12-2012

Functional Analysis of Drosophila Integrator Complex in snRNA 3' End Processing

Jiandong Chen

FUNCTIONAL ANALYSIS OF DROSOPHILA INTEGRATOR COMPLEX IN SNRNA 3' END PROCESSING

by

Jiandong Chen, M.S.

APPROVED:

_____
Eric J.Wagner, Ph.D.
Supervisory Professor

_____
Ambro van Hoof, Ph.D.

_____
Michelle C.Barton, Ph.D.

_____
Michael R.Blackburn, Ph.D.

_____
Xiaobing Shi, Ph.D.

APPROVED:

_____
Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

FUNCTIONAL ANALYSIS OF DROSOPHILA INTEGRATOR COMPLEX IN SNRNA 3' END PROCESSING

A

DISSERTATION

Presented to the Faculty of The University of Texas Health Science Center at Houston
and
The University of Texas MD Anderson Cancer Center
Graduate School of Biomedical Sciences
In Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

by

Jiandong Chen, M.S.

Houston, Texas

December, 2012

ACKNOWLEDGEMENTS

I would like to sincerely thank my mentor, Dr. Eric J. Wagner, for his constant help, support and encouragement during my graduate study. It was a great opportunity for me to work with him in the past four years. His enthusiasm for science, critical thinking, positive attitude and open personality leading to a friendly lab environment have all guaranteed an integrate training for me to become an independent scientist and will for sure benefit me throughout my life. I would like to also thank my committee members: Drs. Michael R. Blackburn, Michelle C. Barton, Ambro van Hoof, Xiaobing Shi, Joseph L. Alcorn, and Ann-Bin Shyu, for their contributions and advice which are indispensable.

I would like to thank the current and former members of Dr. Wagner's laboratory, Dr. Nader Ezzeddine, Dr. David Baillat, Dr. Chioniso P. Masamha, Todd Albrecht, Natoya Peart, Anupama Sataluri, Sarah May. Dr. Ezzeddine has been a great friend to me and guided me through at the beginning of my graduate study. Dr. Baillat has provided tremendous advice on my thesis. I would like to extend my gratitude to the faculty, staff, and students from the Department of Biochemistry and Molecular Biology for their help and kind sharing of reagents and equipment. I would like to thank Dr. Victoria Knutson, associate Dean for academic affairs, for her patience and kind help with all the paperworks. I would like to thank Dr. Dinghai Zheng, former Ph.D. student in Dr. Ann-Bin Shyu' lab, for helping me settle down at the very beginning.

A special thank goes to my beloved wife, Qing, for her dedication to my personal life as well as academic studies. Without her, nothing could be done smoothly. I would like to also thank my family members, especially my parents-in-law for their constant solicitude and encouragement. I am getting there!

FUNCTIONAL ANALYSIS OF DROSOPHILA INTEGRATOR COMPLEX IN SNRNA 3' END PROCESSING

Publication No._____
Jiandong Chen, M.S.
Supervisory Professor: Eric J. Wagner, Ph.D.

Uridine-rich small nuclear RNAs (U snRNAs) play essential roles in eukaryotic gene expression by facilitating the removal of introns from mRNA precursors and the processing of the replication-dependent histone pre-mRNAs. Formation of the 3' end of these snRNAs is carried out by a poorly characterized, twelve-membered protein complex named Integrator Complex.

In the effort to understand Integrator Complex function in the formation of the snRNA 3' end, we performed a functional RNAi screen in Drosophila S2 cells to identify protein factors required for snRNA 3' end formation. This screen was conducted by using a fluorescence-based reporter that elicits GFP expression in response to a deficiency in snRNA processing. Besides scoring the known Integrator subunits, we identified Asunder and CG4785 as additional core members of the Integrator Complex. Additionally, we also found a conserved requirement for Cyclin C and Cdk8 in both fly and human snRNA 3' end processing. We have further demonstrated that the kinase activity of Cdk8 is critical for snRNA 3' end processing and is likely to function independent of its well-documented function within the Mediator Cdk8 module. Taken together, this work functionally defines the Drosophila Integrator Complex and demonstrates a novel function for Cyclin C/Cdk8 in snRNA 3' end formation.

This thesis work has also characterized an important functional interaction mediated by a microdomain within Integrator subunit 12 (IntS12) and IntS1 that is required for the activity of the Integrator Complex in processing the snRNA 3' end. Through the development of a reporter-based functional RNAi-rescue assay in Drosophila S2 cells, we analyzed domains within IntS12 required for snRNA 3' end formation. This analysis unexpectedly revealed that an N-terminal 30 amino acid region and not the highly conserved central PHD finger domain, is required for snRNA 3' end cleavage. The IntS12 microdomain (1-45) functions autonomously, and is sufficient to interact and stabilize the putative scaffold protein IntS1.

Our findings provide more details of the Integrator Complex for understanding the molecular mechanism of snRNA 3' end processing. Moreover, these results lay the foundation for future studies of the complex through the identification of a novel functional domain within one subunit and the identification of additional subunits.

TABLE OF CONTENTS

Page No.

LIST OF FIGURES

Page no.

Appendix. IntS12 is Preferentially Recruited to Promoter Region of snRNA Genes.

LIST OF TABLES

Chapter 1. Introduction

INTRODUCTION AND BACKGROUND

Overview: Cellular functions and biogenesis of uridine-rich small nuclear RNA (U snRNA).

U snRNAs are a population of non-coding small RNAs in eukaryotic cells that are highly expressed, non-polyadenylated, and function in the nucleoplasm. Based on consensus sequence features and bound protein cofactors, U snRNAs are grouped into two different classes. The Sm-class contains a tri-methylguanosine cap at the 5' end, a stem-loop structure at the 3' end, and a central consensus protein binding site that serves as a binding platform for the heteroheptameric Sm ring (Sm site) (Figure 1.1A). The Lsm-class, in contrast, have a 5' mono-methylphosphate cap, a 3′ stem-loop, and a stretch of uridines at the 3' end that serves as a binding site for a distinct heptameric ring of Lsm proteins (Figure 1.1B).

The Sm-class snRNAs are transcribed by RNA polymerase II (RNAPII) and include U1, U2, U4, U5, U7, U11, U12, and U4atac. In contrast, the Lsm-class snRNAs are RNA polymerase III (RNAPIII) transcripts and include U6 and U6atac. U snRNAs are best known for their function in forming the RNA core of the spliceosome to facilitate the removal of introns from mRNA precursors (pre-mRNA). During the splicing reaction, specific and dynamic complementary interactions between the substrate and snRNA take place in the context of over a hundred of accessory proteins (Figure 1.2A). Accurate removal of intronic sequences is achieved by sequential synergistic actions of distinct small nuclear ribonucleoproteins (snRNPs). The early interactions in the splicing reaction include the U1 snRNP recognition of the 5'-splice sites of the pre-mRNA by base pairing, the U2 snRNP recognition of the branch point sequences, and identification of the 3' splice site by the U2 auxiliary factors. Once these recognition events have been established, the U4/U6/U5 tri-snRNP enters the complex to ultimately form the catalytic spliceosome (Will and Luhrmann 2011). The aforementioned snRNAs are the components of the cellular major spliceosome, whereas the U11, U12, U4atac and U6atac are found in the minor spliceosome and catalyze the removal of the non-canonical "atac" introns (Will and Luhrmann 2005). One exception within this snRNA group is the U7 snRNA, which is required for histone pre-mRNA 3′ processing and plays no role in pre-mRNA splicing. The U7 snRNP can basepair with the histone downstream element (HDE) that is located downstream of the cleavage site, and facilitates the 3' end endonucleolytic cleavage of the histone pre-mRNA (Figure 1.2B) (Marzluff et al. 2008).

Recent studies of the U1 snRNP reveal that it also functions beyond pre-mRNA splicing. The U1 snRNP is shown to activate gene transcription through a combination of interactions with the pre-mRNA transcript and with the transcription initiation machinery (Kwek et al. 2002). It is also found to repress the levels of viral transcripts by suppression of RNA

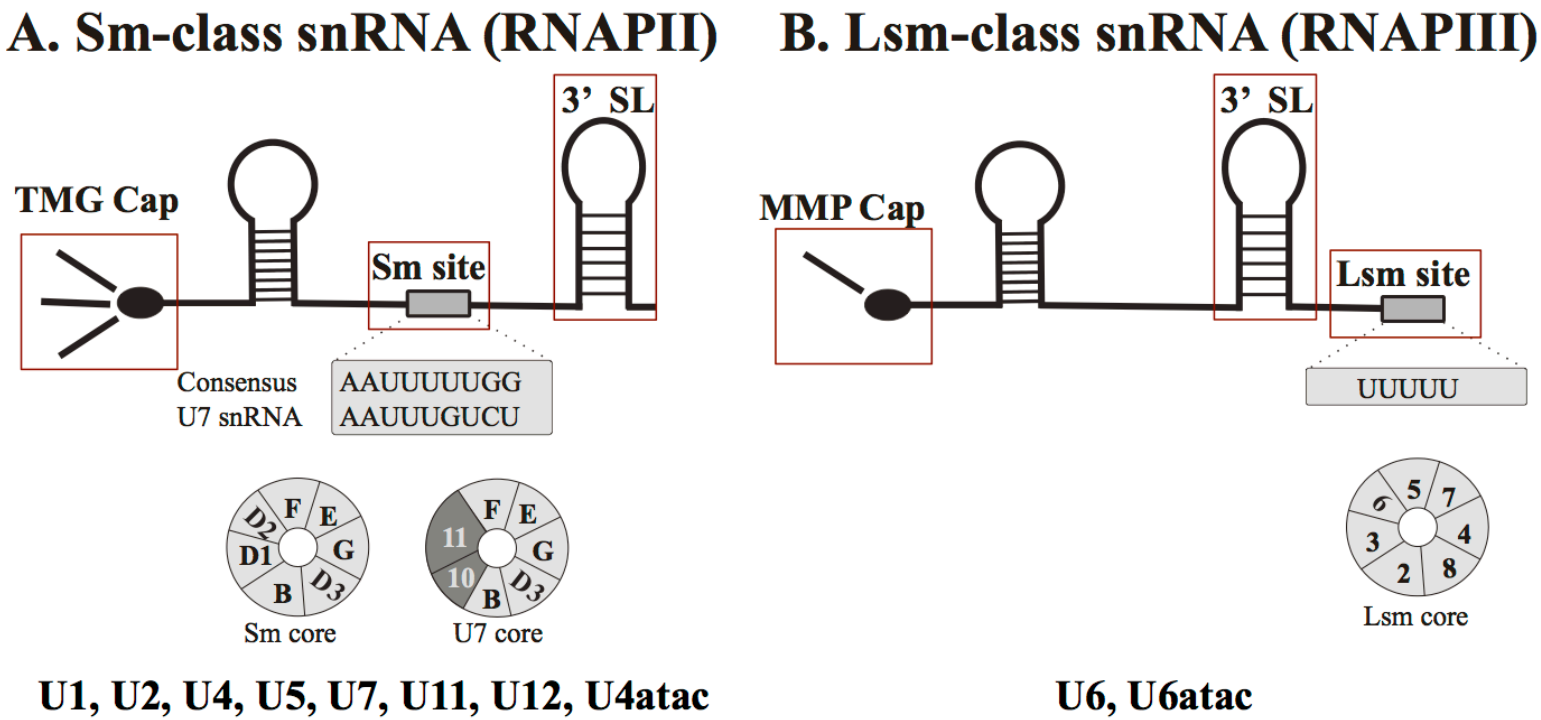


Figure 1.1 Anatomical features of Sm- and Lsm-class small nuclear RNAs.

Adapted from *Nature Reviews Molecular Cell Biology* Matera *et al. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs.* 8, 209–220 (March 2007).

Figure 1.1 Anatomical features of Sm- and Lsm-class small nuclear RNAs. (**A**). Sm-class small nuclear RNAs are transcribed by RNAPII, and have a 5’ trimethylguanosine (TMG) cap, a 3’ stem loop and a Sm protein binding site in the middle. The U7 snRNA has a noncanonical Sm site that is bound by a U7 Sm core. (**B**). Lsm-class snRNAs are transcribed by RNAPIII and have a 5’ monomethylphosphate (MMP) cap, a 3′ stem loop, and a stretch of uridines at the very 3’ end that serve as a binding site for a distinct heptameric ring of Lsm proteins.

