion of transposon sequence. ~200bp remained in the 5'UTR of Sym^{58} 5'UTR. We also recovered an allele that contained a deletion of *symplekin*. This deletion extended beyond *symplekin* to the intergenic region upstream of the adjacent gene *castor*. We did not recover a *symplekin* allele that disrupted the *symplekin* coding region without affecting neighboring essential genes. We also isolated flies from a precise excision event. We subsequently used one of these alleles, Sym^{53}, as our wildtype control to ensure that any observations were not a consequence of other lesions on the chromosome.

We first compared the developmental phenotypes of our excision alleles with that of the original $Sym^{EY20504}$ fly (Figure 4.2B). We focused on the strongest genetic combination of *symplekin* mutants by crossing each of our alleles to $Sym^{DF(3R)7283}$. Sym^{152} deleted *symplekin* and was embryonic lethal, while all of our other *symplekin* alleles hatched (Figure 4.2C). Sym^{152} deletion did not remove the coding region of the essential gene *castor (cas)*, however this gene is controlled by numerous regulatory elements upstream of the gene. To test if the observed lethality was a consequence of disrupting *cas* regulation, we asked if Sym^{152} was complemented by a lethal castor allele. This genetic combination was also lethal (data not shown), so we concluded that the Sym^{152} lethal phenotype was not due solely to mutation of symplekin.

Our other symplekin alleles, Sym^{441}, Sym^{619} and Sym^{58}, contained internal deletions of the EY transposon (Table 4.2). Sym^{441} and Sym^{619} retained a significant amount of transposon sequence in the 5'UTR and displayed the same larval lethality phenotype as $Sym^{EY20504}$. Sym^{58} resulted from a larger deletion of transposon sequence, leaving ~200bp of exogenous sequence in the 5'UTR of the symplekin gene. This allele was viable but sterile. Sym^{58} flies were also developmentally delayed, as they eclosed 1-6 days later than their siblings that contained a wildtype copy of *symplekin* (Figure 4.1D). We therefore hypothesized that the different amounts of transposon sequence in the 5'UTR affected Symplekin expression and consequently development.

To compare disruption of Symplekin expression across our allelic series, we measured Symplekin protein in 16-20h embryos. We isolated protein from staged, GFP negative mutant embryos and detected Symplekin levels by western blot (Figure 4.1E). At this stage, all of our *symplekin* alleles have less protein than the wildtype control, as compared to a non-specific cross reacting band that we have previously observed with this antibody (*). Symplekin was detected in Sym^{152}, suggesting that there is still some maternal supply of Symplekin throughout embryogenesis (Figure 4.1E, lane 2). The increased level of protein detected in our other alleles must therefore be produced zygotically from the p element chromosome. As expected, the amount of Symplekin expressed correlated with the developmental phenotype of the animal. I could therefore next explore the consequences of different in concentrations of a critical trans factor in forming the 3'ends of polyadenylated and histone mRNAs.

Figure 4.1

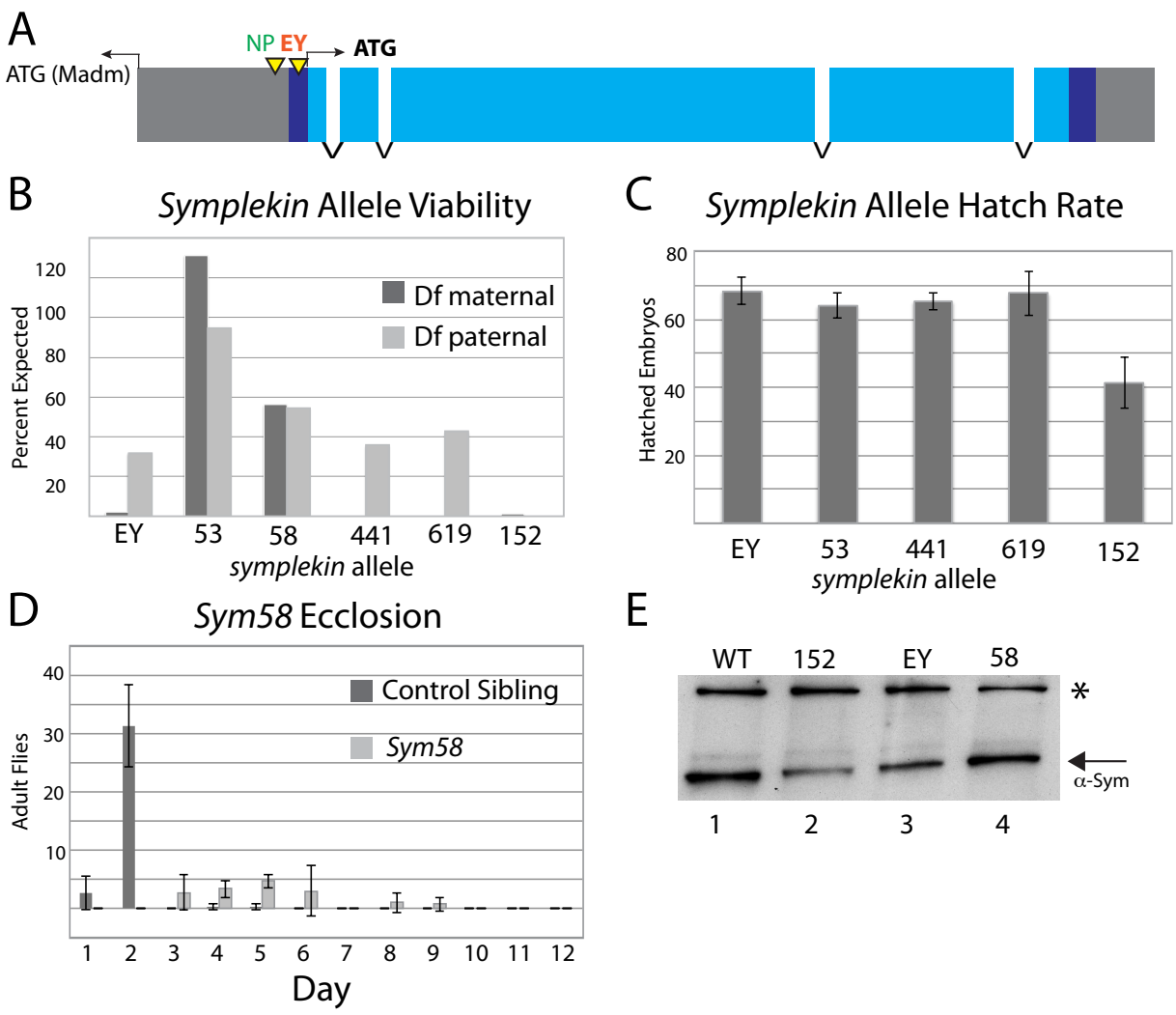


Figure 4.1. Characterization of *Symplekin* Allelic Series.

A. Schematic of *symplekin* gene structure. Insertion sites of $Symp^{EY20504}$ and $Symp^{NP2964}$ are indicated.
B. Quantification of viability of *symplekin* mutants. Each mutant (X axis labels) was recovered in *trans* to either a maternal or paternal $Sym^{DF(3R)7283}$ chromosome. The percent expected calculation was based on the recovered number of control siblings.
C. Quantification of *symplekin* allele hatch rates. The number of viable progeny of paternal *symplekin* alleles (X axis labels) was quantified 48 hours after egg collection. Bars represent the average of three individual experiments, of which each contained 100 embryos. Error bars indicate standard deviation.
D. Quantification of Sym^{58} ecclosion. $Symp^{58}$ mutant flies and control siblings were counted each day for 12 days. Each bar represents the average of three vials and the error bars indicate standard deviation.

E. Symplekin protein levels of *symplekin* mutants during embryogenesis. 16-20 h embryos of the indicated genotype were isolated and protein levels were extracted for western blot analysis. Arrow indicates Symplekin protein. * represents an unrelated cross reacting band.

Table 4.2 *Sym* Allele DNA Sequences

Lesion	Sequence
EY20504	5'UTR START: CCTTGCCATCCGTGCCG / GCAGCTCTTAAAAATATCCGAA
NP2964	CAAAAAATAGCTAGATGGT / CCGGAAATATGCAGAATTACC
152	Sym-CCTTGCCATCCGTGCC / GGCAGCTCCATGATGATCTCCCCCTTCT / TCCGGGGGGTAGAA-cas
619	EY 5' CATGATGAAATAA[A] / TCCGATCATCGGATAGGCAATCG …
441	EY 5' … ATCCGCACCATGGCTTG / GAGGCTCGTGCATAGAATGCA …
58	EY 5' … CAAAGCTGTGACTGGA / TGTTATTTCATCATG

Reducing Symplekin Affects both Poly(A) and Histone pre-mRNA Processing

Biochemical experiments indicate that Symplekin is a critical component of both the poly(A) machinery as well as the histone cleavage complex (HCC). I therefore asked if reducing Symplekin in vivo affected either of these processes. With the allelic series, I could also ask if the extent of Symplekin depletion modulated the severity of any RNA phenotype. I first assessed a gene, *ribosomal protein L32 (Rp49),* that was known to be sensitive to disruption of the poly(A) factor, CstF77 (*Su(F)*) (Benoit et al., 2002). It was previously shown by RT-PCR that when cleavage and polyadenylation was disrupted, RNA Pol II continued transcribing and resulted in a detectable read through RNA species (Benoit et al., 2002). Primers upstream of the cleavage site measure the total amount of the transcript, while primers downstream detect the aberrant transcript (Figure 4.2A). The relative amount of each of these species can be measured and expressed as a ratio. We extracted RNA from Sym^{58} well as control RNA from WT and *U7* mutant flies as well as cells depleted by RNAi of the polyadenylation factor CPSF-30 as a positive control. After performing quantitative-RT-PCR, I detected readthrough RP49 RNA in the Sym^{58} mutant, but not in the histone processing *U7 snRNA* mutant (Figure 4.2B).

I assessed histone mRNA processing with a similar assay. Both the $Sym^{EY20504}$ and Sym^{58} alleles had misprocessed histone mRNA, however it was not a significant fraction of the total histone mRNA transcript (Figure 4.2C). To directly visualize the histone mRNA

species in our *symplekin* mutant flies, we performed an S1 nuclease protection assay (Figure 4.2A). Complementary to the quantitative qRT-PCR result, we primarily detect properly processed histone mRNA in both of these genotypes. We do detect longer, aberrant transcripts in both genotypes. In the *Symp*^{58} genotype, we also detect accumulation of mRNA that extends past the region of complementarity to the probe. This phenotype has been observed in S2 cultured cells depleted of Symplekin (Sullivan et al., 2009). Because Symplekin is required for both histone pre-mRNA processing, as well as cleavage and polyadenylation, we interpret accumulation of these species as an indication of disruption of both processes. We therefore conclude that reducing Symplekin levels in an animal results in misprocessing of both histone and poly(A) mRNA.

Figure 4.2

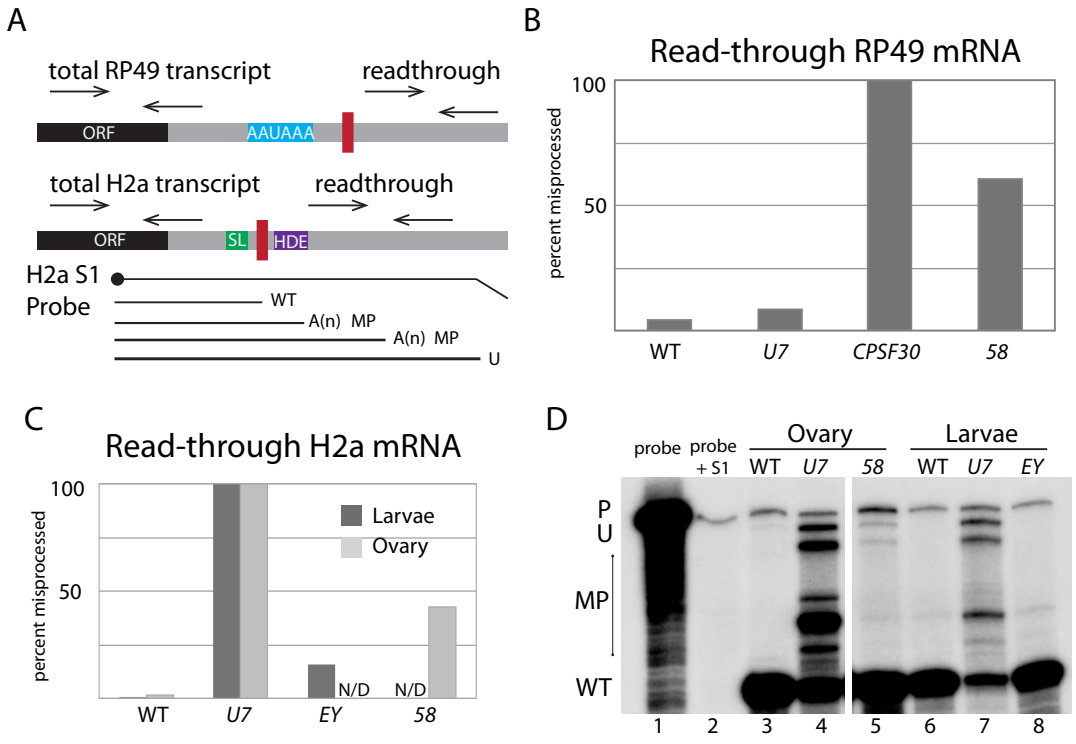


Figure 4.2. Characterization of *Symplekin* mutant RNA phenotypes.
A. Schematic of RNA assays. Representative diagrams indicating location of primers used to detect total and readthrough RP49 and H2a mRNA by quantitative-RT-PCR. Also

depicted is the S1 protection assay probe and representative RNA species that could hybridize to the radiolabeled, DNA probe.
B. Measurement of readthrough RP49 mRNA in *Sym*^{58} ovaries. The levels of readthrough mRNA were first calculated as a percentage of total mRNA and are presented in relation to the positive control, RNA from CPSF30 depleted S2 cells, which was set to 100 percent. RP49 readthrough message is detected in *Sym*^{58}.
C. Measurement of readthrough H2a mRNA in either larvae or ovaries. The levels of readthrough mRNA were first calculated as a percentage of total mRNA and are presented in relation to the positive control, *U7*^{14}, which was set to 100 percent. Readthrough H2a RNA is detected in both *Sym*^{EY20504} larvae and *Sym*^{58} ovaries
D. Visualization of H2a RNA species by S1 protection assay. The 5'-labeled DNA probe used in the S1 nuclease protection assay is complementary to the H2a message starting in the coding region through to the convergent H4 HDE. Properly processed histone message is cleaved in between the stem loop and HDE resulting in a single protected fragment of 340bp. Misprocessed transcripts are longer and polyadenylated, The H2a gene contains multiple cryptic polyadenylation signals, so several protected products represent aberrant histone processing. The poly(A) tail is not complementary to the probe, so each band represents transcripts that have been cleaved at the same site. The probe also contains vector sequence, so any transcripts that extend beyond the complementary region will collapse the probe to a single band that is distinct from undigested probe. Properly processed histone mRNA is detected in both *Sym*^{58} and *Sym*^{EY20504} alleles. *Sym*^{58} also contains mis-processed and unprocessed histone mRNA.

Symplekin Localization Is Dynamic

In my effort to understand the many roles of Symplekin in a complex organism, I next asked if Symplekin localization was related to function. Symplekin is enriched in the HLB, forming a focus that co-localizes with V5-tagged Lsm11 in a cellularising embryo, as well as being ubiquitously expressed throughout the nucleoplasm, in agreement with its role in polyadenylation (Figure 4.3A). This agrees with a previous report that myc-tagged Symplekin localizes to the HLB (Wagner et al., 2007) in cultured cells. We next asked if symplekin was localized in the HLB throughout the cell cycle.

To quantify Symplekin localization to the HLB during the cell cycle, we focused on the follicle cells of stage eight ovaries. These endocycling cells oscillate between S and G phase of the cell cycle. We visualized the HLB with two markers: anti-mute, which labels the HLB in both S and G phases of the cell cycle and anti-Mpm-2, a reagent that detects phosphorylation of another HLB component, Mxc, only during S phase (Figure 4.3B). Of all of the Mute defined HLBs, we found that Symplekin and MPM-2 co-localize with mute in

7.2% of stage 8 follicle cells (Figure 4.3B). We also detected co-localization of just mute and MPM-2 and just mute and Symplekin (13.6 and 7.5% respectively). A previous study reported that ~30% of follicle cells are in S phase at this stage (Calvi et al., 1998). Combining our three observed co-localization classes shows that 28.3% of cells are MPM-2 and or Symplekin positive and we interpret this to mean that Symplekin, though out of phase with MPM-2, is enriched in the HLB during S phase. This result classifies Symplekin as a transient resident of the HLB.

To further confirm that Symplekin localizes to the HLB in S phase, we hypothesized that Symplekin would not be detected in cells that have differentiated and exited the cell cycle. Ovarian stage 10B follicle cells no longer endocycle, though they are amplifying the chorion gene loci. When we examined Symplekin localization in this population of cells, we no longer detect Symplekin enrichment in the HLB (Figure 4.3C). We still observe nuclear staining but surprisingly there was intense staining at the septate junction, specifically the tri-cellular junction, but not the junctions between adjacent cells. Thus symplekin does not localize to HLBs in cells that have exited the cell cycle. We confirmed our observation that Symplekin localized to the junction by expressing ectopic HA tagged Symplekin and assaying its localization (Figure 4.4C). Like the endogenous protein, we detected HA:Symplekin at the junction of stage 10B follicle cells (Figure 4.4B).

Figure 4.3

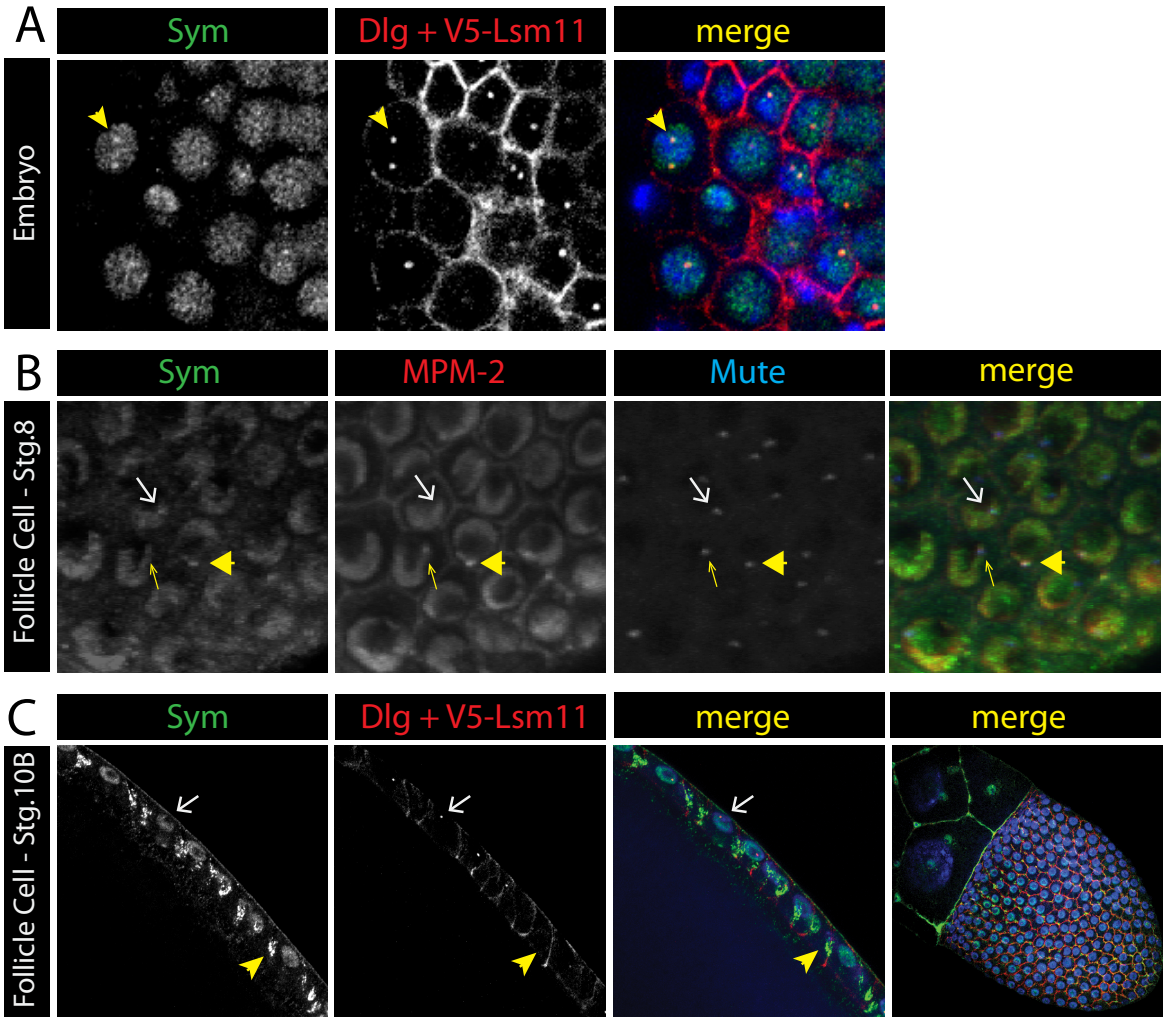


Figure 4.3. Characterization of Symplekin Localization During Development.
A. Stage 5 Lsm11c02047, P[V5-Lsm11+] homozygous embryos were stained with anti-Symplekin, anti-Discs Large antibodies, to visualize cell boundaries, and anti-V5 antibodies (center panel, both red in merge) to visualize the HLB. Anti-mouse secondary antibodies were used to simultaneously detect V5-Lsm11 and Discs Large and anti-rabbit secondary antibodies were used to detect Symplekin. Embryos were also stained with DAPI (blue in merge). Note that V5-Lsm11 and Symplekin co-localize (yellow arrowhead).
B. Representative images used to quantify Symplekin localization to the HLB during the cell cycle. Stage 8 ovarian follicle cells were stained with anti-Symplekin, anti-MPM-2, to visualize S-phase HLBs, and anti-Mute to visualize all HLBs. Of 558 mute positive HLBs, 400 (71.6%) were only mute positive, 76 (13.6%) contained Mute and MPM-2 (yellow arrow), 42 (7.53%) contained Mute and Symplekin white arrow, and 40 (7.17%) contained all three HLB components (yellow arrowhead). This data was contributed by Alison Witowski.

C. Symplekin localizes to the junction and not the HLB in stage 10B ovarian follicle cells. Stage 10 ovarian follicle cells were stained as described in **Figure 3A.** White arrow indicates the V5-Lsm11 positive HLB, which does not contain Symplekin. Yellow arrowhead indicates Symplekin localization to the junction. Fourth panel indicates stage of the other images in part C.