A. snRNPs function in pre-mRNA splicing

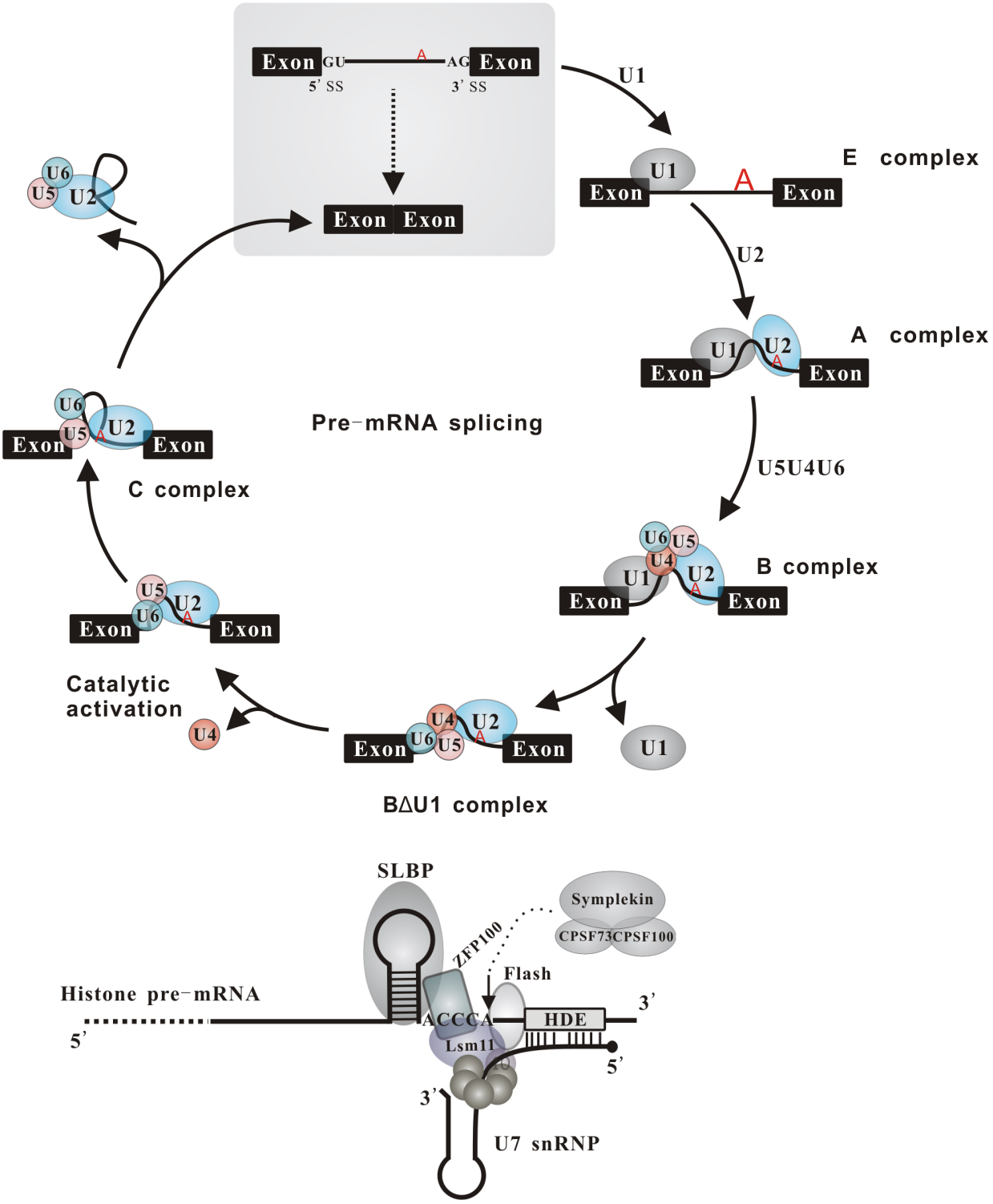


Figure 1.2 Cellular functions of spliceosomal snRNP and U7 snRNP.

Figure 1.2 Cellular functions of spliceosomal snRNP and U7 snRNP. (**A**). Schematic representation of the spliceosome assembly pathway during pre-mRNA splicing. Introns are excised from pre-mRNA by the spliceosome that is assembled by stepwise integration of U1, U2, and U4/U6.U5 snRNPs. (**B**). Schematic representation of U7 snRNP involved in histone mRNA 3' end formation. The U7 snRNP binds histone downstream element (HDE) through basepairing between 5' U7 snRNA and HDE sequence, and the stem loop binding protein (SLBP) binds the 3' stem-loop, together with ZFP100 and FLASH to stabilize the U7 snRNP and help recruiting the endonuclease to cleave the histone pre-mRNA.

polyadenylation (West 2012), and in human cells, it has been shown to protect the widespread premature cleavage and polyadenylation of pre-mRNA transcripts through U1 snRNA binding to the 5'-splice site-like sequences present in primary transcripts (Kaida et al. 2010). These studies are all consistent with the fact that the U1 snRNA is expressed at significantly higher levels than other snRNAs.

The Lifecycle of the snRNP.

Introduction. The biogenesis of snRNPs is an extensively studied research topic. The primary focus of these investigations has been to determine how snRNPs are assembled as well as how the mature snRNPs function in pre-mRNA splicing. In contrast, less is known about the events that precede snRNP incorporation into the spliceosome. The biogenesis of the functional Sm-class snRNPs includes a sophisticated nuclear-cytoplasmic life cycle (Figure 1.3A), while the Lsm-class snRNPs are only subject to a nuclear maturation process (Figure 1.3B).

snRNA Transcription. Transcription of the RNAPII transcribed snRNA genes is related in several ways to other transcription reactions but also possesses several important distinctions from mRNA-encoding genes (for more comprehensive reviews, see (Hernandez 2001; Egloff et al. 2008)). The RNAPII-transcribed snRNA genes have a TATA-less promoter that is featured by the presence of a well-conserved DNA element called the proximal sequence element (PSE). Transcription initiation at snRNA genes is mediated by a group of snRNA gene-specific proteins that have been referred to differently in the literature: the snRNA activating protein complex (SNAPc), the PSE-binding transcription factor (PTF), PSE-binding protein (PBP) (Hung and Stumph 2011). The SNAPc complex is thought to recruit RNAPII to the snRNA promoter by binding to the PSE element, similar to how general transcription factors recruit RNAPII, but the events that follow are snRNA-specific. For protein-encoding genes, transcription initiates from the recruitment of RNAPII by transcription factors, and is followed by phosphorylation of Ser5 within C-terminal domain of the heptad repeat ($YSPTS^5PS$) of the RNAPII large subunit (Rpb1) by the cyclin-dependent kinase 7 (Cdk7) of transcription factor IIH (TFIIH), which is critical for promoter clearance (reviewed in (Phatnani and Greenleaf 2006; Buratowski 2009)). Subsequent to promoter clearance there is an increase in the phosphorylation of Ser2 within the CTD heptad repeat by CDK9 of the positive transcription elongation factor-b complex (P-TEFb) that is concomitant with a reduction in the levels of pSer5. This change in phosphorylation status is thought to increase elongation efficiency and facilitate the loading of RNA-processing factors onto the RNAPII

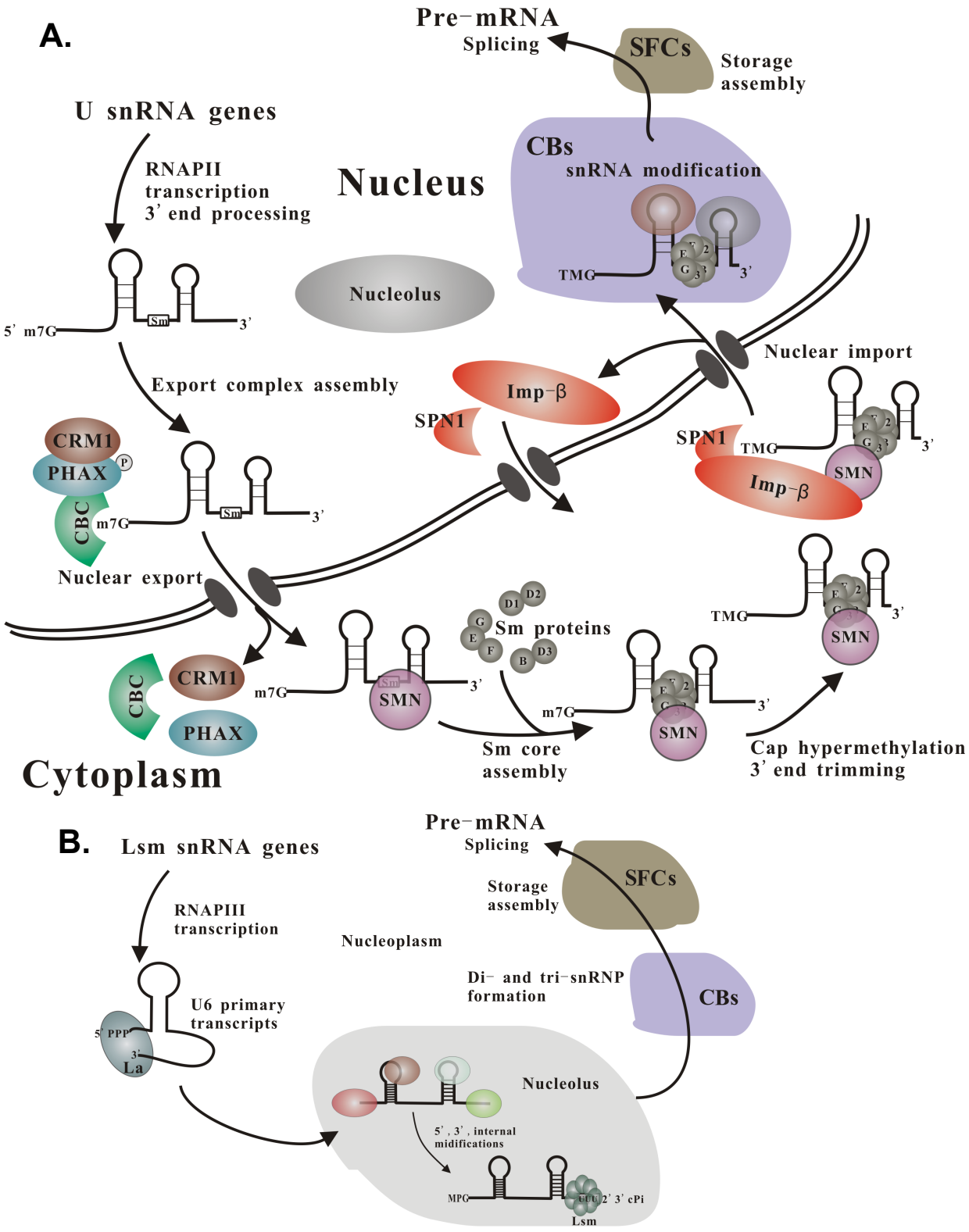


Figure 1.3 Sm-class and Lsm-class snRNP biogenesis pathways.

Adapted from *Nucleic Acids Research*, Patel S and Bellini M, *The assembly of a spliceosomal small nuclear ribonucleoprotein particle.* 2008, Vol. 36, No. 20, 6482–6493.

Figure 1.3 Sm-class and Lsm-class snRNP biogenesis pathways. (**A**). The biogenesis of Sm-class snRNPs occurs both in the nucleus and the cytoplasm. The snRNA primary transcripts are generated by RNAPII, cleaved by the Integrator complex and then exported to the cytoplasm by the PHAX-containing snRNA export complex. In the cytoplasm, the SMN complex recruits the heteroheptameric Sm proteins ring and tethers it to snRNA to form the Sm-core RNP. Following assembly of the Sm core, the 7-methylguanosine (m^7G) cap is hypermethylated and the 3' end is trimmed. The newly formed TMG cap serves as a signal for assembling the import complex containing the import adaptor Snurportin-1 (SPN) and the import receptor Importin-β (Imp-β). Once returned to the nucleus, the Sm-class snRNPs are targeted to Cajal bodies for snRNA-based modification, which is requisite for spliceosomal function. The matured snRNPs then leave the Cajal bodies and either are stored in the splicing factor compartments (SFCs) or participate in transcription-coupled pre-mRNA splicing at perichromatin fibrils (PFs). (**B**). The biogenesis of Lsm-class snRNP (U6 snRNP) is an exclusively nuclear process. The U6 snRNA are transcribed by RNAPIII and the 3’ end is formed by transcription termination with stretch of uridines. The newly transcribed U6 snRNA is bound by the La antigen at both 5’ and 3’ ends and then is targeted to nucleoli, where its 5’ cap and 3’ end are modified. The modification discharges La protein binding and enables Lsm protein binding. Once the Lsm core is assembled onto U6 snRNA, the U6 snRNP is targeted to Cajal bodies for further maturation. The mature U6 snRNP either are stored in the splicing factor compartments (SFCs) or participate in transcription-coupled pre-mRNA splicing process.

CTD in order to catalyze downstream RNA processing reactions (Zhang et al. 2005; Zhang and Gilmour 2006). Unlike its role in protein-encoding genes, P-TEFb is found to be dispensable for efficient elongation of snRNA genes, which is not unexpected given the relative short length and lack of polyadenylate tails in snRNAs (Medlin et al. 2005). Surprisingly, however, phosphorylation of Ser2 in the heptad repeat of the RNAPII CTD has been reported to be required for proper snRNA 3′ end formation, thereby implicating a distinct requirement for P-TEFb in snRNA 3' end formation (Medlin et al. 2005). Recently, an additional phosphorylation at Ser7 of the RNAPII CTD has been reported by the Murphy laboratory to be essential for snRNA processing (Chapman et al. 2007; Egloff et al. 2007), and, surprisingly, TFIIH has been found to phosphorylate Ser7 of the CTD, expanding its established substrate specificity beyond Ser5 (Akhtar et al. 2009).

snRNA nuclear export. The primary snRNA transcript is cleaved at the 3' end to facilitate RNAPII termination and recycling. This cleavage event is governed by a poorly conserved cis-regulatory element called the 3′-box, which is located several nucleotides downstream the cleavage site (Hernandez 1985). The 3′ box is likely to serve as a recognition site for the Integrator Complex (described in more detail below) that is responsible for the nascent snRNA cleavage from the elongating polymerase. Once the snRNA is cleaved from the DNA template, it is exported to the cytoplasm by an snRNA-specific export complex for maturation.

The cytoplasmic export of snRNA is carried out by a protein complex that is comprised of the snRNA-specific phosphorylated adaptor for RNA export (PHAX), the cap-binding complex (CBC), the chromosome region maintenance-1 (CRM1/Xpo1), and RanGTP (Ohno et al. 2000). The 5′ mono-methylguanosine cap structure and length of the snRNA are key determinants in snRNA nuclear export while the compartmentalized phosphorylation status of PHAX controls the directionality of export (Ohno et al. 2000). Phosphorylated PHAX is localized in the nucleus and together with CBC complex bridges the snRNA to CRM1/Xpo1 for export. Once PHAX enters into the cytoplasm it is dephosphorylated causing export complex disassembly and PHAX is then recycled to the nucleus through the import receptor importin-β (Segref et al. 2001). The nuclear kinase and cytoplasmic phosphatase that regulate this process have been identified to be Casein Kinase 2 (CK2) and protein phosphatase 2A, respectively (Kitao et al. 2008). There have been no PHAX orthologues found in fungal genomes suggesting that either the fungal snRNAs are not exported to cytoplasm or they use a different system for snRNA export.

SMN and snRNP assembly. Once exported to the cytoplasm, the snRNAs are loaded onto the Sm-core particles under the facilitation of the Survival of Motor Neuron (SMN) protein

complex. Proper function of SMN is critical, as loss or mutation of *SMN1* gene is known to cause selective dysfunction of motor neurons leading to Spinal Muscular Atrophy (SMA) (Burghes and Beattie 2009; Coady and Lorson 2011; Workman et al. 2012). Given the importance of SMN in snRNP biogenesis, one explanation is that mutation of SMN results in splicing alterations of specific genes that are required for neuromuscular junction function. The SMN complex (SMN, Germin2-8 and UNR-interacting protein) bridges the heptameric Sm protein ring and the newly exported snRNA precursors to form the snRNPs. The binding specificity of SMN complex to snRNAs is mediated through the interaction between the Germin5 WD repeat and Sm site of snRNA (Lau et al. 2009; Yong et al. 2010). It is noteworthy that the U7 snRNA has a nonconsensus Sm site, which determines the inclusion of two alternative Sm-like proteins (Lsm10 and Lsm11) in the U7 core (Figure 1.1A). It is still not known how the SMN complex facilitates loading the alternative heptameric Sm core onto the U7 snRNA.

Once the Sm core/SMN complex are assembled on the Sm sites, the cap structure of snRNA is hypermethylated by the tri-methylguanosine synthase-1 (TGS1) to form a 2,2,7-trimethylguanosine (TMG) cap (Mouaikel et al. 2002). The snRNA 3' end is then further trimmed several nucleotides by a yet-to-be identified cytoplasmic exonuclease. The newly formed TMG cap serves as a nuclear import signal that is recognized by the adaptor protein Snurportin-1, and together with the SMN complex, the snRNP is re-imported into the nucleus through the import receptor importin-β (Imp-β), which directly interacts both Snurportin-1 and SMN complex (Segref et al. 2001).

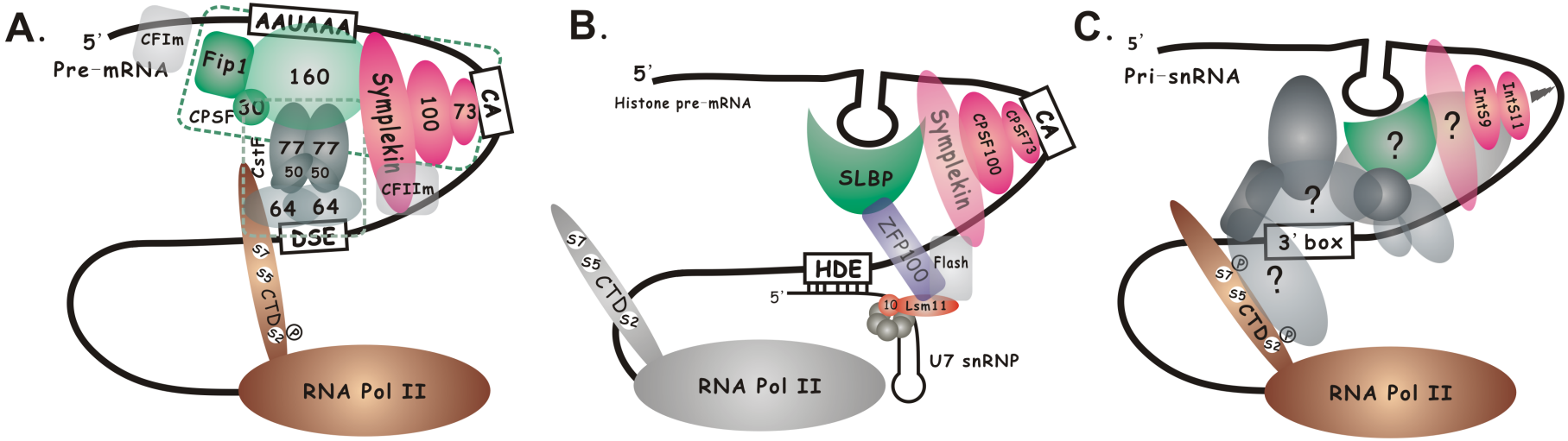
Once imported back into the nucleus, the Sm-class snRNPs are sorted into Cajal bodies for further maturation through site-specific modification and assembly of snRNP-specific proteins (Matera et al. 2007; Patel and Bellini 2008). A recent study also indicates that the production of snRNAs is a critical event for maintaining the integrity of the nuclear Cajal bodies in human cells (Takata et al. 2012). This suggests that there is an interplay between the formation of the Cajal Body and the levels of snRNAs that transit through the Cajal Body. Functional snRNPs then leave the Cajal bodies and function in splicing at perichromatin fibrils (PFs) or are stored in splicing factor compartments (SFCs) for later use.

The RNAPIII-transcribed U snRNAs (*U6* and *U6atac*) use a unique extragenic RNAPIII type III promoter for transcription (Schramm and Hernandez 2002) and their 3' ends terminate with a run of uridines that serves as the Lsm-binding site (Figure 1.1B) as well as the Pol III transcription terminator. These snRNAs remain in the nucleus for maturation (Figure 1.3B) (detailed review, see (Patel and Bellini 2008)).

3' end formation of RNAPII transcripts

Introduction. In eukaryotes, RNA polymerase II transcribes a variety of RNA transcripts that serve distinct functions within the cell. The three major types of RNAPII-transcribed RNA are: the polyadenylated mRNAs, the nonpolyadenylated histone mRNAs, and the small nuclear RNAs. The presence of a poly(A) tail at the 3' end is a prominent feature of the eukaryotic mRNAs, which is important for mRNA export, stability and translation. However, the metazoan replication-dependent histone mRNAs and the ubiquitously expressed snRNAs have a stem-loop structure at their 3' end. These non-polyadenylated RNAs utilize mechanisms to form their 3' ends that are distinct from poly(A)+ mRNA but they also share certain similarities for the 3' end formation that include the arrangement of cis-regulatory sequence elements and the chemistry of the cleavage event itself (Reviewed in (Chen and Wagner 2010)).

***The polyadenylated messenger RNA*.** The 3' end formation of polyadenylated mRNAs is carried out by the cleavage/polyadenylation machinery and initially requires recognition of the two cis-regulatory 3' end processing signals and then cleavage of the pre-mRNA (Figure 1.4). The initial binding of the highly conserved AAUAAA polyadenylation signal (PAS) is carried out by the 160 kDa subunit of the five membered Cleavage and Polyadenylation Specificity Factor (CPSF) complex (Murthy and Manley 1995). This occurs in parallel or in concert with the recognition of the downstream G/U-rich sequence element (DSE) through the 64 kDa subunit of the trimeric Cleavage Stimulation Factor (CstF) complex (MacDonald et al. 1994). Recognition of these two conserved sequence elements by the CPSF and CstF complexes facilitates the recruitment of a cleavage factor that then cleaves the pre-mRNA at the cleavage site, which is preferably a CA dinucleotide located between the PAS and DSE. The cleavage factor for polyadenylated mRNA comprises of the endonuclease CPSF73, the catalytically inactivated CPSF100, and a large scaffold protein called Symplekin (Mandel et al. 2006). The crystal structure of CPSF73 demonstrates that it belongs to the metallo-β-lactamase (MBL) and metallo-β-lactamase-associated CPSF, Artemis, SNM1/PSO2 (β-CASP) subfamily chelating zinc ions, and cleavage assay further revealed that it is a hydrolase capable of cleaving single stranded RNA substrates *in vitro* (Mandel et al. 2006). Interestingly, the crystal structure of CPSF100 clearly demonstrates that it adopts the same fold as CPSF73, but it does not coordinate zinc ions in its active site and cleavage analysis shows that it has no endonuclease activity to RNA tested *in vitr*o, consistent with the observation that it lacks critical amino acids thought to be required for catalytic activity. Symplekin is an ARM/HEAT repeats containing protein and data from its crystal structure indicates that its HEAT domain is likely to function as a scaffold for protein-protein interactions essential to the mRNA maturation process in mammalian cells (Andrade et al. 2001). Besides the factors mentioned above,



Poly(A)+ mRNA 3' end processing machinery

Poly(A)- histone mRNA 3' end processing machinery

snRNA 3' end processing machinery

Figure 1.4 Comparison of the three 3' end processing complexes for RNA polymerase II transcripts. The molecular machineries governing 3' end processing for polyadenylated mRNA (**A**), nonpolyadenylated histone mRNA (**B**) and small nuclear RNA (**C**) are presented. The three 'cleavage factor' proteins are denoted in pink for all three systems with the Integrator cleavage factor having one of its unknown subunits labeled with a question mark. The dark-grey proteins represent factors involved in either CTD recognition or downstream *cis* element (DSE, HDE or 3' box) binding, and the green proteins represent factors for upstream 3' end sequence element (DES, stem-loop) recognition. CstF, cleavage stimulation factor; HDE, histone downstream element; RNA Pol II, RNA polymerase II.

cleavage factor I (CFIm) and cleavage factor II (CFIIm) complexes have also been implicated in the cleavage reaction of pre-mRNAs lacking the canonical PAS signal by stimulating binding of the CPSF complex to the pre-mRNA through a functional interaction with Fip1 (Brown and Gilmartin 2003; Venkataraman et al. 2005). Finally, the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (Rpb1) is required for the pre-mRNA cleavage step likely through its ability to recruit cleavage factors to sites of transcription. Tightly coupled to the cleavage step is the addition of a poly(A) tail to the newly formed mRNA 3' end by the poly(A) polymerase (PAP) to form the mature mRNA with a 200-250 nucleotide poly(A) tail.