Localization of Symplekin to the Junction Depends on YPS

Symplekin was discovered because it localized in polarized epithelial cells to the tight junction (Keon et al., 1996). It was subsequently characterized as a component of the poly(A) machinery, HCC and cytoplasmic polyadenylation machinery (Barnard et al., 2004; Kolev and Steitz, 2005; Takagaki and Manley, 2000; Wagner et al., 2007).

Studies of Symplekin at the tight junction showed that it and the Y box transcription factor, ZONAB, co-localize at both the junction and nucleoplasm of an intestinal epithelial cell line, HT-29 (Kavanagh et al., 2006). This study also showed direct interaction between recombinant Symplekin and ZONAB. I was curious if this interaction was conserved in flies and searched the genome for homologous *Drosophila* proteins. The highest similarity of ZONAB to an RNA binding protein was another Y box protein, - (*YPS*) (Figure 4.4A). YPS interacts with RNA trafficking proteins and is a component of a developmentally localized mRNP, oskar (Mansfield et al., 2002; Wilhelm et al., 2000). We hypothesized that if the YPS-Symplekin interaction were conserved, we would detect it at the junctions of stage 10B follicle cells.

Like Symplekin, cytoplasmic YPS also was enriched at the tri-cellular junction of stage 10B follicle cells (Figure 4.4D). This suggested conservation of the interaction between these two proteins. To test if Symplekin localization was dependent on YPS, we stained YPS mutant ovaries with Symplekin. In mutant stage 10B follicle cells, Symplekin was no longer enriched at the tri-cellular junction and instead was dispersed in the cytoplasm (Figure 4.4E). We conclude that YPS is required for Symplekin localization to the septate junction in this cell type.

YPS binds to an RNA that is regulated by cytoplasmic polyadenylation. It is therefore possible that we detect Symplekin at the junction because RNAs are localized there. To determine whether other factors involved in cytpplasmic polyadenylation are localized at the septate junction, we stained ovaries with another RNA processing factor that is involved in cytoplasmic polyadenylation, CPSF-73 (Figure 4.4F). Like Symplekin and YPS, this protein was also localized to the junctions only in stage 10B follicle cells. CstF64 is another RNA processing factor, and Symplekin interacting protein, that has not been implicated in cytoplasmic polyadenylation. This factor does not localize to the junction, further supporting the distinct accumulation of the cytoplasmic polyadenylation machinery (data not shown). These results suggest that Symplekin localizes to the junction as a component of an mRNP(s) that contain YPS.

Figure 4.4

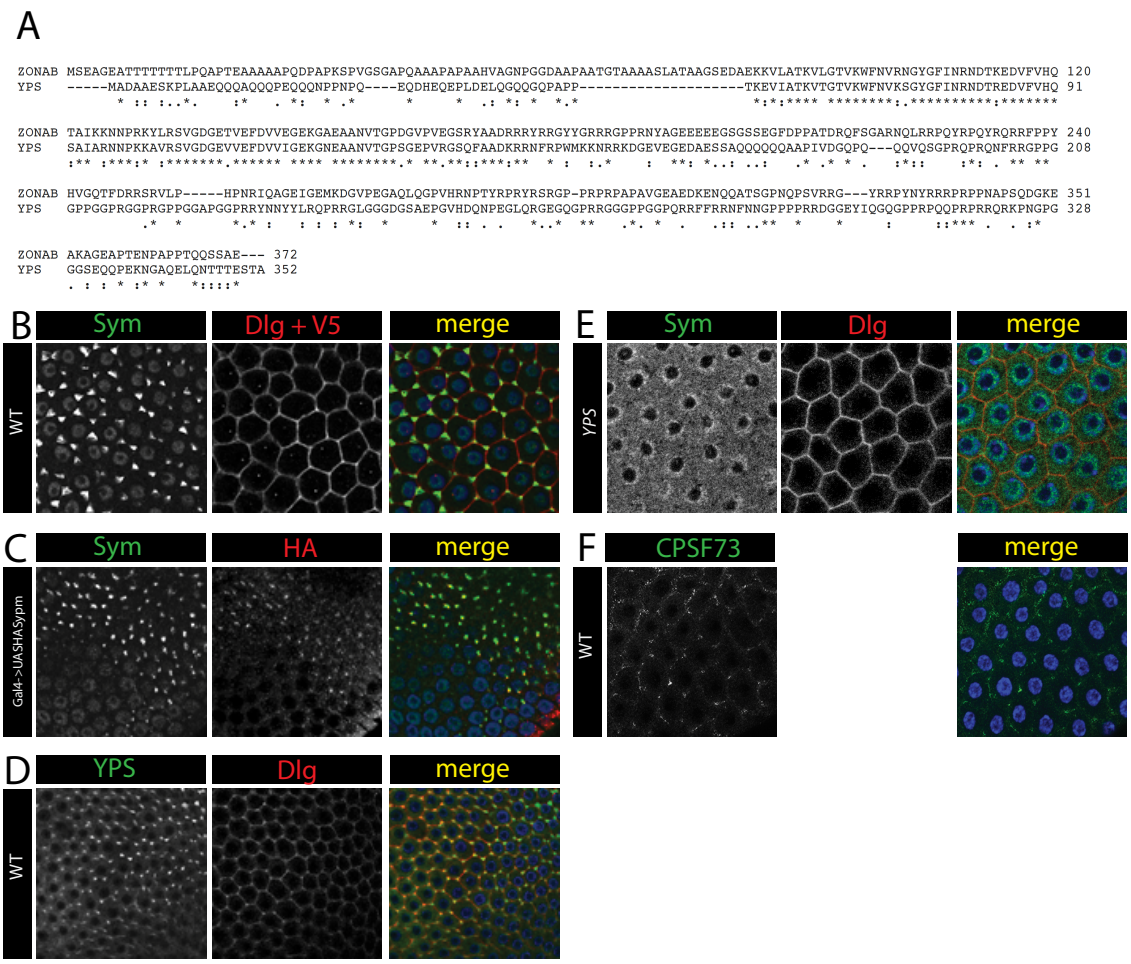


Figure 4.4. Characterization of Symplekin Localization to the Septate Junction

A. Alignment of hZONAB and dYPS. Alignment was performed with Clustal W.

B. Stage 10B ovarian follicle cells were stained as described in **Figure 2C.**

C. Ectopic full length Symplekin localizes to the septate junction. UAS-HA:Symplekin was expressed with c329b-Gal4, which is a stage 10B follicle driver.

D. YPS localizes to the septate junction and accumulates at the tricellular junction. Anti-YPS was visualized with anti-rabbit secondary antibodies. This image was contributed by Lindsay Rizzardi.

E. Symplekin localization to the junction requires YPS. Anti-Symplekin was stained in *YPS/Df* mutant ovaries. The junctions were visualized with ant-Dlg. **(F)** CPSF-73 localizes to the septate junction. Stage 10B follicle cells were stained with anti-CPSF 73 and were visualized with anti-rabbit secondary antibodies.

Candidate Symplekin Separation of Function Mutation

Symplekin interacts with a number of complexes involved in RNA biology. We hypothesize that specific regions of Symplekin enable it to incorporate into distinct machinery, and other distinct regions may be important for its function in polyadenylation and histone pre-mRNA processing. Manley and coworkers identified a strong interaction between Symplekin and CstF-64 (Takagaki and Manley, 2000). However the same region of CstF64 interacts with both CstF77 for polyadenylation and with Symplekin, suggesting that this reaction is not involved in polyadenylation. Schumperli and coworkers showed that this interaction was likely required for histone pre-mRNA processing in vivo in mammalian cells and dispensable for polyadenylation (Ruepp et al., 2011). Zbig Dominski showed that both Symplekin and CstF64 are in the HCC in mammalian cells and Drosophila (Sabath et al., 2013; Yang et al., 2013) and provided evidence that the CstF64/Symplekin complex might contact the FLASH/Lsm11 heterodimer. I have designed a strategy to test the importance of this interaction for histone pre-mRNA processing *in vivo*.

I first asked if the region of Symplekin required for the CstF-64 interaction was conserved. With assistance from Dr. Kevin Slep, we identified amino acids 383-465 of dSymplekin (Figure 4.5B) as the orthologous region that interacts with CstF64 in mammals. This fragment of Symplekin was labeled by *in vitro* translation with ^{35}S-methionine and incubated with full length GST-tagged dCstF64 (Figure 4.5C). Visualization of the interaction by autoradiography indicated that this 80 amino acid region of Symplekin binds to CstF64. Whether this interaction occurs in the context of the full length protein is yet to be determined.

I next designed four constructs to test the functional consequences of disrupting this interaction *in vivo* (Figure 4.5A). All four constructs contain the endogenous Symplekin promoter, 5' and 3' UTRs, first two exons and endogenous polyadenylation sequences. Two of the constructs contain C-terminal HA tags, and two are untagged. The CstF-64 binding

region (amino acids 388-453) was deleted from a tagged and untagged construct. Insertion of these various versions of the *symplekin* locus will allow us to test the ability of each protein to rescue *symplekin* mutant phenotypes, when expressed from the symplekin promoter. Replacement of the 3'UTR with that of another gene, such as OpeI, in these constructs would also allow for cell culture analysis, as endogenous symplekin can be efficiently depleted with double stranded RNA directed to the Symplekin 3' UTR (Figure 4.5D), and we can test the ability of the different symplekin constructs to rescue the effect of the knockdown, testing restoration of polyadenylation and histone 3' processing. Overall, this establishes a system for testing the role of different regions of Symplekin *in vivo,* which should be amenable for a genetic screen.

Figure 4.5

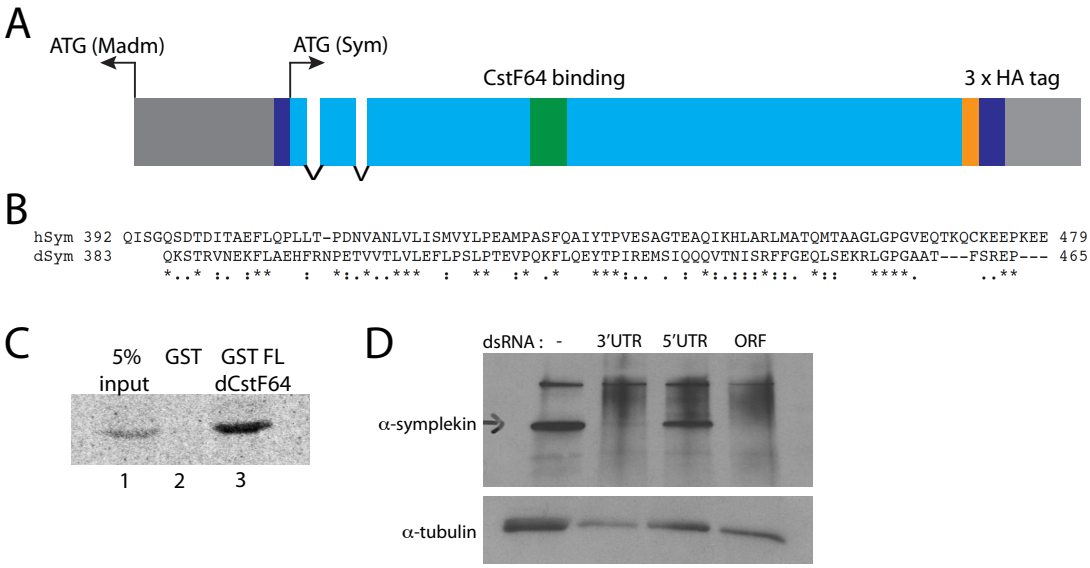


Figure 4.5. System for Testing Candidate Symplekin Interactions.
A. Schematic identify feature of the Symplekin rescue constructs. Grey indicates regulatory regions. Dark blue represents the 5' and 3' UTRs. The Symplekin CDS is light blue. The CstF-64 binding region, which may or may not be present in the clone, is marked in green. The HA tag, which also may not be included is represented in orange.
B. Alignment of amino acids 383-465 of dSymplekin with hSymplekin. Clustal W was used to align the fragment of dSymplekin with FL hSymplekin. Only the aligned region is shown.

C. GST pulldown of *in vitro* transcribed and translated Sym383-465 with full length dCstF-64. Symplekin was incubated with either GST or GST-CstF-64 followed, resolved by SDS-PAGE and visualized by autoradiography.
D. Depletion of Symplekin in cultured S2 cells. dsRNA targeting the 3' UTR reduces Symplekin protein. The reduction is not as efficient as with dsRNA directed to the ORF. dsRNA directed to the 5'UTR did not result in Symplekin depletion.

DISCUSSION

Symplekin participates in histone pre-mRNA processing, cleavage and polyadenlyation, cleavage followed by STAR-PAP mediated polyadenylation and cytoplasmic polyadenylation, and may also play a role in mRNA localization. As a potential interface between specialized *trans* factors and general components of 3' end processing machinery, we hypothesize that Symplekin incorporates into each of these multi protein complexes by binding specific *trans* factors. Understanding these networks will connect what is known about specific interactions with mRNA *cis* elements and functional studies of these processes. As each region of binding between Symplekin and members of the listed complexes is identified, it will be possible to test the functional consequences of disrupting that interaction.

Assembling Symplekin into the Histone pre-mRNA Processing Complex

Depletion of dSymplekin results in mis-processed and unprocessed histone mRNA (Sullivan et al., 2009; Wagner et al., 2007) and our analysis of two *symplekin* mutants indicate that Symplekin is required for accurate histone 3' end formation in the animal. These longer polyadenylated or unprocessed transcripts do not accumulate to the levels of the *U7 snRNA* mutant fly. The fact that Symplekin is also required for cleavage and polyadenylation affects production of the mis-processed mRNA; however it is also worth noting that in the animals of this study, as well as the previously reported Symplekin depleted S2 cells (Sullivan et al., 2009), a significant amount of properly processed histone mRNA was detected. As these

are hypomorphic mutants, it is likely that the amount of Symplekin generated from the mutant chromosome is sufficient for complete, rapid histone mRNA processing. In vivo there is only small amounts of the HCC, determined by the levels of U7 snRNP, which requires only 1% of the CstF64 or symplekin in the cell (Sabath et al., 2013; Yang et al., 2013). Thus low levels of Symplekin may be sufficient for histone pre-mRNA processing, if the symplekin CstF 64 interaction is stronger than other interactions symplekin is involved in.

The amount of properly processed histone mRNA in the larval lethal $Sym^{Ey20504}$ allele exceeds that of the viable *U7* mutant control. We conclude that the lethality of $Sym^{EY20504}$ is likely due to a general defect in expression of polyadenylated mRNAs, as we have observed a defect for both, *RP49* and *H2a* mis-processing in this allele. Mutation of a CstF-77, a polyadenylation factor, results in larval lethality (Benoit et al., 2002), however mutation of CPSF-160 causes embryonic lethality (Salinas et al., 1998), indicating general defects in polyadenylation is lethal. The phenotypes, both developmental and for pre-RNA processing, of a molecular null allele of *symplekin* is yet to be determined.

dSymplekin forms a tight complex with CPSF-73 (Sullivan et al., 2009), which is involved in polyadenylation but also involved in histone pre-mRNA processing. Determining how Symplekin interacts with FLASH and or U7 snRNP could indicate the architecture of the HCC complex just before the enzymatic reaction. It was shown that a complex of recombinant FLASH and Lsm11 can immunopreciptate Symplekin (Sabath et al., 2013; Yang et al., 2013). Whether this interaction only occurs in S phase, or throughout the cell cycle is not known. Controlling this interaction could provide a layer of temporal regulation for histone pre-mRNA processing. Our analysis of HLB localization in ovarian follicle cells indicates that Symplekin is enriched in the HLB during S phase, while U7 snRNP is concentrated there throughout the cell, suggesting assembly of the active U7 snRNP may be regulated during the cell cycle.

The Role of Symplekin at the Septate Junction

I identified a tissue where Symplekin accumulates at the septate junction, together with CPSF73 and YPS, but not the general polyadenylation factor CstF64. This has the hallmarks of a cytoplasmic polyadenylation complex. The genetic interaction between Symplekin and YPS suggests that detection of Symplekin at the junction is a consequence of mRNA trafficking. YPS has been implicated in translational repression of Osk mRNA and is part of the mRNP that traffics the mRNA along microtubules (Mansfield et al., 2002; Wilhelm et al., 2000). Preventing translation until the mRNA reaches its destination ensures localized expression. This is mediated through extension of the poly(A) tail through cytoplasmic polyadenylation, which can modulate translation.

The absence of YPS results in diffuse, but detectable Symplekin staining in the cytoplasm of stage 10B follicle cells. This extensive accumulation outside of the nucleus contrasts with all other Drosophila cells characterized thus far, where Symplekin staining appears exclusively nuclear. That is not to say that cytoplasmic polyadenylation, or participation of Symplekin in that processes, does not occur, but if it is spread throughout the cytoplasm, it is likely below our threshold of detection. Our ovarian sample preparation conditions do not result in Symplekin penetrance into nurse cells or the developing oocyte, where we might detect additional cells with cytoplasmic or strongly positioned enrichment of Symplekin protein.

If Symplekin and YPS localization to the septate junction indicates localized regulation of mRNA translation, there is likely a highly abundant transcript(s) localized to this region. Stage 12-14 ovarian follicle cells produce the chorion proteins for the egg shell and during stage 10B of oogenesis, these cells produce Yolk and Vitellin, so these transcripts could be localized to the cell periphery. Microarray analysis of follicle cells during the late stages of oogenesis has provided a complete identification of follicle cell specific mRNAs, which are candidates for localized mRNAs (Tootle et al., 2011). Performing *in situ*

hybridization to a candidate list of mRNA might identified a localized transcript(s). Since YPS was implicated in translational repression preceding Osk protein synthesis, it is possible that this transcript(s) might be stored at the septate junction and then not translated until a later stage of development.

YPS was identified through a BLAST search for homologues to ZONAB. This transcription factor binds to the tight junction protein ZO-1 (Balda and Matter, 2000). A recent study showed that ZO-1 mRNA is regulated by cytoplasmic polyadenylation. ZO-1 requires CPEB (Orb in *Drosophila)* binding sites in the 3' UTR for localized translation and consequently tight junction integrity and polarization of the cells (Nagaoka et al., 2012). It is therefore possible that a structural protein is locally translated at in this cell type. Identification of a localized mRNA will confirm our hypothesis that Symplekin localizes to this region as part of the cytoplasmic polyadenylation machinery. This hypothesis can now also be tested in cells where Symplekin localization to the tight junction has been observed.

Overall, our characterization of Symplekin during development indicates that it is an essential gene with multiple *in vivo* roles. Our results also imply that the role of Symplekin at the tight junction is possibly related to RNA biology. Expression of Symplekin mutant proteins and characterization of developmental, mRNA and localization phenotypes will help dissect its various roles in gene expression. This work provides a foundation for such studies.

ACKNOWLEDGMENTS

I thank James Wilhelm for kindly sending the YPS antibody and mutant flies. From the Marzluff lab, I thank Xiao Yang for the GST-CstF-64 and GST protein, Mindy Steiniger for the CPSF-30 depleted RNA sample and Brandon Burch for H2a total and H2a readthrough primers.

CHAPTER 5: OVERALL CONCLUSIONS AND SIGNIFICANCE

Every eukaryotic cell division requires synthesis of new histone protein to package the newly replicated genome. The unique 3' end of metazoan histone mRNA mediates precise expression of this special class of genes. How the cell coordinates biosynthesis of these distinct transcripts for all five classes of histone proteins is still unknown. In this thesis, I approached this question from two directions. One goal was to determine the contribution of the HLB to histone mRNA expression. My other aim was to understand how general 3' end formation factors assemble into the distinct histone pre-mRNA processing machinery. Using *Drosophila* as a model system, I tested the contribution of the histone locus in HLB assembly and expression (Chapter 2), asked if localizing a histone pre-mRNA processing factor to the HLB influenced 3' end formation (Chapter 3) and established a system to test Symplekin function in histone pre-mRNA processing (Chapter 4). From these studies, I demonstrate a role for the histone locus in HLB maturation and consequently show that formation of the nuclear body is tied to histone mRNA expression. I also show that raising the local concentration of FLASH at the site of histone mRNA synthesis increases the rate of histone pre-mRNA processing. These observations, and my characterization of Symplekin during development, advance our understanding of how the cell coordinates gene expression through nuclear organization.

The Histone Locus Participates in HLB Assembly

The highly expressed *Drosophila* replicative-histone genes are grouped and repeated 100 times on the second chromosome and have a distinct mechanism of 3' end formation. It is therefore possible that the HLB forms as a consequence of robust, clustered transcription, which would require the recruitment of processing factors to the site of transcription. Yet, studies of the HLB during the cell cycle as well as in embryogenesis uncouple HLB formation and histone mRNA expression. Two factors presumed to be involved in transcription, Mxc and Mute, and two pre-mRNA processing components, FLASH and U7snRNP, are present in the HLB throughout the cell cycle, and the HLB persists outside of S phase and when cells are not proliferating or expressing histone mRNA (White et al., 2007; White et al., 2011). This suggests that participants in histone mRNA biosynthesis are

In Chapter 2, I show how the histone locus plays an essential role in forming the HLB. This study defined a sequence encompassing the *H3-H4* promoter and corresponding 5'UTRs that contributed to HLB assembly in two ways. First, a *cis* element(s) within this region stabilized an HLB intermediate, containing Mxc and FLASH, which I term the proto-HLB. Consistent with these two factors accumulating during an early phase of HLB assembly, FLASH and Mxc co-localize in cycle 10 of embryogenesis, one cell cycle before the onset of zygotic histone mRNA transcription (White et al., 2011). Currently, FLASH and Mxc are the only known HLB components enriched at the histone locus at this stage of development. Second, transcription initiation from the *H3-H4* promoter was required for HLB maturation, but no sequences in the histone genes, including the histone 3' end signals were required for assembly of a complete HLB. This activity coincided with maturation of the HLB and was required for the activation of the neighboring *H2a-H2b* genes present in the cluster. These observations define a pathway for HLB assembly that both implicates the locus as a physical platform for building the structure and provides insight into the

relationship between HLB assembly and histone gene transcriptional activation. Adding the *H3-H4* promoter to our model of HLB formation as a mediator of the transition from proto to mature HLB opens new questions regarding HLB biogenesis and its participation in histone gene expression.

Currently, we define the proto-HLB as a nuclear body containing FLASH and Mxc that lacks U7snRNP and Mute. Detection of FLASH-Mxc foci in the absence of the histone locus suggests that a specific DNA sequence in the locus is not essential for initial formation of the proto-HLB, which must be triggered by an unknown event. This HLB intermediate, however, is unstable in the absence of the histone genes as it is not detected 5 to 6 cell cycles later in germ band extended mutant embryos. Preventing activity from the H3-H4 promoter through mutation of the TATAA box, however, stabilized this FLASH-Mxc intermediate, suggesting that a *cis* element in the H3-H4 promoter is a component of the proto-HLB.