The replication-dependent histone mRNA. The mature replication-dependent metazoan histone mRNAs terminate with a 3' end stem-loop structure that is generated by single endonucleolytic cleavage but it is distinctly not followed by polyadenylation. Histone genes are intronless, thus this single cleavage reaction is the only RNA processing event required to form mature histone mRNAs. The 3' end of histone mRNA is formed by a distinct processing machinery that recognizes two conserved sequence elements present in histone pre-mRNAs: the 3' end stem-loop and a purine-rich histone downstream element (HDE) (Dominski and Marzluff 2007) (Figure 1.4). The stem-loop sequence is highly conserved throughout evolution and consists of a 6-base pair stem and a 4-nucleotide loop. In contrast, the HDE sequence of vertebrate histone pre-mRNAs is less conserved but all contains a purine-rich core, which has a typical sequence of PuAAAGAGCTG (Pu, purine) located 15-20 nucleotides downstream of the 3' end stem-loop. The conserved stem-loop structure in histone pre-mRNA is recognized by the stem-loop binding protein (SLBP) through its central RNA binding domain (RBD) (Wang et al. 1996), and the downstream HDE sequence is a binding site for the U7 snRNP where the 5' end of U7 snRNA basepairs with HDE sequence (Mowry and Steitz 1987; Dominski et al. 2003). Recognition of these two elements by SLBP and U7 snRNP recruits other protein factors, including the cleavage factor CPSF73/CPSF100 and Symplekin, to cleave the histone pre-mRNA at the cleavage site, typically an adenosine between the stem-loop and the HDE sequences. Thus the unique histone mRNA 3' end processing machinery includes at least the U7 snRNP, SLBP, a 100-kDa zinc-finger protein (ZFP100) and a same large complex consisting of CPSF and CstF complexes as the effector of RNA cleavage.

The U7 snRNA has an unstructured 5' terminus that is involved in formation of a duplex with the histone HDE sequence, and thus is essential for histone pre-mRNA 3' end formation. In contrast, the 3' end of U7 snRNA folds into a stem-loop structure, which has been shown to be resistant to alterations at the sequence level suggesting that the structure itself is important rather than the content. The central region of the U7 snRNA contains its noncanonical Sm protein-binding site that guides inclusion of a different Sm core called the U7 core. Instead of

the classical Sm core consisting of seven Sm proteins: B, D1, D2, D3, E, F and G, the U7 core forms a heptameric ring by replacing the SmD1, D2 proteins with Lsm10 and Lsm11, respectively (Pillai et al. 2001; Pillai et al. 2003). Lsm10 resembles other Sm proteins in size and structure and shares the highest sequence homology to SmD1 but is not observed to be included in spliceosomal snRNPs and is required for U7 snRNP function. Interestingly, Lsm11 has an unusual large size with a long N-terminus, and it has been shown to directly participate in recruiting an essential processing factor for histone pre-mRNA 3' end formation (Yang et al. 2012). The U7 snRNP that contains either N-terminal truncated Lsm11 or N-terminal mutated Lsm11 is defective in histone pre-mRNA 3' end processing (Pillai et al. 2003). Recognition of the HDE sequence by the U7 snRNP is an essential event for histone pre-mRNA 3' end processing and has been proposed to serve as a molecular ruler that specifies the site of cleavage. Mutations introduced into the HDE sequence that compromise U7 snRNA complimentary base paring consistently lead to complete inhibition of the 3' end cleavage reaction and alteration of the distance between the stem-loop and the HDE sequence shifts the location of the cleavage site (Mowry et al. 1989; Scharl and Steitz 1994).

Mutation of the stem-loop (SL) sequence or genetic loss of SLBP leads to misprocessing of histone pre-mRNA underscoring the essential nature of this element and its RNA binding protein (Wang et al. 1996). However, mutations introduced into the stem-loop sequence in the context of an HDE that has very high complementarity with the U7 snRNA does not significantly affect the efficiency of 3' end processing tested *in vitro* (Dominski et al. 1999). This suggests that the recognition of the stem-loop by SLBP is used to stabilize binding of U7 snRNP to the HDE, especially in histone pre-mRNAs with relatively weak complementarity of the HDE sequence to U7 snRNA 5' end. This notion is further supported by the evidence that ZFP100 interacts with the SL/SLBP complex and Lsm11 thus bridging both factors (Dominski et al. 2002; Azzouz et al. 2005). Additional interactions have been identified with the recent finding that binding between a 220 kDa proapoptotic protein FLASH and Lsm11 in the U7 snRNP is essential for histone pre-mRNA 3' end processing (Yang et al. 2009). The N-terminal regions of these two proteins interact and form a platform for a unique combination of polyadenylation factors including all CPSF subunits, Symplekin and CstF subunits, providing a molecular mechanism for recruitment the similar cleavage factor containing the endonuclease CPSF73 to histone pre-mRNAs (Burch et al. 2011; Yang et al. 2012). Finally, in stark contrast to its role in protein-encoding genes, the CTD of Rpb1 is likely not required for coupling transcription and 3' end processing of histone pre-mRNA. Evidence shows that inhibition of the CTD Ser2 kinase Cdk9 affects neither transcription of histone genes nor 3' end processing of histone pre-mRNAs (Medlin et al. 2005).

The uridine-rich small nuclear RNAs. U snRNA primary transcripts made by RNAPII are very similar to their histone pre-mRNA counterparts in that both transcripts are short, intronless, and contain a stem-loop structure at their 3' end. For snRNA, however, the size and content of this 3' stem-loop is not well conserved, and the 3' end processing signal downstream of the cleavage site is an AU-rich cis-regulatory sequence element termed "3' box" (Hernandez 1985). The 3' box is located 9-19 nucleotides downstream of the 3' end of mature snRNA and has a typical sequence of $GTTTN_{0-3}AAAPuN_2AGA$ (N, any nucleotide; Pu, purine). Unlike the rigid requirement for the HDE element in 3' end processing of the histone pre-mRNAs, the snRNA 3' box is required but more tolerant to mutation in that no single point mutation significantly compromises efficient snRNA 3' end processing (Ach and Weiner 1987; Ezzeddine et al. 2011). This suggests that other mechanisms are likely to contribute to the cleavage specificity. Indeed, mutagenesis analysis of *Drosophila U7* snRNA reveals that its 3' end stem-loop structure is also required for efficient 3' end formation, which is analogous to the role of the 3' end stem-loop in histone pre-mRNA 3' end formation (Ezzeddine et al. 2011). Accurate 3' end formation of pre-snRNA in metazoan cells also requires that transcription initiates from the unique TATA-less snRNA promoters. Replacing the snRNA promoter with other RNAPII promoters completely abolishes proper pre-snRNA 3' end formation, indicating that unique factors important for accurate 3' end processing must load onto the snRNA promoter early in the transcription cycle (de Vegvar et al. 1986; Hernandez and Weiner 1986; Ezzeddine et al. 2011).

In addition to the promoter-coupled 3' end formation is the fact that the CTD of Rpb1 is also required for snRNA gene expression. Inhibition of Cdk9 activity by kinase inhibitors does not affect snRNA gene transcription but abolishes their RNA 3' end processing suggesting that phosphorylation of serine 2 of the CTD is selectively required for snRNA 3' end processing but is dispensable for transcription elongation (Medlin et al. 2003; Jacobs et al. 2004). Indeed, follow-up studies in mammalian cells demonstrate that a double phosphorylation mark on the RNAPII CTD, pSer2 and pSer7 is required for recruitment of the snRNA specific 3' end processing machinery-the Integrator complex to snRNA gene promoters (Egloff et al. 2007; Egloff et al. 2010).

The snRNA 3' -end processing complex: Integrator Complex

The Integrator complex was serendipitously discovered by Baillat et al from the Shiekhattar laboratory to associate with the CTD of RNAPII and serves as the 3' end processing machinery for metazoan RNAPII transcribed U snRNAs (Baillat et al. 2005). The name "Integrator" refers to integrating the CTD of RNAPII largest subunit with the 3' end processing of small nuclear RNAs U1 and U2. The original biochemical purification and mass

spectrometric analysis identified twelve different polypeptides that exhibited very little homology with known proteins. A domain schematic of all 12 known human Integrator proteins analyzed by Pfam is shown in Figure 1.5 and well-defined domains are shown in black and regions with high sequence conservation are marked in grey. Proteins in the complex are annotated in numerical order based on migration size on the SDS-PAGE gel, with Integrator 1 (IntS1) migrating the slowest with a molecular weight of ~ 244 kDa and IntS12 migrating as the fastest at ~48 kDa. The majority of proteins in the Integrator complex were not studied at the time of purification, and none of the proteins have an identifiable RNA-binding domain. However, a close examination reveals the presence of a von Willebrand factor type A (VWA) domain in IntS6, Armadillo repeats (ARM) in IntS4 and IntS7, and a plant homeodomain (PHD) finger in IntS12. IntS11/RC-68 and IntS9/RC-74 both contain the MBL/β-CASP domains and display high sequence similarity to the cleavage and polyadenylation specificity factor CPSF73 and CPSF100, which are known to function in the cleavage reaction of other RNAPII transcripts (Dominski et al. 2005a; Dominski et al. 2005b). Interestingly, IntS11 is predicted to be the catalytically active endonuclease analogous to CPSF73 that enables to cleave single strand RNA, and IntS9 analogous to CPSF100 has critical residues required for the activity of the MBL/β-CASP domain mutated, making it catalytically inactive (Mandel et al. 2006). As predicted, IntS9 and IntS11 have been shown to form heterodimer through their C-terminal regions using yeast two-hybrid analysis and pulldown assays and this interaction is important for snRNA 3' end formation in vivo (Albrecht and Wagner 2012). Based on these results, IntS9/11 are proposed to form the core that performs the catalytic cleavage reaction of the pre-snRNA. An *in vitro* RNA cleavage assay or a co-crystallization of the snRNA substrate and the endonuclease will provide final evidence to support this model. The details of the Integrator Complex molecular mechanism remain to be determined. More specifically, it is not known which Integrator subunit is responsible for recognition of the 3' box or the 3' stem-loop structure and which subunit facilitates complex recruitment to the snRNA promoter to couple transcription with 3' end formation.

Several studies have emerged to provide cellular and biological details of the Integrator complex in different organisms. Some of these studies are consistent with a function of snRNA biogenesis, whereas others suggest that Integrator proteins may function beyond that. Targeted disruption of mouse IntS1/KIAA1440 causes embryonic lethality at the blastocyst stage, demonstrating a fundamental role of IntS1 in mouse development that is not compensated by other genes (Hata and Nakayama 2007). The observation of accumulation of unprocessed, primary U2 snRNA transcript but decrease of the mature U2 snRNA transcript in IntS1 null embryos implies that the developmental role of IntS1 in mouse is likely through controlling snRNA biogenesis pathway.