Although the H3-H4 promoter scaffolds HLB assembly, FLASH and or Mxc could possibly localize to other regions of the genome. In the histone deletion animals, we observed pairing of the unstable FLASH-Mxc foci during development, suggesting that they are associated with chromatin. Identifying additional gene targets of these factors in both WT and histone deletion animals could be determined by ChIP–seq analysis. This approach would test the hypothesis that these factors localize to other genes, and could also provide additional sequences that recruit HLB factors for comparison with the *H3-H4* promoter. This type of experiment would aid in interpreting FLASH and Mxc mutant phenotypes as well as contributing to our understanding of DNA motifs that nucleate the HLB.

It is likely that the proto-HLB contains additional components. Neither FLASH nor Mxc have been shown to directly bind DNA, so it is possible that a factor that directly

interacts with DNA participates in this early stage of HLB assembly. Testing additional transgenes with portions of the H3-H4 promoter deleted or mutated with the ectopic HLB assembly system could identify which region(s) nucleate the body. This approach could reveal a binding motif for potential transcription factors. Narrowing down the essential sequence could also result in identification of a piece of DNA that is small enough for biochemical approaches to identify DNA binding factors. For example, it is possible that a biotinylated DNA oligonucleotide containing the HLB assembly sequence could allow isolation of components of the HLB for identification through mass spectrometry. Adding to the collection of known HLB components could provide insight into the mechanisms of both transcription and processing of histone mRNA.

Our characterization of the *H3-H4* promoter also indicated that transcription initiation participates in HLB maturation. In addition to detection of Mute and U7 snRNP, staining of both the early embryo (White et al., 2011), and our ectopic HLB assembly analysis indicate that FLASH and Mxc detection is significantly more consistent and robust after transcription initiation. One interpretation of this increased signal is spreading of the HLB, which could be mediated by an actively moving polymerase, modification of the adjacent chromatin structure or signal in the 5'UTR of the pre-mRNA. Our minimal DNA fragment required for HLB nucleation contained the 5' UTRs of the *H3* and *H4* genes. It is therefore not known if specific *cis* elements in the UTR of these RNA contribute to HLB formation. This could be tested through assessment of HLB assembly on a DNA fragment lacking the UTR sequence. This experiment would focus subsequent analysis of HLB assembly. In one scenario, where RNA sequence plays no role, assembly of transcription factors and initiation could trigger HLB assembly. Additionally, transcription would physically alter the histone locus as nucleosomes are moved or modified to accommodate the oncoming polymerase. If the 5' UTR sequences were necessary, TATA box mutation would indicate if the *cis* element

directed HLB assembly through DNA or RNA. If the RNA was deemed necessary, it is possible that interaction between the mRNA 5' UTR and an unknown element could mediate the transition to a mature HLB. These future experiments will add details to the mechanism of HLB assembly as well as expand our grasp of histone gene expression.

While it is theoretically possible that the 5' UTR of histone mRNA participates in HLB assembly, our analysis shows that recruitment of histone pre-mRNA processing factors to the HLB does not require the *cis* elements of the 3' UTR. This demonstrates that at least some components of the histone processing machinery is recruited to the histone genes upstream of transcribing the *cis* elements that assemble the complex.

Overall, our studies demonstrate the role of the histone locus in assembling the HLB. We have refined our understanding of the components required for HLB formation. Identifying the critical contribution of transcription to this processes further supports a model where the HLB assembles in distinct, hierarchical steps. We also show that HLB assembly is critical for histone locus activation, indicating that the nuclear body forms to regulate gene expression.

The HLB Coordinates Histone Gene Expression and Ensures Efficient 3' End Formation

The interplay between nuclear body assembly and its associated biology presents a challenge to assessing how the nuclear body contributes to a reaction. Detailed analyses of how the body assembles and is maintained are an essential foundation for such studies. During the course of this thesis, I have shown two functions of the HLB: coordinate expression of the core histone genes and ensuring efficient 3' end formation.

In the absence of formation of a mature HLB, (when the TATAA boxes are mutated in the *H3-H4* promoter) neither the *H3-H4* genes, not the histone *H2a-H2b* genes are expressed, although a proto HLB is formed. This correlation suggests a connection between HLB maturation and gene activation. If possible, defining which aspect of transcription induces recruitment of Lsm11 and Mute, and determining if that process precedes *H2a-H2b* activation will further our understanding of how the HLB contributes to histone mRNA expression.

In Chapter 3, I provide evidence that concentration of FLASH and Lsm11 in the HLB is important for the efficiency of 3' end formation, demonstrating that nuclear bodies can function by raising the local concentration of factors to increase the rate of a biochemical reaction. In our system, I found that when FLASH was present in the nucleus, but not concentrated in the HLB, nascent, un-processed histone mRNA transcripts accumulated. The levels of the histone mRNA in the cells was normal. This suggests that the majority of the histone mRNA is, ultimately, properly processed. We interpret the presence of these longer precursor mRNAs as an indication of a slower rate of histone 3' end formation, as RNA polymerase has time to continue transcribing the substrate of the histone pre-mRNA processing machinery. More surprising was that there was very little polyadenylation of histone mRNA under these conditions.

There are at least two feasible mechanisms for generating processed histone mRNA in animals expressing mislocalized FLASH. One possibility is that histone pre-mRNA processing occurs on an RNA substrate of similar length to that in wildtype. During transcription the RNA polymerase stalls at a site after transcription of the HDE (Adamson and Price, 2003), and normally terminates at or near that site based on the lack of RNA polymerase detected by CHIP-Seq in the intergenic regions. Without FLASH readily available at the site of the reaction, at some frequency, the RNA polymerase releases from

the pause and continues transcribing. The longer transcripts could then be polyadenylated, generating misprocessed histone mRNA, or could generate non functional read through transcripts that are never processed and rapidly degraded, or could generate longer pre-mRNAs that are ultimately processed. My results suggests that the third possibility occurs. There is not extensive polyadenylation of the histone mRNA, as occurs in mutants in the U7 snRNP or SLBP, and the levels of normal histone mRNA that accumulates is close to normal. This result suggests that polyadenylation is prevented on the read-through transcripts, and that this pre-mRNA is still recognized by the histone processing machinery and is cleaved between the stem loop and HDE. Alternatively the histone mRNAs produced with low efficiency could have a much longer half-life than normal. Analysis of the ultimate fate of the downstream cleavage products would prove which mechanism predominates. These transcripts are rapidly degraded, but in exosome mutants they would be much more stable, and should accumulate if they are not ultimately processed.

Cleavage of the longer transcript (and the lack of polyadenylated histone mRNA in these genes, suggests suppression of the poly(A) signals. Misprocessed histone mRNA are detected when the histone processing machinery is disrupted (Godfrey et al., 2006; Godfrey et al., 2009; Lanzotti et al., 2002; Sullivan et al., 2001; Wagner et al., 2007), including when the levels of FLASH are greatly reduced by RNAi (Burch et al., 2011) . In contrast, when FLASH is mis-localized, all of the factors in the cell are competent for the processing reaction, but there is not a high concentration of FLASH at the site of histone pre-mRNA transcription. One possibility is that SLBP and the U7snRNP associate normally with the pre-mRNA, and that this association commits the transcript to processing just after the stem loop, and somehow blocks the polyadenylation reaction. When FLASH cannot interact with Lsm11 or the HCC, this reaction fails and the cleavage and polyadenylation machinery ultimately generate the 3' end of the mRNA. When FLASH is present in the cell with an

intact processing domain, proper histone 3' end formation can occur, but it is delayed for a few seconds (the stall time is not known but the time required for transcription of this region is <10 seconds, assuming 25 nts/sec. Stalling can't be too long or the levels of histone mRNA would go down because of blocking new transcription). This model also suggests that recruitment of the cleavage and polyadenylation machinery to histone mRNA is slow, or that the presence of the SLBP and U7 snRNP on the transcript interferes with this reaction.

Commitment to a cleavage site through association with trans factors has been observed for mRNA controlled by alternative polyadenylation, such as ELAV determined site selection in *Drosophila* (Hilgers et al., 2012). It is possible that for histone mRNA, the timing of this event is critical for proper 3' end formation. In a *Drosophila H2aV* mutant or after H2aV depletion by RNAi, Lsm11 does not accumulate in the HLB (Wagner et al., 2007). Consequently, misprocessed histone mRNA accumulates even though Lsm11 levels are not altered, and extracts made from the depleted cells processes a synthetic pre-mRNA substrate (Wagner et al., 2007). This observation indicates that localization of Lsm11 to the HLB is critical for efficient *in vivo* histone 3' end formation, possibly through committing the transcript to cleavage between the stem loop and HDE.

These two sets of experiments provide compelling evidence for HLB function in both histone mRNA transcription and 3' end formation. While the observation that HLB maturation coincides with locus activation is currently a correlation, identification of additional factors contributing to this transition will provide insight to the contribution by the HLB. Our understanding of FLASH function in histone pre-mRNA processing as well as HLB assembly allowed us to test the consequences of mis-localizing this critical component while maintaining an intact nuclear body. In addition to establishing the phenotype of un-processed histone mRNA in this mutant background, we also found that mis-localized FLASH also sensitized the animal to hypomorphic perturbations of the processing

machinery such as FLASH^{LDIY71} and our V5-tagged Lsm11 rescue line (data not shown). This genetic background provides a system for uncovering additional contributions to histone pre-mRNA processing that, under wildtype circumstance, would go unnoticed.

The Histone Replacement Platform as a System for Understanding *cis* Control of Histone mRNA Biosynthesis

Metazoans carry multiple copies of each histone gene Marzluff et al., 2008), therefore determining *in vivo* requirements for *cis* elements within a gene, or the locus is not trivial. Recently, development of a strategy in *Drosophila* enables these types of studies by taking advantage of the fact that all the replicative histone genes in this animal are clustered (Gunesdogan et al., 2010). The 100 copies (haploid) of the histone repeat can be deleted and a removal of this locus can be replaced by two genomic insertions containing 3 histone cluster repeats (resulting twelve histone gene clusters at two ectopic sites), and expression from these transgenes results in viable, fertile flies. This powerful system provides a tool for understanding histone modifications as well as testing and extending the observations made in this thesis.

The ability to replace the entire histone locus with 12 copies, indicates that the entire 100 copy locus is not essential. Since heterozygotes containing a histone deletion are also completely viable and fertile, 100 total copies is clearly sufficient. It is not known how many of the 100 copies of the histone genes are expressed (or whether the number expressed varies in different tissues. The length of S-phase varies in different cell types, and the multiple histone genes may all be expressed in cells with the shortest S-phase. It is also possible that endoreduplicating cells (salivary gland, nurse cells in the ovary), may underreplicate the histone locus resulting in effectively reducing the number of histone genes.

In Chapter 2, we demonstrated that the *H3-H4* promoter nucleated the HLB and was required for activation of the *H2a-H2b* gene pair. This observation was made at an ectopic locus in an animal that contained wildtype histone gene clusters on the second chromosome. This platform allows us to determine necessity. We showed that the activity of this promoter could be transferred to the *H2a-H2b* genes, so would there be a consequence to having this promoter activity driving all five histone genes? The promoter of *H1* does not contain a TATA box, and expression of this gene is mediated through TRF2 (Isogai et al., 2007), so it is possible that there would be consequences to altering regulation of this gene. As sequences within the histone locus driving HLB assembly and transcription of the mRNA are identified, there is now a genetic system for testing their role in the animal.

In *Drosophila*, cryptic polyadenylation sites follow the canonical processing signals for every histone gene. As these genes are organized in a compact, in some cases convergent array, it is possible that transcription termination is essential, and these signals serve to ensure termination before the polymerase reaches the next gene. Read through transcription could result in the generation of double stranded RNA and possibly gene silencing. This could be tested through mutation of the cryptic polyadenylation signals.