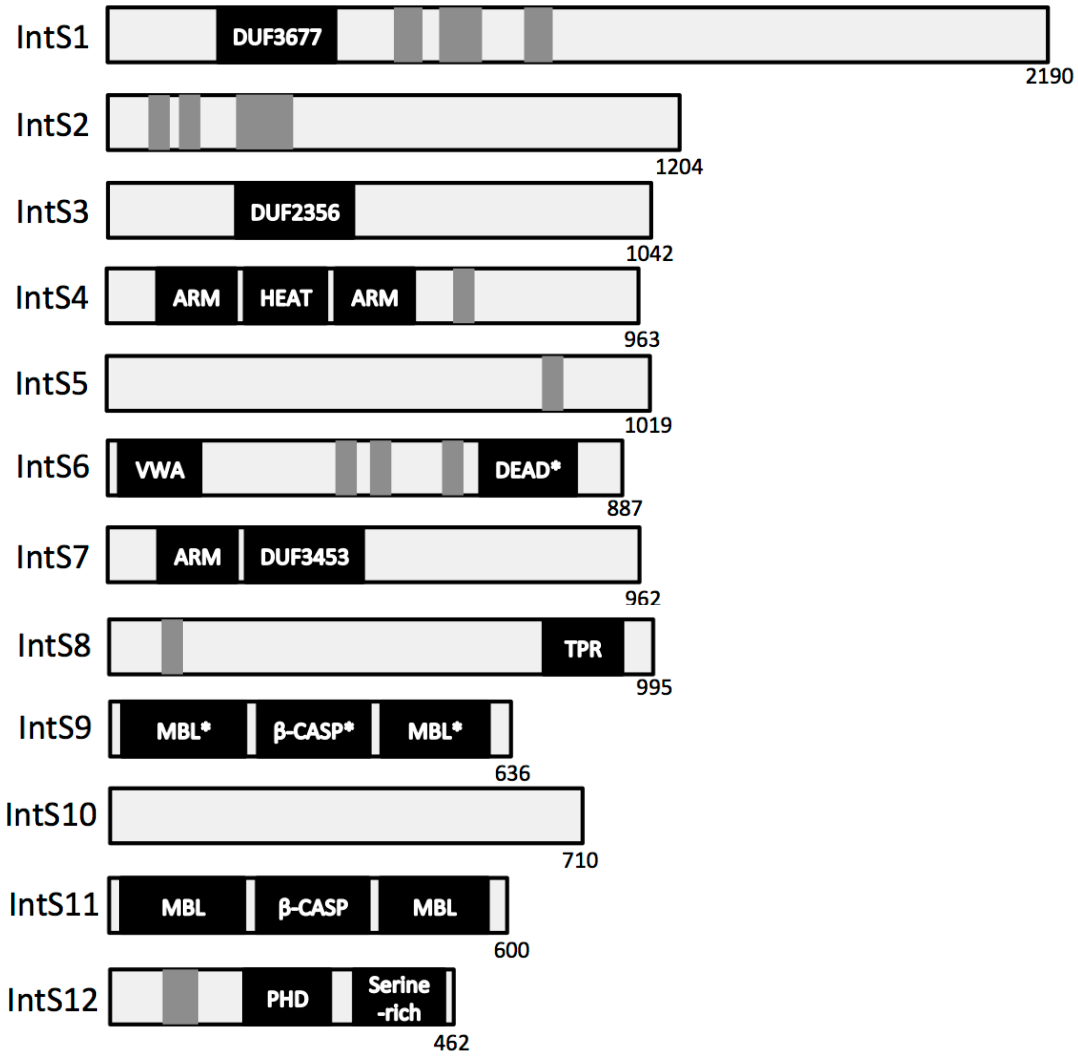


Figure 1.5 Schematic representations of domains and motifs found in Integrator protein. Protein domain and conserved motif search is conducted by using Pfam analysis, and protein domains are labeled in black whereas conserved motif (>80%) over a stretch of ten amino acids are labeled in grey. The species used for conservation analysis are human, cow, chicken, *Drosophila* and zebrafish. Abbreviations: ARM, Armadillo-fold repeats; MBL, metallo-β-lactamase domain; β-CASP, metallo-β-lactamase-associated CPSF Artemis SNM1/PSO2; DEAD, RNA helicase DEAD box; DUF, domain of unknown function; HEAT, HEAT repeat units; PHD, plant homeodomain finger; TPR repeats, tetratricopeptide repeats; VWA, von Willebrand factor type A domain; IntS, Integrator subunit; * represents that certain critical residues are altered that renders the domain inactive.

Similarly, IntS4 or IntS7/*deflated* null flies exhibit various developmental defects at early stages and in both cases die at the stage of late second instar larvae. Elevated levels of unprocessed primary U snRNAs are also observed in these Integrator null flies, suggesting that an essential developmental role of the snRNA biogenesis controlled by the Integrator complex (Rutkowski and Warren 2009; Ezzeddine et al. 2011). Moreover, antisense morpholino-mediated depletion of IntS5 (or IntS9 and IntS11) in zebrafish embryos causes specific red blood cell differentiation arrest. This is reported to be a requirement of the Integrator complex for the bone morphogenetic protein (BMP) signaling pathway, a pathway essential for the specification and proliferation of blood progenitors. Depletion of the Integrator subunits affects snRNA processing, which in turn leads to aberrant splicing of *smad1* and *smad5* pre-mRNAs, two downstream effectors of BMP signaling important for hematopoiesis (Tao et al. 2009). The observed developmental defect in zebrafish is surprisingly specific and one possible explanation is that splicing of these particular pre-mRNAs at specific developmental stage is sensitive to the levels of functional snRNA in cells.

Recent studies from several different laboratories have identified IntS3 as DNA damage response gene as IntS3 localizes to DNA damage foci once human cells are subject to genotoxic stress (Huang et al. 2009; Li et al. 2009; Skaar et al. 2009). It is found to be one component of the heterotrimeric sensor of single-stranded DNA (SOSS) complex that is likely involved in maintenance of genome stability. IntS3 may function as a platform to form the complex and its N-terminal region including the highly conserved DUF domain shown in Figure 1.5 is required for both interactions with other two components in the complex (Huang et al. 2009). The SOSS complex appears to be distinct from the known Integrator complex, as SOSS subunits other than IntS3 were neither found in the original Integrator purification nor in the recent high-throughout proteomic analysis of Integrator immunoprecipitations (Malovannaya et al. 2010; Malovannaya et al. 2011). Recently, a DNA damage response screen also identified IntS7 as one candidate protein that is recruited to the sites of DNA damage (Cotta-Ramusino et al. 2011). But it is still unknown how IntS7 participates in the DNA damage response. IntS6/DICE1 (deleted in cancer 1) is the only member in the Integrator complex that had been functionally studied before the purification of the Integrator complex (Wieland et al. 1999; Han et al. 2006; Filleur et al. 2009). *IntS6* is a putative tumor suppressor gene, as it is often found to be deleted or downregulated in the majority of non-small cell lung carcinomas and prostate cancers and overexpression of IntS6 in prostate cancer cells or human non-small cell lung carcinoma cells inhibits colony formation and causes a cell-cycle arrest (Wieland et al. 1999; Filleur et al. 2009). No mutations in IntS6 were reported in human non-small cell lung carcinomas and prostate cancers, and the reduced levels of expression in prostate cancers is due to CpG hypermethylation of the IntS6 promoter (Ropke et al. 2005). It

is unexpected that loss or downregulation of a specific Integrator protein, which would presumably affect snRNA biosynthesis, correlates to tumorigenesis. The possible explanation is that IntS6 may form different complexes that function in different pathways. IntS6 contains an N-terminal well-conserved VWA (von Willebrand factor type A) domain, which is present in both extracellular and intracellular proteins. The extracellular VWA domains are found to mediate adhesion via metal ion-dependent adhesion sites (MIDAS), and the intracellular VWA proteins are commonly found in multiprotein complexes, which are involved in a variety of functions such as transcription, DNA repair and the proteasome (Whittaker and Hynes 2002). Absence of a signal motif and association with the RNAPII CTD suggests that IntS6 plays an intracellular function. Work on the worm orthologue of IntS6 (DIC-1) however, suggests that DIC-1 is a mitochondrial protein located in the inner membrane of mitochondria, and is required for formation of the normal morphology of mitochondria. DIC-1 is required for worm development as RNAi-mediated depletion of DIC-1 leads to defective oogenesis and inviable embryos (Han et al. 2006). This observation is unexpected, as Integrator proteins have been shown to function in the nucleus in snRNA 3' end formation. Finally, IntS4 (or IntS11) has been found to be required for the Integrity of the nuclear Cajal bodies as RNAi-mediated depletion of IntS4 significantly abolishes Cajal body formation in HeLa cells (Takata et al. 2012). Given the importance of the Integrator complex in snRNA 3' end formation, it is very likely that Integrator complex contributes to nuclear Cajal body formation through controlling snRNA maturation.

Besides the known Integrator subunits, a recent high-throughput proteomic analysis of immunoprecipitates from human cells using specific antibodies against different Integrator subunits confirms the presence of all 12 Integrator proteins in the complex and demonstrates the existence of additional protein factors that form an expanded complex (Malovannaya et al. 2010; Malovannaya et al. 2011). However, no functional studies have been done to determine the importance of these factors, especially their roles in snRNA 3' end formation, making their importance to snRNA biogenesis unknown.

SIGNIFICANCE OF THE STUDY

RNA processing is a major engine driving diversity within the proteome that is absent in the comparatively simpler genome. Not surprisingly, mutations in many of the genes involved in RNA processing can lead to human diseases manifesting in almost any tissue type. One of the most dramatic examples of such a disease is SMA, which is caused by attenuation in the activity of the SMN protein that is responsible for snRNP assembly (Coady and Lorson 2011; Workman et al. 2012). SMA and a host of other splicing-related diseases underscore the need for further understanding of snRNA biogenesis.

The combined efforts of several laboratories have created the paradigm that the 3' end formation of metazoan RNAPII-transcribed snRNA requires the snRNA promoter itself as well as the 3' box and the potential 3' stem-loop structure (Hernandez 1985; Ezzeddine et al. 2011). It is clear that the 3' box is not a termination site as RNAPII has been found to transcribe much further downstream (Cuello et al. 1999). Rather, it is a cis-regulatory element that likely binds a complex of proteins that carry out the 3' end cleavage reaction but the components of this complex had remained unknown until 2005.

The identification of the Integrator complex as the snRNA 3' end processing machinery represents a significant advancement in our knowledge of how snRNA 3' end formation occurs (Baillat et al. 2005). The Integrator subunits were purified due to their association with the CTD of the large subunit RNAPII. The initial purification identified twelve polypeptides that bear little resemblance to the 3' processing factors known to participate in both poly(A) mRNA and histone mRNA 3' end formation with the exception of IntS9/11. The predicted snRNA 3' end processing model (Figure 1.4), which is analogous to the counterpart for poly(A) mRNA or histone mRNA 3' end processing, involves recognition of the 3' box and 3' stem-loop elements by Integrator proteins, which subsequently recruit the endonuclease IntS11 to perform the cleavage reaction.

Since no identifiable RNA-binding domain is present in the original purified Integrator complex and a more expanded biochemical Integrator complex has been identified, we set out to identify the candidate protein factors that are functionally required for snRNA 3' end formation by a genome-wide functional RNAi screen in *Drosophila* S2 cells. Chapter 3 describes the genome-wide screen data and it shows that a majority of the known Integrator subunits (10/12) are required for reporter U7 snRNA 3' end formation. Importantly, this screen identified Asunder and CG4785 as two additional core Integrator subunits that are functional required for snRNA 3' end formation and biochemically interact with the known Integrators. The screen also identified a conserved requirement for Cyclin C/Cdk8 in snRNA 3' end formation. This work is important for understanding snRNA 3' end formation because it provides genome-wide analysis from a functional perspective, and thus complements the

existing biochemical data. This work also lays the foundation to further understand the mechanism of the snRNA 3' end machinery through the identification of additional critical factors required for cis-regulatory sequence element recognition and cleavage complex assembly.