Changing the poly(A) signals following each histone 3'UTR could also affect histone pre-mRNA processing. As discussed in the previous section, we interpret the lack of mis-processed histone mRNA in our FLASH localization mutant as an indication of slow recruitment and or activation of the cleavage and polyadenylation machinery to the histone pre-mRNA. This could be due to the fact that the signals are not the preferred AAUAAA sequence. This system could test the ability of a stronger poly(A) signal to compete with the histone *cis* elements. If done in the mis-localized FLASH mutant background, it is possible that more polyadenylated histone mRNA would accumulate. It is also possible that SLBP

and U7snRNA binding upstream of the poly(A) sites maintain cleavage through the canonical pathway.

A final example of a sequence element that could be tested with this system is the RNA polymerase pause site(s) that were identified just after the HDE (Adamson and Price, 2003). Examination of *in vitro* transcription reactions indicated that when processing was inhibited, longer, relatively uniform RNA transcripts accumulated, suggesting pausing of RNA polymerase 15nt downstream of the HDE. This pause could enhance SLBP and U7snRNA binding to the pre-mRNA and could also ensure efficient transcription termination, as slowing the polymerase would shorten the length of the downstream cleavage product. While RNA polymerase is not detected in the intergenic downstream region of each histone gene (Celniker et al., 2009; Sullivan et al., 2009), it is not known if the pause site mediates efficient termination. Again, the contribution of this *cis* element is now testable with the Histone Gene Replacement Platform.

Assembly of the Histone Processing Complex *in vivo*

How the general components of 3' end formation interact with specific histone pre-mRNA processing factors remains elusive. In Chapter 3, we mutated various regions of FLASH implicated in histone pre-mRNA *in vitro* and in cell culture (Burch et al., 2011; Sabath et al., 2013; Yang et al., 2013). While mutation of the Lsm11 binding region debilitated FLASH and could not rescue the mutant phenotype, changing the LDIY71 motif to alanines supported FLASH function. This discrepancy between the *in vitro* and *in vivo* requirement of these residues aided in uncovering a role for concentrating FLASH in the HLB, but also highlighted the challenge of studying the histone processing machinery *in vivo*. Characterization of the LDIY71 motif also unveiled a contribution of LDIY45 residues in histone pre-mRNA

processing when combined with mutation of LDIY71. Alone, mutation of LDIY45 had no effect on FLASH function in histone pre-mRNA, and its contribution, if any, under wildtype conditions remains unknown.

Although we have not yet generated separation of function alleles for Symplekin, our *in vivo* analysis of FLASH emphasizes the critical final step of determining the function of an identified domain in an animal. Therefore, as shown in Chapter 4, our characterization of Symplekin mutant phenotypes across an allelic series will provide a background for testing candidate mutations of Symplekin protein. The observation that Symplekin localizes to the HLB in the germline suggests that there is a mechanism for recruiting Symplekin to the histone locus independently of the nascent transcript. We hypothesize that this is through specific residues in Symplekin. Therefore, testing candidate interactions between Symplekin and other 3' end formation could identify how Symplekin assembles into the histone processing complex.

One such candidate interaction is Symplekin and CstF-64. Complementation of various versions of Symplekin protein in cells where endogenous Symplekin was depleted indicated that this association was required for histone pre-mRNA processing, but dispensable for cleavage and polyadenylation (Ruepp et al., 2011). The location of CstF-64 binding to Symplekin is conserved between humans and the fly. If this interaction is critical for histone pre-mRNA processing and dispensable for polyadenylation, we hypothesize that the levels of polyadenylated histone mRNA would increase. We also predict restoration of wild type cleavage and polyadenlyation of mRNAs such as RP49 and SOP. Mindful of how the animal tolerates disruption of processing histone pre-mRNA, we could also carry out such experiments in the mis-localized FLASH background to test for subtle contributions to histone pre-mRNA processing.

Despite a dispensable role *in vivo*, biochemical approaches to determine the protein that interacts with LDIY71 will advance our understanding of how the histone pre-mRNA processing complex assembles. While binding under stringent conditions implicates the Symplekin/CstF64 as the interacting protein, it is possible that an unknown component of the histone pre-mRNA processing machinery remains to be discovered, which binds to the Symplekin/CstF64 complex. Combining biochemical and molecular characterization of processing factors with functional assays to determine genetic and developmental requirements will provide a comprehensive overview of histone pre-mRNA processing.

Concluding Remarks

The goal of this thesis was to examine the relationship between nuclear architecture and gene expression. By focusing on a nuclear body and expression of a unique class of genes, we have enhanced our understanding of HLB assembly and consequently histone mRNA expression. Production of histone mRNA is coordinated and restricted to S phase of the cell cycle, and we have shown that a sequence in the histone locus nucleates the proto-HLB. Maturation of this nuclear body, as well as activation of the neighboring *H2a-H2b* gene pair, requires transcription initiation from the *H3-H4* bidirectional promoter, and this correlation suggests a possible mechanism for precise activation of the locus.

Characterization of two histone pre-mRNA processing factors in the animal, FLASH and Symplekin, advanced our understanding assembly of the *in vivo* 3' end formation machinery. Notably, we provide explicit experimental evidence that localizing a protein, FLASH, to a nuclear body, the HLB, increases the rate of a reaction, histone pre-mRNA processing. A number of models for nuclear body assembly exist and the biological reactions associated with individual nuclear bodies vary. That said, we propose that the

themes highlighted by the experiments of this thesis can inform function of other nuclear bodies and mechanisms of mRNA biosynthesis.

REFERENCES

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

Adamson, T.E., Shutt, D.C., and Price, D.H. (2005). Functional coupling of cleavage and polyadenylation with transcription of mRNA. J. Biol. Chem. *280,* 32262-32271.

An, S., Kumar, R., Sheets, E.D., and Benkovic, S.J. (2008). Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. Science *320,* 103-106.

Anamika, K., Gyenis, A., Poidevin, L., Poch, O., and Tora, L. (2012). RNA polymerase II pausing downstream of core histone genes is different from genes producing polyadenylated transcripts. PLoS One *7,* e38769.

Avgousti, D.C., Palani, S., Sherman, Y., and Grishok, A. (2012). CSR-1 RNAi pathway positively regulates histone expression in C. elegans. EMBO J. *31,* 3821-3832.

Aviv, H., and Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. U. S. A. *69,* 1408-1412.

Balda, M.S., and Matter, K. (2000). The tight junction protein ZO-1 and an interacting transcription factor regulate ErbB-2 expression. EMBO J. *19,* 2024-2033.

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

Barnard, D.C., Ryan, K., Manley, J.L., and Richter, J.D. (2004). Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. Cell *119,* 641-651.

Bateman, J.R., and Wu, C.T. (2007). DNA replication and models for the origin of piRNAs. Bioessays *29,* 382-385.

Benoit, B., Juge, F., Iral, F., Audibert, A., and Simonelig, M. (2002). Chimeric human CstF-77/Drosophila Suppressor of forked proteins rescue suppressor of forked mutant lethality and mRNA 3' end processing in Drosophila. Proc. Natl. Acad. Sci. U. S. A. *99,* 10593-10598.

Bernardi, R., and Pandolfi, P.P. (2007). Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. Nat. Rev. Mol. Cell Biol. *8,* 1006-1016.

Birchmeier, C., Grosschedl, R., and 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.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. U. S. A. *104,* 3312-3317.

Boisvert, F.M., van Koningsbruggen, S., Navascues, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat. Rev. Mol. Cell Biol. *8,* 574-585.

Bongiorno-Borbone, L., De Cola, A., Barcaroli, D., Knight, R.A., Di Ilio, C., Melino, G., and De Laurenzi, V. (2010). FLASH degradation in response to UV-C results in histone locus bodies disruption and cell-cycle arrest. Oncogene *29,* 802-810.

Bongiorno-Borbone, L., De Cola, A., Vernole, P., Finos, L., Barcaroli, D., Knight, R.A., Melino, G., and De Laurenzi, V. (2008). FLASH and NPAT positive but not Coilin positive Cajal Bodies correlate with cell ploidy. Cell. Cycle *7,* 2357-2367.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324,* 1729-1732.

Brangwynne, C.P., Mitchison, T.J., and Hyman, A.A. (2011). Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc. Natl. Acad. Sci. U. S. A. *108,* 4334-4339.

Bulchand, S., Menon, S.D., George, S.E., and Chia, W. (2010). Muscle wasted: a novel component of the Drosophila histone locus body required for muscle integrity. J. Cell. Sci. *123,* 2697-2707.

Burch, B.D., Godfrey, A.C., Gasdaska, P.Y., Salzler, H.R., Duronio, R.J., Marzluff, W.F., and Dominski, Z. (2011). Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila. RNA *17,* 1132-1147.

Callan, H.G., and Gall, J.G. (1991). Association of RNA with the B and C snurposomes of Xenopus oocyte nuclei. Chromosoma *101,* 69-82.

Calvi, B.R., Lilly, M.A., and Spradling, A.C. (1998). Cell cycle control of chorion gene amplification. Genes Dev. *12,* 734-744.

Carmo-Fonseca, M., and Rino, J. (2011). RNA seeds nuclear bodies. Nat. Cell Biol. *13,* 110-112.

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M.*, et al.* (2009). Unlocking the secrets of the genome. Nature *459,* 927-930.

Chen, F., MacDonald, C.C., and Wilusz, J. (1995). Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acids Res. *23,* 2614-2620.

Chen, M., Galvao, R.M., Li, M., Burger, B., Bugea, J., Bolado, J., and Chory, J. (2010). Arabidopsis HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. Cell *141,* 1230-1240.

Cheng, B., Li, T., Rahl, P.B., Adamson, T.E., Loudas, N.B., Guo, J., Varzavand, K., Cooper, J.J., Hu, X., Gnatt, A., Young, R.A., and Price, D.H. (2012). Functional association of Gdown1 with RNA polymerase II poised on human genes. Mol. Cell *45,* 38-50.

Christofori, G., and Keller, W. (1989). Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. Mol. Cell. Biol. *9,* 193-203.

Christofori, G., and Keller, W. (1988). 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP. Cell *54,* 875-889.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature *389,* 399-402.

Decker, C.J., Teixeira, D., and Parker, R. (2007). Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in Saccharomyces cerevisiae. J. Cell Biol. *179,* 437-449.

Decker, M., Jaensch, S., Pozniakovsky, A., Zinke, A., O'Connell, K.F., Zachariae, W., Myers, E., and Hyman, A.A. (2011). Limiting amounts of centrosome material set centrosome size in C. elegans embryos. Curr. Biol. *21,* 1259-1267.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet. *14,* 390-403.

Denissov, S., Lessard, F., Mayer, C., Stefanovsky, V., van Driel, M., Grummt, I., Moss, T., and Stunnenberg, H.G. (2011). A model for the topology of active ribosomal RNA genes. EMBO Rep. *12,* 231-237.

Deryusheva, S., and Gall, J.G. (2009). Small Cajal body-specific RNAs of Drosophila function in the absence of Cajal bodies. Mol. Biol. Cell *20,* 5250-5259.

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

Dominski, Z., Yang, X.C., Purdy, M., Wagner, E.J., and Marzluff, W.F. (2005). A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. Mol. Cell. Biol. *25,* 1489-1500.

Dominski, Z., Zheng, L.X., Sanchez, R., and Marzluff, W.F. (1999). Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. Mol. Cell. Biol. *19,* 3561-3570.