The work described in Chapter 4 characterizes the role of one of the Integrator subunits, IntS12, in snRNA 3' end formation. IntS12 is the smallest subunit in the Integrator Complex and contains a centrally located and well-conserved plant homeodomain (PHD) finger. Data in Chapter 4 demonstrate the surprising finding that there is an N-terminal "microdomain" within IntS12 that is capable of mediating its function in snRNA 3' end formation. The IntS12 microdomain mediates its function through an interaction with IntS1, while the centrally located conserved PHD finger is functionally dispensable. These results suggest that IntS12 contributes to snRNA 3' end formation through a requisite binding to IntS1, which is the largest Integrator subunit. Chapter 5 further characterizes the biochemical features of IntS12, focusing specifically on its PHD finger and its affinity toward histones. Data in this chapter show that though functionally not required for processing of the reporter snRNA 3' end, the IntS12 PHD finger avidly binds histone H3 *in vitro*, and enhances interaction with other Integrators in cells. The cellular function of these biochemical interactions is not known yet, but we speculate that the PHD finger is playing a role beyond snRNA processing.

At the beginning of my doctoral research, very little was known about the human Integrator Complex and essential nothing was known about the fly complex. Taken together, data presented in these studies functionally define the *Drosophila* Integrator complex to comprise of fourteen core subunits (IntS 1-12, Asu/IntS13, IntS14), and of a regulatory CycC/Cdk8 kinase. These features are likely conserved in other metazoan species. This work also provides the first detailed structural-functional analysis of the smallest Integrator subunit and reveals an unexpected autonomous microdomain mediating the activation of the Integrator complex in snRNA 3' end formation. All of the work presented here provides a foundation for future studies to characterize the network of interactions within the Integrator Complex that are important for snRNA 3 end formation.

Chapter 2. Materials and Methods

(Partial contents described in this chapter have been published in ***RNA December 2012 18: 2148-2156***, and usage permissions have been granted from the publishers.)

Plasmids and Cell Lines Used

Plasmids and oligonucleotides (Sigma) used in these studies are described in Table 2.1 and Table 2.2, respectively. The pUB-Myc/HA/FLAG *Drosophila* expression vector was generated by cloning the ubiquitin 63E promoter (ubi-p63E) and the viral OpIE2 polyadenylation signal into the bacterial plasmid pUC19 backbone. The Green Fluorescent Protein (GFP) based snRNA readthrough reporters were constructed by cloning the snRNA gene (including the promoter, coding sequence and 3' flanking sequence) upstream of the EGFP Open Reading Frame (ORF) into the promoterless pUB (Drosophila snRNA-GFP reporter) or pcDNA6 (human U7-GFP reporter) vectors as described in detail previously (Ezzeddine et al. 2011; Albrecht and Wagner 2012). The GFP-based histone H3 reporter and Actin5C promoter driven H3 reporter were described in detail earlier (Wagner et al. 2007; Yang et al. 2009). The *Drosophila* expression clones (pUB-Myc/HA/FLAG), the yeast two hybrid clones (pGADT7 and pGBKT7) and bacterial recombinant protein clones (pET49) were constructed by standard PCR cloning procedures. Site-directed mutation-containing clones were created using the QuickChange method (Stratagene).

All cell lines and bacterial and yeast strains used in these studies are listed in Table 2.3. D.mel-2 (S2) cells from Invitrogen (Catalog No. 10831-014) were used for most of studies described here. S2 stable cell lines were generated by co-transfection of the plasmids expressing FLAG-tagged protein with a Blastocidin expressing plasmid (w/w,19:1) and selection was performed in *Drosophila* media (Invitrogen) containing 10%FBS and 25 ug/mL blastocidin. *Escherichia coli* XL-1 blue was used for standard cloning, and *E.coli* BL-21 (DE3) was used for expression of recombinant IntS12 proteins. Yeast strain AH109 (a/α) was used to perform the yeast two-hybrid assay (Clontech, Matchmaker 3).

Cell Culture and RNA Interference

S2 cells were maintained in Sf-900 II SFM medium (Invitrogen) unless otherwise mentioned following the manufacture's instructions. Double-stranded RNA (dsRNA) used for the RNA interference (RNAi) experiment in S2 cells were created by in vitro transcription of the T7 DNA templates using T7 RNA polymerase (Fermentas). All T7 DNA templates were created by PCR amplification of targeted genes off of genomic DNA using primers containing T7 promoters at the 5' end. In all cases, PCR primers were designed to amplify exons to preclude the inclusion of intronic sequences in the dsRNA. To minimize off-target effects (OTEs), the online dsRNA design tools "Snapdragon" (http://www.flyrnai.org/cgi-bin/RNAi_find_primers.pl) and "Find OTEs" (http://www.flyrnai.org/RNAi_find_frag_free.html)

Table 2.1: Plasmids used			
Name	**Description**	**Marker**	**Origin**
pUB	pUC19 backbone, ubi-63E promoter, OpIE2 polyA site	Amp	Wagner lab
pIZ-U7GFP	Promoterless pIZ vector, U7 promoter, coding region, 3' UTR followed by eGFP coding sequence (ref)	Zeo	Wagner lab
pUB-U4GFP	Promoterless pUB vector, U4 promoter, coding region, 3' UTR followed by eGFP ORF	Amp	Wagner lab
pcDNA-hU7GFP	Described by Albrecht and Wagner 2012 (ref)	Amp	Wagner lab
pIZ-H3GFP	Described by Wagner and Marzluff 2007 (Ref)	Zeo	Wagner lab
pIZ-ActH3GFP	Described by Ezzeddine and Wagner 2011 (ref)	Zeo	Wagner lab
pUB-Myc-Cdk8	Ubi-p63E promoter, *Drosophila* Cdk8 ORF	Amp	CO3, CO4
pUB-Myc-Cdk8(DN)	Ubi-p63E promoter, *Drosophila* kinase activity dead Cdk8 ORF (D173A)	Amp	WO127, WO128
pUB-Myc-IntS12	Ubi-p63E promoter, Myc, *Drosophila* IntS12 ORF	Amp	CO63, CO20
pUB-Myc-IntS12*	Ubi-p63E promoter, Myc, *Drosophila* RNAi-resistant IntS12 ORF	Amp	CO13, CO20
pUB-Myc-ΔC	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 2-251	Amp	CO13, CO14
pUB-Myc-NP	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 2-185	Amp	CO13, CO21
pUB-Myc-N	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 2-130	Amp	CO13, CO15
pUB-Myc-C	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 184-328	Amp	CO20, CO22
pUB-Myc-PC	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 124-328	Amp	CO16, CO20
pUB-Myc-ΔN	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 46-328	Amp	CO17, CO20
pUB-Myc-P	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 124-185	Amp	CO16, CO21
pUB-Myc-ΔP	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 2-130 and 184-328	Amp	CO13-CO18, CO19-CO20, CO13-CO20
pUB-Myc-ΔN15	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 16-328	Amp	CO41, CO20
pUB-Myc-ΔN30	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 31-328	Amp	CO23, CO20
pUB-Myc-N45	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 2-45	Amp	CO13, CO24
pUB-Myc-IntS12*Mt1	Ubi-p63E promoter, Myc, *Drosophila* IntS12* (16DPVLK20->AAAAA)	Amp	CO25, CO26
pUB-Myc-IntS12*Mt2	Ubi-p63E promoter, Myc, *Drosophila* IntS12* (21KAIKL25->AAAAA)	Amp	CO27, CO28
pUB-Myc-IntS12*Mt3	Ubi-p63E promoter, Myc, *Drosophila* IntS12* (26LHSSN30->AAAAA)	Amp	CO29, CO30
pUB-Myc-IntS12*Mt4	Ubi-p63E promoter, Myc, *Drosophila* IntS12* (31PTSAA35->AAAAA)	Amp	CO31, CO32

Table 2.1: Plasmids used (continued)			
Name	**Description**	**Marker**	**Origin**
pUB-Myc-IntS12*Mt5	Ubi-p63E promoter, Myc, *Drosophila* IntS12* (36ELRLL40->AAAAA)	Amp	CO33, CO34
pUB-Myc-IntS12*Mt6	Ubi-p63E promoter, Myc, *Drosophila* IntS12* (41LDEAL45->AAAAA)	Amp	CO35, CO36
pUB-Myc-IntS12*NMt1	Ubi-p63E promoter, Myc, *Drosophila* IntS12*N (16DPVLK20->AAAAA)	Amp	CO25, CO26
pUB-Myc-IntS12*NMt2	Ubi-p63E promoter, Myc, *Drosophila* IntS12*N (21KAIKL25->AAAAA)	Amp	CO27, CO28
pUB-Myc-IntS12*NMt3	Ubi-p63E promoter, Myc, *Drosophila* IntS12*N (26LHSSN30->AAAAA)	Amp	CO29, CO30
pUB-Myc-IntS12*NMt4	Ubi-p63E promoter, Myc, *Drosophila* IntS12*N (31PTSAA35->AAAAA)	Amp	CO31, CO32
pUB-Myc-IntS12*NMt5	Ubi-p63E promoter, Myc, *Drosophila* IntS12*N (36ELRLL40->AAAAA)	Amp	CO33, CO34
pUB-Myc-IntS12*NMt6	Ubi-p63E promoter, Myc, *Drosophila* IntS12*N (41LDEAL45->AAAAA)	Amp	CO35, CO36
pUB-Myc-NS2829	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 1-129 (S2829A)	Amp	CO37, CO38
pUB-Myc-NS33	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 1-129 (S33A)	Amp	CO39, CO40
pUB-Myc-NT56	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 1-129 (T56A)	Amp	CO47, CO48
pUB-Myc-NT76	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 1-129 (T76A)	Amp	CO51, CO52
pUB-Myc-NT91	Ubi-p63E promoter, Myc, *Drosophila* IntS12* residues 1-129 (T91A)	Amp	CO49, CO50
pUB-HA-IntS12	Ubi-p63E promoter, HA, *Drosophila* IntS12	Amp	CO5, CO6
pUB-FLAG-Asu	Ubi-p63E promoter, FLAG, *Drosophila* Asunder ORF	Amp	CO80, CO81
pUB-FLAG-CG4785	Ubi-p63E promoter, FLAG, *Drosophila* CG4785 ORF	Amp	CO82, CO83
pUB-FLAG-mCherry	Ubi-p63E promoter, FLAG, mCherry ORF	Amp	CO53, CO54
pUB-FLAG-Cdk8	Ubi-p63E promoter, FLAG, *Drosophila* Cdk8 ORF	Amp	CO3, CO4
pUB-FLAG-CycC	Ubi-p63E promoter, FLAG, *Drosophila* Cyclin C ORF	Amp	CO1, CO2
pUB-FLAG-IntS12*	Ubi-p63E promoter, FLAG, *Drosophila* IntS12*	Amp	CO13, CO20
pUB-FLAG-Mt3	Ubi-p63E promoter, FLAG, *Drosophila* IntS12* (26LHSSN30->AAAAA)	Amp	CO29, CO30
pUB-FLAG-Mt5	Ubi-p63E promoter, FLAG, *Drosophila* IntS12* (36ELRLL40->AAAAA)	Amp	CO33, CO34
pUB-FLAG-ΔN	Ubi-p63E promoter, FLAG, *Drosophila* IntS12* residues 46-328	Amp	CO17, CO20
pUB-FLAG-N45	Ubi-p63E promoter, FLAG, *Drosophila* IntS12* residues 1-45	Amp	CO13, CO24
pUB-FLAG-N45Ch	Ubi-p63E promoter, FLAG, *Drosophila* IntS12* residues 1-45 fused to mCherry	Amp	CO13-CO55, CO53-CO54
pUB-FLAG-12NmCh	Ubi-p63E promoter, FLAG, Drosophila IntS12 residues 1-129, mCherry	Amp	CO13-CO56, CO53-CO54
pUB-FLAG-12NT76mCh	Ubi-p63E promoter, FLAG, Drosophila IntS12 residues 1-129 (T76A), mChe	Amp	CO61, CO62