Dundr, M. (2011). Seed and grow: a two-step model for nuclear body biogenesis. J. Cell Biol. *193,* 605-606.

Dundr, M., Hebert, M.D., Karpova, T.S., Stanek, D., Xu, H., Shpargel, K.B., Meier, U.T., Neugebauer, K.M., Matera, A.G., and Misteli, T. (2004). In vivo kinetics of Cajal body components. J. Cell Biol. *164,* 831-842.

Fong, K.W., Li, Y., Wang, W., Ma, W., Li, K., Qi, R.Z., Liu, D., Songyang, Z., and Chen, J. (2013). Whole-genome screening identifies proteins localized to distinct nuclear bodies. J. Cell Biol. *203,* 149-164.

Frey, M.R., Bailey, A.D., Weiner, A.M., and Matera, A.G. (1999). Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. Curr. Biol. *9,* 126-135.

Frey, M.R., and Matera, A.G. (2001). RNA-mediated interaction of Cajal bodies and U2 snRNA genes. J. Cell Biol. *154,* 499-509.

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

Fung, J.C., Marshall, W.F., Dernburg, A., Agard, D.A., and Sedat, J.W. (1998). Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations. J. Cell Biol. *141,* 5-20.

Gall, J.G. (2000). Cajal bodies: the first 100 years. Annu. Rev. Cell Dev. Biol. *16,* 273-300.

Gall, J.G., Stephenson, E.C., Erba, H.P., Diaz, M.O., and Barsacchi-Pilone, G. (1981). Histone genes are located at the sphere loci of newt lampbrush chromosomes. Chromosoma *84,* 159-171.

Galli, G., Hofstetter, H., Stunnenberg, H.G., and Birnstiel, M.L. (1983). Biochemical complementation with RNA in the Xenopus oocyte: a small RNA is required for the generation of 3' histone mRNA termini. Cell *34,* 823-828.

Gao, L., Frey, M.R., and Matera, A.G. (1997). Human genes encoding U3 snRNA associate with coiled bodies in interphase cells and are clustered on chromosome 17p11.2 in a complex inverted repeat structure. Nucleic Acids Res. *25,* 4740-4747.

Gebauer, F., and Richter, J.D. (1995). Cloning and characterization of a Xenopus poly(A) polymerase. Mol. Cell. Biol. *15,* 3460.

Ghazy, M.A., He, X., Singh, B.N., Hampsey, M., and Moore, C. (2009). The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing. Mol. Cell. Biol. *29,* 2296-2307.

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., and 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. *105,* 16964-16969.

Gick, O., Kramer, A., Keller, W., and Birnstiel, M.L. (1986). Generation of histone mRNA 3' ends by endonucleolytic cleavage of the pre-mRNA in a snRNP-dependent in vitro reaction. EMBO J. *5,* 1319-1326.

Gick, O., Kramer, A., Vasserot, A., and Birnstiel, M.L. (1987). Heat-labile regulatory factor is required for 3' processing of histone precursor mRNAs. Proc. Natl. Acad. Sci. U. S. A. *84,* 8937-8940.

Gil, A., and Proudfoot, N.J. (1987). Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. Cell *49,* 399-406.

Gil, A., and Proudfoot, N.J. (1984). A sequence downstream of AAUAAA is required for rabbit beta-globin mRNA 3'-end formation. Nature *312,* 473-474.

Gilmartin, G.M., and Nevins, J.R. (1989). An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev. *3,* 2180-2190.

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

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

Gunesdogan, U., Jackle, H., and Herzig, A. (2010). A genetic system to assess in vivo the functions of histones and histone modifications in higher eukaryotes. EMBO Rep. *11,* 772-776.

Gunjan, A., and Verreault, A. (2003). A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in S. cerevisiae. Cell *115,* 537-549.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., and McKnight, S.L. (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149,* 768-779.

Handwerger, K.E., and Gall, J.G. (2006). Subnuclear organelles: new insights into form and function. Trends Cell Biol. *16,* 19-26.

Heidemann, M., Hintermair, C., Voss, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim. Biophys. Acta *1829,* 55-62.

Hernandez-Verdun, D. (2006). The nucleolus: a model for the organization of nuclear functions. Histochem. Cell Biol. *126,* 135-148.

Hernandez-Verdun, D., Robert-Nicoud, M., Geraud, G., and Masson, C. (1991). Behaviour of nucleolar proteins in nuclei lacking ribosomal genes. A study by confocal laser scanning microscopy. J. Cell. Sci. *98 (Pt 1),* 99-105.

Hilgers, V., Lemke, S.B., and Levine, M. (2012). ELAV mediates 3' UTR extension in the Drosophila nervous system. Genes Dev. *26,* 2259-2264.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395,* 93-96.

Hofmann, I., Schnolzer, M., Kaufmann, I., and Franke, W.W. (2002). Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in Xenopus laevis oocytes. Mol. Biol. Cell *13,* 1665-1676.

Hu, J., Lutz, C.S., Wilusz, J., and Tian, B. (2005). Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA *11,* 1485-1493.

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

Ishov, A.M., Sotnikov, A.G., Negorev, D., Vladimirova, O.V., Neff, N., Kamitani, T., Yeh, E.T., Strauss, J.F.,3rd, and Maul, G.G. (1999). PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. J. Cell Biol. *147,* 221-234.

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

Kaiser, T.E., Intine, R.V., and Dundr, M. (2008). De novo formation of a subnuclear body. Science *322,* 1713-1717.

Karpen, G.H., Schaefer, J.E., and Laird, C.D. (1988). A Drosophila rRNA gene located in euchromatin is active in transcription and nucleolus formation. Genes Dev. *2,* 1745-1763.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J*., et al.* (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149,* 753-767.

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

Kavanagh, E., Buchert, M., Tsapara, A., Choquet, A., Balda, M.S., Hollande, F., and Matter, K. (2006). Functional interaction between the ZO-1-interacting transcription factor ZONAB/DbpA and the RNA processing factor symplekin. J. Cell. Sci. *119,* 5098-5105.

Kennedy, S.A., Frazier, M.L., Steiniger, M., Mast, A.M., Marzluff, W.F., and Redinbo, M.R. (2009). Crystal structure of the HEAT domain from the Pre-mRNA processing factor Symplekin. J. Mol. Biol. *392,* 115-128.

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

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T.*, et al.* (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471,* 480-485.

Klingauf, M., Stanek, D., and Neugebauer, K.M. (2006). Enhancement of U4/U6 small nuclear ribonucleoprotein particle association in Cajal bodies predicted by mathematical modeling. Mol. Biol. Cell *17,* 4972-4981.

Kohwi, M., Lupton, J.R., Lai, S.L., Miller, M.R., and Doe, C.Q. (2013). Developmentally regulated subnuclear genome reorganization restricts neural progenitor competence in Drosophila. Cell *152,* 97-108.

Kolev, N.G., and Steitz, J.A. (2005). Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. Genes Dev. *19,* 2583-2592.

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

Kugler, J.M., and Lasko, P. (2009). Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during Drosophila oogenesis. Fly (Austin) *3,* 15-28.

Laishram, R.S., and Anderson, R.A. (2010). The poly A polymerase Star-PAP controls 3'-end cleavage by promoting CPSF interaction and specificity toward the pre-mRNA. EMBO J. *29,* 4132-4145.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10,* R25-2009-10-3-r25. Epub 2009 Mar 4.

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

Leach, T.J., Mazzeo, M., Chotkowski, H.L., Madigan, J.P., Wotring, M.G., and Glaser, R.L. (2000). Histone H2A.Z is widely but nonrandomly distributed in chromosomes of Drosophila melanogaster. J. Biol. Chem. *275,* 23267-23272.

Lecuyer, E., Parthasarathy, N., and Krause, H.M. (2008). Fluorescent in situ hybridization protocols in Drosophila embryos and tissues. Methods Mol. Biol. *420,* 289-302.

Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131,* 174-187.

Lee, C., Zhang, H., Baker, A.E., Occhipinti, P., Borsuk, M.E., and Gladfelter, A.S. (2013). Protein aggregation behavior regulates cyclin transcript localization and cell-cycle control. Dev. Cell. *25,* 572-584.

Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev.

Lifton, R.P., Goldberg, M.L., Karp, R.W., and Hogness, D.S. (1978). The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. Cold Spring Harb. Symp. Quant. Biol. *42 Pt 2,* 1047-1051.

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

Liu, J.L., Wu, Z., Nizami, Z., Deryusheva, S., Rajendra, T.K., Beumer, K.J., Gao, H., Matera, A.G., Carroll, D., and Gall, J.G. (2009). Coilin is essential for Cajal body organization in Drosophila melanogaster. Mol. Biol. Cell *20,* 1661-1670.

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

MacDonald, C.C., Wilusz, J., and 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. Mol. Cell. Biol. *14,* 6647-6654.

Machyna, M., Heyn, P., and Neugebauer, K.M. (2013). Cajal bodies: where form meets function. Wiley Interdiscip. Rev. RNA *4,* 17-34.

Mais, C., Wright, J.E., Prieto, J.L., Raggett, S.L., and McStay, B. (2005). UBF-binding site arrays form pseudo-NORs and sequester the RNA polymerase I transcription machinery. Genes Dev. *19,* 50-64.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. Cell Mol. Life Sci. *65,* 1099-1122.

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

Mangone, M., Manoharan, A.P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S.D., Mis, E., Zegar, C., Gutwein, M.R., Khivansara, V.*, et al.* (2010). The landscape of C. elegans 3'UTRs. Science *329,* 432-435.

Mansfield, J.H., Wilhelm, J.E., and Hazelrigg, T. (2002). Ypsilon Schachtel, a Drosophila Y-box protein, acts antagonistically to Orb in the oskar mRNA localization and translation pathway. Development *129,* 197-209.

Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat. Cell Biol. *13,* 95-101.

Mao, Y.S., Zhang, B., and Spector, D.L. (2011). Biogenesis and function of nuclear bodies. Trends Genet. *27,* 295-306.

Marshall, W.F., Straight, A., Marko, J.F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A.W., Agard, D.A., and Sedat, J.W. (1997). Interphase chromosomes undergo constrained diffusional motion in living cells. Curr. Biol. *7,* 930-939.

Martin, F., Schaller, A., Eglite, S., Schumperli, D., and Muller, B. (1997). The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. EMBO J. *16,* 769-778.

Marzluff, W.F. (2010). Terminating histone synthesis to preserve centromere integrity. Dev. Cell. *18,* 335-336.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. *9,* 843-854.

Matera, A.G., Izaguire-Sierra, M., Praveen, K., and Rajendra, T.K. (2009). Nuclear bodies: random aggregates of sticky proteins or crucibles of macromolecular assembly? Dev. Cell. *17,* 639-647.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385,* 357-361.

McKay, D.J., and Lieb, J.D. (2013). A common set of DNA regulatory elements shapes Drosophila appendages. Dev. Cell. *27,* 306-318.

McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B. (1985). The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucleic Acids Res. *13,* 1347-1368.

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

Mellman, D.L., Gonzales, M.L., Song, C., Barlow, C.A., Wang, P., Kendziorski, C., and Anderson, R.A. (2008). A PtdIns4,5P2-regulated nuclear poly(A) polymerase controls expression of select mRNAs. Nature *451,* 1013-1017.

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J.*, et al.* (2006). Nuclear pore components are involved

in the transcriptional regulation of dosage compensation in Drosophila. Mol. Cell *21,* 811-823.