Table 2.1: Plasmids used (continued)			
Name	**Description**	**Marker**	**Origin**
pUB-FLAG-12NPmCh	Ubi-p63E promoter, FLAG, Drosophila IntS12 residues 1-185, mCherry	Amp	CO13-CO57, CO53-CO54
pUB-FLAG-12NPLmCh	Ubi-p63E promoter, FLAG, Drosophila IntS12 residues 1-185 (L145A), mCherry	Amp	CO74, CO75
pUB-FLAG-12NPEmCH	Ubi-p63E promoter, FLAG, Drosophila IntS12 residues 1-185 (E147A), mCherry	Amp	WO811, WO812
pUB-FLAG-mCh12C	Ubi-p63E promoter, FLAG, mCherry, Drosophila IntS12 residues 184-328	Amp	CO58-CO20, CO59-CO60
pGADT7-Asu	ADH1 promoter, GAL4 activation domain, HA epitope tag, Drosophila Asunder ORF	Amp, LEU2	WO, WO
pGADT7-CG4785	ADH1 promoter, GAL4 activation domain, HA epitope tag, Drosophila CG4785 ORF	Amp, LEU2	WO, WO
pGADT7-IntS12	ADH1 promoter, GAL4 activation domain, HA epitope tag, Drosophila IntS12	Amp, LEU2	CO76, CO77
pGADT7-IntS12Mt3	ADH1 promoter, GAL4 activation domain, HA epitope tag, Drosophila IntS12 (26LHSSN30->AAAAA)	Amp, LEU2	CO76, CO77
pGADT7-IntS12Mt5	ADH1 promoter, GAL4 activation domain, HA epitope tag, Drosophila IntS12 (36ELRLL40->AAAAA)	Amp, LEU2	CO76, CO77
pGADT7-IntS12ΔN	ADH1 promoter, GAL4 activation domain, HA epitope tag, Drosophila IntS12 residues 46-328	Amp, LEU2	CO78, CO77
pGADT7-IntS12N45	ADH1 promoter, GAL4 activation domain, HA epitope tag, Drosophila IntS12 residues 1-45	Amp, LEU2	CO76, CO79
pGBKT7-IntS1	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS1 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS2	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS2 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS3	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS3 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS4	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS4 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS5	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS5 ORF	Kan, TRP1	Warren lab

Table 2.1: Plasmids used (continued)			
Name	**Description**	**Marker**	**Origin**
pGBKT7-IntS6	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS6 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS7	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS7 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS8	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS8 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS9	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS9 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS10	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS10 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS11	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS11 ORF	Kan, TRP1	Warren lab
pGBKT7-IntS12	ADH1 promoter, GAL4 DNA binding domain, c-Myc epitope tag, Drosophila IntS12 ORF	Kan, TRP1	Warren lab
pET49b(+)	T7 promter,N-terminal GST-tag, His-tag	Kan	Novagen
pET49-IntS12PHD	T7 promter,N-terminal GST-tag, His-tag, *Drosophila* IntS12 PHD finger, residues 120-193	Kan	CO9, CO10
pET49-hIntS12PHD	T7 promter,N-terminal GST-tag, His-tag, human IntS12 PHD finger, residues 130-224	Kan	CO11, CO12
pET49-IntS12	T7 promter,N-terminal GST-tag, His-tag, *Drosophila* IntS12	Kan	CO5, CO6
pET-12PHDV	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (V133A)	Kan	WO803, WO804
pET-12PHDM	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (M137A)	Kan	WO805, WO806
pET-12PHDN	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (N143A)	Kan	WO807, WO808
pET-12PHDL	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (L145A)	Kan	CO74, CO75
pET-12PHDE	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (E147A)	Kan	WO811, WO812
pET-12PHDY	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (Y155A)	Kan	WO813, WO814

Table 2.1: Plasmids used (continued)			
Name	**Description**	**Marker**	**Origin**
pET-12PHDH	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (H160A)	Kan	CO7, CO8
pET-12PHDP	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (P162A)	Kan	CO72, CO73
pET-12PHDD	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (D171A)	Kan	CO64, CO65
pET-12PHDW	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (W177A)	Kan	CO66, CO67
pET-12PHDCCH	T7 promter,N-terminal GST-tag, His-tag, Drosophila IntS12 PHD finger, residues 120-193 (C148A,C151A,H156A)	Kan	CO68-CO69, CO70-CO71

Table 2.2: Oligonucleotides used	
Cloning Oligos	**Sequence**
CO1	GATGAATTCTGGCGGGCAATTTTTGGCAG
CO2	GATTCTAGACTAACGCTGAGGCGGTGG
CO3	GCGACTAGTCGACTACGATTTCAAGATGAAAAC
CO4	GCGTCTAGAGTTGAAGCGCTGGAAGTTC
CO5	GCGTCTAGAATGGCCGCAAATATAGCC
CO6	GCGAAGCTTTTACTGCTTGGATCTGCG
CO9	GTAGAATTCTACCGGTGACACCGGCGAC
CO10	GCGAAGCTTTTATGTTGTCCTCCCGCTGCT
CO11	GCGGAATTCTGCTGATGATTTTGCCATG
CO12	CGCAAGCTTTTACTGAGTTTTTTGAGCCAT
CO13	GCGGAATTCTGGCGGCCAACATCGCGGC
CO14	CGCTCTAGAGCTGCTACTGCTGCTGGC
CO15	GCGTCTAGACAGATCTCCGAAATCTCC
CO16	CGCGAATTCTGGGAGATTTCGGAGATCTG
CO17	GATGAATTCTGCTGAAAGCCCGCTTCGGC
CO18	CGTTGGCTTGTTCAGATCTCCGAAATCTCC
CO19	TTCGGAGATCTGAACAAGCCAACGAGCAGC
CO20	CGCTCTAGATTACTGCTTGGATCTGCG
CO21	CGCTCTAGACTTGTTGCAGCACGTGTC
CO22	GCGGAATTCTGAACAAGCCAACGAGCAGC
CO23	GATGAATTCTGCCGACGAGCGCCGCGGAG
CO24	CGCTCTAGACAGGGCTTCATCCAGCAG
CO53	CATGCGGCCGCTGTGAGCAAGGGCGAGGAG
CO54	CGCTCTAGATTACTTGTACAGCTCGTCCATG
CO55	CATGCGGCCGCCAGGGCTTCATCCAGCAG
CO56	CATGCGGCCGCCAGATCTCCGAAATCTCC
CO57	CATGCGGCCGCCTTGTTGCAGCACGTGTC
CO58	CATGCGGCCGCTAACAAGCCAACGAGCAGC
CO59	GCGGAATTCTGGTGAGCAAGGGCGAGGAG
CO60	CATGCGGCCGCCTTGTACAGCTCGTCCATG
CO63	GCGGAATTCTGATGGCCGCAAATATAGCC
CO76	GATGAATTCGCGGCCAACATCGCGGCG
CO77	GCGGGATCCTTACTGCTTGGATCTGCGC
CO78	GCGGAATTCCTGAAAGCCCGCTTCGGC
CO79	CGCGGATCCTTACAGGGCTTCATCCAGCAG
CO80	GGCCACTAGTCATGTTCGAACGCAACCAGAAG
CO81	GGCCGAATTCTTAACTACGTACGGATTCCTCC
CO82	GGCCGGATCCCATGCTGCGCCCGGTGCCGGG
CO83	GGCCGAATTCTCAATACATGTATGCAGGAGC
QuickChange Oligos	
CO7	GGTGCCATGTACGCTCAGGAGTGCCAC
CO8	GTGGCACTCCTGAGCGTACATGGCACC

CO25	CGCGCAGGAGGTCGCCGCGGCCGCCGCAAAAGCGATTAAAC
CO26	GTTTAATCGCTTTTGCGGCGGCCGCGGCGACCTCCTGCGCG
CO27	CCGGTCCTCAAAGCAGCGGCTGCAGCGCTCCATAGCTCC
CO28	GGAGCTATGGAGCGCTGCAGCCGCTGCTTTGAGGACCGG
CO29	GCGATTAAACTGGCCGCTGCCGCCGCCCCGACGAGCGCC
CO30	GGCGCTCGTCGGGGCGGCGGCAGCGGCCAGTTTAATCGC
CO31	CATAGCTCCAACGCGGCGGCCGCCGCGGAGCTC
CO32	GAGCTCCGCGGCGGCCGCCGCGTTGGAGCTATG
CO33	CGAGCGCCGCGGCGGCCGCCGCCGCGCTGGATGAAGC
CO34	GCTTCATCCAGCGCGGCGGCGGCCGCCGCGGCGCTCG
CO35	CTCCGCCTCCTGGCGGCTGCAGCCGCGAAAGCCCGCTTC
CO36	GAAGCGGGCTTTCGCGGCTGCAGCCGCCAGGAGGCGGAG
CO37	AAGTTGCTGCACGCGGCTAATCCCACCTCG
CO38	CGAGGTGGGATTAGCCGCGTGCAGCAACTT
CO39	AGTAATCCCACCGCGGCGGCCGAACTG
CO40	CAGTTCGGCCGCCGCGGTGGGATTACT
CO47	GAGAAAAGTTTGGCCAACAACATGACG
CO48	CGTCATGTTGTTGGCCAAACTTTTCTC
CO49	ATCATCAATTTGGCCAACTCACCGGAC
CO50	GTCCGGTGAGTTGGCCAAATTGATGAT
CO51	GGACGTGCTGCCGCGCCCCCGCAGCAG
CO52	CTGCTGCGGGGGCGCGGCAGCACGTCC
CO61	GGGCCGCGCCGCGGCCCCGCCCCAACAAC
CO62	GTTGTTGGGGCGGGGCCGCGGCGCGGCCC
CO64	AAGGAGGAGGCGGCCGCTGACCAGGAGCAGAAC
CO65	GTTCTGCTCCTGGTCAGCGGCCGCCTCCTCCTT
CO66	GACCAGGAGCAGAACGCGCAGTGCGACACGTGC
CO67	GCACGTGTCGCACTGCGCGTTCTGCTCCTGGTC
CO68	ATGTACGCTCAGGAGGCCCACAAGCCGCCCATA
CO69	TATGGGCGGCTTGTGGGCCTCCTGAGCGTACAT
CO70	AATCGGCTGATTGAGGCCTCCAAGGCCGGTGCC
CO71	GGCACCGGCCTTGGAGGCCTCAATCAGCCGATT
CO72	CAGGAGTGCCACAAGGCGCCCATAACCAAGGAG
CO73	CTCCTTGGTTATGGGCGCCTTGTGGCACTCCTG
CO74	ACGGCCACCAATCGGGCGATTGAGTGCTCCAAG
CO75	CTTGGAGCACTCAATCGCCCGATTGGTGGCCGT
WO811	CCAATCGGCTGATTGCGTGCTCCAAGTGCGG
WO812	CCGCACTTGGAGCACGCAATCAGCCGATTGG
WO127	GCCGCGTAAAAATCGCCGcCATGGGTTTCGCGCGGCTC
WO128	GAGCCGCGCGAAACCCATGGCGGCGATTTTTACGCGGC
WO803	CGACCTCAATTGCTGCGCGTGCGGCGAGATGG
WO804	CCATCTCGCCGCACGCGCAGCAATTGAGGTCG
WO805	CTGCGTGTGCGGCGAGGCGGTTTTCACGGCCAC
WO806	GTGGCCGTGAAAACCGCCTCGCCGCACACGCAG
WO807	GGTTTTCACGGCCACCGCTCGGCTGATTGAGTGC

WO808	GCACTCAATCAGCCGAGCGGTGGCCGTGAAAACC
WO813	CAAGTGCGGTGCCATGGCCCATCAGGAGTGCCAC
WO814	GTGGCACTCCTGATGGGCCATGGCACCGCACTTG
PCR Oligos for endogenous misprocessed/unprocessed snRNA	
U1-pF	GCTGAGTTGACCTCTGCGATTA
U1-pR	CTTTTAAAATTTATTGCAGATGTCGG
U2 pF	CCCGGTATTGCAGTACCGCCGGGA
U2 pR	CAAAGGACACTTTCGACATGTC
U4 pF	GGTGGCAATACCGTAACCAAT
U4 pR	GGCTAAGACAACCGTCATATTAA
U5 pF	CGCCTTTTACTAAAGATTTCCGTGG
U5 pR	CCATGTATATGACCACCAGACC
Rps17 for	CGAACCAAGACGGTGAAGAAG
Rps17 rev	CCTGCAACTTGATGGAGATACC
Realtime PCR Oligos for Integrator mRNAs	
IntS2-1-f	GAGCTGAAGAAGGAGCTGCAG
IntS2-1-r	CAGATCCTCGGGATGCTTGAG
IntS3-1-f	GAGATGCATGATATGTTGTCAC
IntS3-1-r	GGCGTACAAAATCTCTTCCTAG
IntS4-1-f	CGTCAACAGGTGTCATCG
IntS4-1-2-r	GCTCGAGTCCTGCGAGTC
IntS5-2-f	CGGAAAATCGCGAGAACC
IntS5-2-r	CAGTGCGCGCAGTAGCAC
IntS6-1-f	CGCTGGGATCAGCGCTTG
IntS6-1-r	GACGGGGAAGCTGCGCAC
IntS7-1-f	CATCTGGACAAGATCCTCAAC
IntS7-1-r	GTATGCGTACTGGTGTCC
IntS8-2-f	CTGCTGGAGAAGTTCCAGC
IntS8-2-r	CTTGGAGTGACACATCCAGTC
IntS10-1-1-f	GTTGTACATGGTCAAGGAG
IntS10-1-2-r	CAGATTTGGCGAAAGTTCC
IntS13-1-f	CTCCAAGAAGGGACTGGTC
IntS13-1-2-r	GCTGGAGTTGGCGTTCTG
IntS14-2-f	CACGAACGTGATTCCCTGG
IntS14-2-r	GTAAACACAGTCTTAAAGTAAGGAC