Miele, A., Braastad, C.D., Holmes, W.F., Mitra, P., Medina, R., Xie, R., Zaidi, S.K., Ye, X., Wei, Y., Harper, J.W.*, et al.* (2005). HiNF-P directly links the cyclin E/CDK2/p220NPAT pathway to histone H4 gene regulation at the G1/S phase cell cycle transition. Mol. Cell. Biol. *25,* 6140-6153.

Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. Cell *128,* 787-800.

Misteli, T. (2001). The concept of self-organization in cellular architecture. J. Cell Biol. *155,* 181-185.

Moore, C.L., Chen, J., and Whoriskey, J. (1988). Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding. EMBO J. *7,* 3159-3169.

Moore, C.L., and Sharp, P.A. (1985). Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell *41,* 845-855.

Moore, G.D., Sinclair, D.A., and Grigliatti, T.A. (1983). Histone Gene Multiplicity and Position Effect Variegation in DROSOPHILA MELANOGASTER. Genetics *105,* 327-344.

Morimoto, M., and Boerkoel, C.F. (2013). The Role of Nuclear Bodies in Gene Expression and Disease. Biology (Basel) *2,* 976-1033.

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

Mullen, T.E., and 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.

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

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

Nagaoka, K., Udagawa, T., and Richter, J.D. (2012). CPEB-mediated ZO-1 mRNA localization is required for epithelial tight-junction assembly and cell polarity. Nat. Commun. *3,* 675.

Narita, T., Yung, T.M., Yamamoto, J., Tsuboi, Y., Tanabe, H., Tanaka, K., Yamaguchi, Y., and Handa, H. (2007). NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. Mol. Cell *26,* 349-365.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327,* 335-338.

Nizami, Z., Deryusheva, S., and Gall, J.G. (2010). The Cajal body and histone locus body. Cold Spring Harb Perspect. Biol. *2,* a000653.

Novotny, I., Blazikova, M., Stanek, D., Herman, P., and Malinsky, J. (2011). In vivo kinetics of U4/U6.U5 tri-snRNP formation in Cajal bodies. Mol. Biol. Cell *22,* 513-523.

Pandey, N.B., Williams, A.S., Sun, J.H., Brown, V.D., Bond, U., and 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. Mol. Cell. Biol. *14,* 1709-1720.

Paro, R. (2008). Mapping protein distributions on polytene chromosomes by immunostaining. CSH Protoc. *2008,* pdb.prot4714.

Pillai, R.S., Grimmler, M., Meister, G., Will, C.L., Luhrmann, R., Fischer, U., and 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.

Pillai, R.S., Will, C.L., Luhrmann, R., Schumperli, D., and Muller, B. (2001). Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. EMBO J. *20,* 5470-5479.

Prestel, M., Feller, C., Straub, T., Mitlohner, H., and Becker, P.B. (2010). The activation potential of MOF is constrained for dosage compensation. Mol. Cell *38,* 815-826.

Prieto, J.L., and McStay, B. (2008). Pseudo-NORs: a novel model for studying nucleoli. Biochim. Biophys. Acta *1783,* 2116-2123.

Pritchard, D.K., and Schubiger, G. (1996). Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes Dev. *10,* 1131-1142.

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. Curr. Opin. Cell Biol. *16,* 272-278.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev. *25,* 1770-1782.

Proudfoot, N.J. (1976). Sequence analysis of the 3' non-coding regions of rabbit alpha- and beta-globin messenger RNAs. J. Mol. Biol. *107,* 491-525.

Proudfoot, N.J., and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. Nature *263,* 211-214.

Proudfoot, N.J., and Longley, J.I. (1976). The 3' terminal sequences of human alpha and beta globin messenger RNAs: comparison with rabbit globin messenger RNA. Cell *9,* 733-746.

Raabe, T., Bollum, F.J., and Manley, J.L. (1991). Primary structure and expression of bovine poly(A) polymerase. Nature *353,* 229-234.

Raabe, T., Murthy, K.G., and Manley, J.L. (1994). Poly(A) polymerase contains multiple functional domains. Mol. Cell. Biol. *14,* 2946-2957.

Raja, S.J., Charapitsa, I., Conrad, T., Vaquerizas, J.M., Gebhardt, P., Holz, H., Kadlec, J., Fraterman, S., Luscombe, N.M., and Akhtar, A. (2010). The nonspecific lethal complex is a transcriptional regulator in Drosophila. Mol. Cell *38,* 827-841.

Rajendra, T.K., Praveen, K., and Matera, A.G. (2010). Genetic analysis of nuclear bodies: from nondeterministic chaos to deterministic order. Cold Spring Harb. Symp. Quant. Biol. *75,* 365-374.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29,* 24-26.

Rouget, C., Papin, C., Boureux, A., Meunier, A.C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature *467,* 1128-1132.

Ruepp, M.D., Schweingruber, C., Kleinschmidt, N., and Schumperli, D. (2011). Interactions of CstF-64, CstF-77, and symplekin: implications on localisation and function. Mol. Biol. Cell *22,* 91-104.

Ryan, K., Calvo, O., and Manley, J.L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA *10,* 565-573.

Ryan, K., Calvo, O., and Manley, J.L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA *10,* 565-573.

Ryner, L.C., Takagaki, Y., and Manley, J.L. (1989). Multiple forms of poly(A) polymerases purified from HeLa cells function in specific mRNA 3'-end formation. Mol. Cell. Biol. *9,* 4229-4238.

Sabath, I., Skrajna, A., Yang, X.C., Dadlez, M., Marzluff, W.F., and Dominski, Z. (2013). 3'-End processing of histone pre-mRNAs in Drosophila: U7 snRNP is associated with FLASH and polyadenylation factors. RNA

Salinas, C.A., Sinclair, D.A., O'Hare, K., and Brock, H.W. (1998). Characterization of a Drosophila homologue of the 160-kDa subunit of the cleavage and polyadenylation specificity factor CPSF. Mol. Gen. Genet. *257,* 672-680.

Salzler, H.R., Tatomer, D.C., Malek, P.Y., McDaniel, S.L., Orlando, A.N., Marzluff, W.F., and Duronio, R.J. (2013). A sequence in the Drosophila H3-H4 Promoter triggers histone locus body assembly and biosynthesis of replication-coupled histone mRNAs. Dev. Cell. *24,* 623-634.

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

Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T., and Luhrmann, R. (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. EMBO J. *23,* 3000-3009.

Schaufele, F., Gilmartin, G.M., Bannwarth, W., and Birnstiel, M.L. (1986). Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA. Nature *323,* 777-781.

Shevtsov, S.P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. Nat. Cell Biol. *13,* 167-173.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R.,3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Mol. Cell *33,* 365-376.

Smith, K.P., Carter, K.C., Johnson, C.V., and Lawrence, J.B. (1995). U2 and U1 snRNA gene loci associate with coiled bodies. J. Cell. Biochem. *59,* 473-485.

Stanek, D., and Neugebauer, K.M. (2004). Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. J. Cell Biol. *166,* 1015-1025.

Strub, K., and Birnstiel, M.L. (1986). Genetic complementation in the Xenopus oocyte: co-expression of sea urchin histone and U7 RNAs restores 3' processing of H3 pre-mRNA in the oocyte. EMBO J. *5,* 1675-1682.

Strub, K., Galli, G., Busslinger, M., and Birnstiel, M.L. (1984). The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing. EMBO J. *3,* 2801-2807.

Strzelecka, M., Trowitzsch, S., Weber, G., Luhrmann, R., Oates, A.C., and Neugebauer, K.M. (2010). Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. Nat. Struct. Mol. Biol. *17,* 403-409.

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

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

Sullivan, K.D., Steiniger, M., and 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. Mol. Cell *34,* 322-332.

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

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1989). Four factors are required for 3'-end cleavage of pre-mRNAs. Genes Dev. *3,* 1711-1724.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell *52,* 731-742.

Tan, D., Marzluff, W.F., Dominski, Z., and Tong, L. (2013). Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hExo ternary complex. Science *339,* 318-321.

Tootle, T.L., Williams, D., Hubb, A., Frederick, R., and Spradling, A. (2011). Drosophila eggshell production: identification of new genes and coordination by Pxt. PLoS One *6,* e19943.

Tucker, K.E., Berciano, M.T., Jacobs, E.Y., LePage, D.F., Shpargel, K.B., Rossire, J.J., Chan, E.K., Lafarga, M., Conlon, R.A., and Matera, A.G. (2001). Residual Cajal bodies in coilin knockout mice fail to recruit Sm snRNPs and SMN, the spinal muscular atrophy gene product. J. Cell Biol. *154,* 293-307.

van Daal, A., and Elgin, S.C. (1992). A histone variant, H2AvD, is essential in Drosophila melanogaster. Mol. Biol. Cell *3,* 593-602.

Van de Vosse, D.W., Wan, Y., Wozniak, R.W., and Aitchison, J.D. (2011). Role of the nuclear envelope in genome organization and gene expression. Wiley Interdiscip. Rev. Syst. Biol. Med. *3,* 147-166.

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

Wahle, E. (1991). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. Cell *66,* 759-768.

Walker, M.P., Tian, L., and Matera, A.G. (2009). Reduced viability, fertility and fecundity in mice lacking the cajal body marker protein, coilin. PLoS One *4,* e6171.

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

Wei, Y., Jin, J., and 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.

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

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

Wilhelm, J.E., Mansfield, J., Hom-Booher, N., Wang, S., Turck, C.W., Hazelrigg, T., and Vale, R.D. (2000). Isolation of a ribonucleoprotein complex involved in mRNA localization in Drosophila oocytes. J. Cell Biol. *148,* 427-440.

Wilhelm, J.E., and Smibert, C.A. (2005). Mechanisms of translational regulation in Drosophila. Biol. Cell *97,* 235-252.

Winters, M.A., and Edmonds, M. (1973). A poly(A) polymerase from calf thymus. Characterization of the reaction product and the primer requirement. J. Biol. Chem. *248,* 4763-4768.

Winters, M.A., and Edmonds, M. (1973). A poly(A) polymerase from calf thymus. Purification and properities of the enzyme. J. Biol. Chem. *248,* 4756-4762.

Wippich, F., Bodenmiller, B., Trajkovska, M.G., Wanka, S., Aebersold, R., and Pelkmans, L. (2013). Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. Cell *152,* 791-805.

Wu, C.H., and Gall, J.G. (1993). U7 small nuclear RNA in C snurposomes of the Xenopus germinal vesicle. Proc. Natl. Acad. Sci. U. S. A. *90,* 6257-6259.

Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M.M., Manley, J.L., and Tong, L. (2010). Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. Nature *467,* 729-733.

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

Yang, X.C., Sabath, I., Debski, J., Kaus-Drobek, M., Dadlez, M., Marzluff, W.F., and Dominski, Z. (2013). A complex containing the CPSF73 endonuclease and other polyadenylation factors associates with U7 snRNP and is recruited to histone pre-mRNA for 3'-end processing. Mol. Cell. Biol. *33,* 28-37.

Ye, X., Wei, Y., Nalepa, G., and Harper, J.W. (2003). The cyclin E/Cdk2 substrate p220(NPAT) is required for S-phase entry, histone gene expression, and Cajal body maintenance in human somatic cells. Mol. Cell. Biol. *23,* 8586-8600.

Zaessinger, S., Busseau, I., and Simonelig, M. (2006). Oskar allows nanos mRNA translation in Drosophila embryos by preventing its deadenylation by Smaug/CCR4. Development *133,* 4573-4583.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. *39,* 1512-1516.

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

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

Zhong, S., Muller, S., Ronchetti, S., Freemont, P.S., Dejean, A., and Pandolfi, P.P. (2000). Role of SUMO-1-modified PML in nuclear body formation. Blood *95,* 2748-2752.