Table 2.3: Cell Lines		
Cell line	**Description**	Reference
D.mel-2 (S2) cell	Drosophila Schneider 2 (S2) Cells (Catalog no. R690-07)	Invitrogen
S2 (FLAG)	S2 stable cell line expressing FLAG-tagged	This Study
S2 (FLAG-mCh)	S2 stable cell line expressing FLAG-tagged mCherry	This Study
S2 (FLAG-Asu)	S2 stable cell line expressing FLAG-tagged Asunder/IntS13	This Study
S2 (FLAG-CG4785)	S2 stable cell line expressing FLAG-tagged CG4785/IntS14	This Study
S2 (FLAG-CycC)	S2 stable cell line expressing FLAG-tagged cycline C	This Study
S2 (FLAG-Cdk8)	S2 stable cell line expressing FLAG-tagged Cdk8	This Study
S2 (FLAG-IntS12*)	S2 stable cell line expressing FLAG-tagged RNAi-resistant full-length IntS12	This Study
S2 (FLAG-Mt3)	S2 stable cell line expressing FLAG-tagged IntS12* mutant (Mt3)	This Study
S2 (FLAG-Mt5)	S2 stable cell line expressing FLAG-tagged IntS12* mutant (M5)	This Study
S2 (FLAG-ΔN)	S2 stable cell line expressing FLAG-tagged IntS12* microdomain deletion mutant	This Study
S2 (FLAG-N45)	S2 stable cell line expressing FLAG-tagged IntS12* microdomain	This Study
S2 (FLAG-N45Ch)	S2 stable cell line expressing FLAG-tagged mCherry fused IntS12 microdomain	This Study
S2 (FLAG-12NmCh)	S2 stable cell line expressing FLAG-tagged mCherry fused IntS12 N-terminus	This Study
S2 (FLAG-12NT76mCh)	S2 stable cell line expressing FLAG-tagged mCherry fused IntS12 N-terminus phosphorylation mutant	This Study
S2 (FLAG-12NPmCh)	S2 stable cell line expressing FLAG-tagged mCherry fused IntS12 C-terminus deletion mutant	This Study
S2 (FLAG-12NPLmCh)	S2 stable cell line expressing FLAG-tagged mCherry fused IntS12 C-terminus deletion mutant with point mutation in L145	This Study
S2 (FLAG-12NPEmCH)	S2 stable cell line expressing FLAG-tagged mCherry fused IntS12 C-terminus deletion mutant with point mutation in E147	This Study
S2 (FLAG-mCh12C)	S2 stable cell line expressing FLAG-tagged mCherry fused IntS12 C-terminus	This Study
HeLa	HeLa cells (Catalog no. CCL-2)	ATCC
Strains	**Genotype**	Reference
AH109	MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2 : : GAL1UAS-GAL1TATA-HIS3, GAL2UAS-GAL2T A T A-ADE2, URA3 : : MEL1UAS-MEL1 TATA-lacZ	James et al., 1996
E.coli XL-1	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB laclqZΔM15 Tn10 (Tetr)]	Stratagene
E.coli BL-21	E. coli B F- dcm ompT hsdS(rB- mB-) gal λ(DE3[lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	New England BioLabs

from Harvard Drosophila RNAi Screening Center (DRSC) were used. *In vitro* transcribed dsRNA was purified using TRIzol reagent (Invitrogen) and the quality of dsRNA was assessed by electrophoresis or using the Nanodrop spectrophotometer (Thermo Scientific). To perform RNAi in S2 cells, $5x10^4$ cells per 96 well in 100µl of Sf-900 II media or $3x10^5$ cells per 24 well in 300µl medium were plated in each well and supplemented with 1µg (96-well) or 3µg (24-well) dsRNA. The next day, an additional 1µg or 3µg of dsRNA was added and then cells were harvested in the fourth day for qRT-PCR or Western blot analysis. Cells were harvested for analysis using TRIzol for total RNA extraction or radioimmunoprecipitation assay (RIPA) buffer [50 mM Tris-HCl pH 8, 150 mM NaCl, 1% NP-40, 0.5% Na deoxycholate, 0.1% SDS] for protein lysates.

HeLa cells were cultured in Dulbecco's modified Eagle medium (DMEM) (Invitrogen, Carlsbad CA) supplemented with 1% penicillin-streptomycin (pen/strep) (Invitrogen) and 10% fetal bovine serum (FBS; Phenix Research, NC). RNAi experiments in HeLa cells were performed by the use of the manufacturer's instructions for Lipo-fectamine 2000 with minor modifications. Briefly, 3µl of 20 µM siRNAs were mixed with 47µl of Opti-MEM (Invitrogen) in tube A, and 12µl of Opti-MEM and 3µl of Lipofectamine 2000 were mixed in Tube B. Both tubes were incubated at room temperature for 7 min, and then the contents were mixed and incubated at room temperature (RT) for an additional 25 min. After this incubation, 38µl of Opti-MEM was added and the mixture (100µl/tube) was pipetted dropwise into the wells. For siRNA transfection, $8.5x10^4$ cells per well were initially plated in 24-well plates, and in second day after the RNAi treatment, cells were re-plated into a 6-well plate for each well containing 2 ml DMEM complete media. The next day, the same amount of siRNA and 500 ng human U7-GFP reporter cells were co-transfected into each well under the conditions described above. Two days after the second siRNA transfection, cells were harvested for Western blot analysis.

Quantitative Real Time-PCR (qRT-PCR) Analysis

Total RNA was isolated from *Drosophila* S2 cells using TRIzol reagent, and 2 µg of RNA was treated with DNase I (Fermentas, Burlington, ON, Canada) in accordance with the manufacturer's recommendations. Reverse transcription was performed on the 2 µg of RNA using random hexamer primers and MMLV reverse transcriptase (Invitrogen) in a total volume of 20µl at 37°C for 1 h, followed by incubation at 95°C for 5 min. In each case, 2µl of the reverse transcription product was used for real-time PCR reaction using SYBR green master mix (Fermentas) and specific primers to each amplicon tested (for oligonucleotides, see Table 2.2). Data were acquired using a Stratagene Mx3000P real-time PCR machine, and were analyzed by the $\Delta\Delta CT$ (CT, threshold cycle) method using the equation fold change = $2^{-[(CT_{GOI} - CT_{norm\,control}) - (CT_{GOI} - CT_{norm\,treated})]}$, where CT_{GOI} is the CT for the gene of interest and CT_{norm} is the normalized CT. Triplicate experiments were performed for each Integrator knockdown, and data are plotted as fold increases in amplicon CT values normalized to the reference gene *RpS17* versus those for the control dsRNA-treated cells, which were also normalized to *RpS17*. (*An excerpt from: Ezzeddine, N., et al. (2011). "A subset of Drosophila*

integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation." Mol Cell Biol ***31****(2): 328-341.)*

Immunoprecipitation

For immunoprecipitation of FLAG-tagged protein, 1 mg nuclear extracts were incubated with 10 µl of anti-FLAG antibody-conjugated agarose beads (anti-FLAG M2 affinity gel, Sigma) with constant rotation for 2 hrs at 4 °C in buffer D [20 mM HEPES pH7.9, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF] plus 0.1% Triton X-100. Beads were washed twice with 1XTris Buffered Saline [50 mM TrisHCl pH 7.4, 150 mM NaCl] and then twice with buffer D plus 0.1% Triton X-100. During last wash, switch to new tube. For immunoprecipitation of endogenous protein, 2 µg antigen purified antibody was pre-incubated with 20 µl Protein A/G beads (Santa Cruz Biotechnology, Inc.) with constant rotation in buffer D at 4 °C for 1 hr, and then 400 µg nuclear extracts were added and incubated with rotation for another 2 hrs at 4 °C. Wash condition is same as above mentioned. Finally 50 µl of 1XSDS loading buffer was added to the beads and boil at 95 °C for 3 minutes and resolved in 12.5% SDS-PAGE gel. Western blot analysis was performed using standard procedure. (*An excerpt from: Chen, J., et al. (2012). "An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation."* RNA*.)*

Yeast Two-Hybrid Analysis

Full-length coding regions of the twelve *Drosophila* Integrator subunits were obtained from the *Drosophila* Genomics Resource Center and cloned in frame into the pGBKT7 vector (BD), Asu/Ints13 and CG4785/Ints14 were cloned into pGADT7 (AD). IntS12 full length, first 45 amino acid microdomain (N45), N-terminal truncation (ΔN), and functional deficient alanine-scanning mutants Mt3 and Mt5 were cloned into pGADT7 (AD) using standard methods. Pairwise co-transformations of AD and BD constructs into yeast strain AH109 were according to manufacturer's instructions (Matchmaker 3 system, Clontech). Empty vectors were co-transformed to control for construct autoactivation. Positive interactions were analyzed using nutritional selection by spotting four serial ten-fold dilutions on SD- plates lacking Leucine/Tryptophan (vector control), or additionally Histidine (medium stringency) and Histidine/Adenine (high stringency). Images were taken after three or five days incubation at 30 °C. To determine the HA-tagged protein expression, yeast transformants were harvested at OD600 of ~0.8 in SC-Leu liquid culture, lysed using the glass beads method and tagged proteins were detected by western blotting using anti-HA antibody (Covance, Princeton NJ). (*An excerpt from: Chen, J., et al. (2012). "An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation."* RNA*.)*

Genome-wide RNAi Screen in S2 Cells

The dsRNA library was purchased from Open Biosystems (Thermo Scientific, Waltham, MA) as 5 µg per well resuspended in 100 µl of water. The dsRNA was realiquoted into 167 96-well plates (Greiner Bio-One, Monroe, NC) at 1µg per well and allowed to dry down. Then $5x10^4$ cells were plated into each well and allowed to incubate for three days. On the fourth day, the 30 ng of U7-GFP reporter was transfected into each well using Effectene according to manufacturer's instructions (Qiagen, Hilden, Germany). Two days following reporter transfection, cells were imaged live using a Cellomics Array Scan V highthroughput automated microscope (Thermo Scientific). Exposure times were calculated on a plate-to-plate basis and

fluorescence levels were calculated using pre-written algorithms incorporated into the manufacturer's software. The two algorithms used in the data presentation are the standard deviation in fluorescence intensity and the mean differential intensity as these two methods were most effective in blindly identifying positive controls in pilot assays. (*An excerpt from: Chen, J., et al. (2012). "An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation." RNA.)*

S2 Cell Nuclear Extracts Preparation

S2 cell nuclear extracts were prepared using previously described methods (Dignam et al. 1983) with minor modifications. In brief, harvested S2 cells were washed twice with cold 1xPBS and resuspended in 5 times packed cell volume (PCV) of ice-cold buffer A [10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF] for 30 min on ice to swell. Swollen cells were further disrupted by grinding 35 strokes using the glass dounce tissue grinder and then cell nuclei were pelleted by centrifuging at 2000 rpm for 5 min. The crude nuclei pellet were further resuspended in ½ pellet volume of buffer C [20 mM HEPES pH 7.9, 25% glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF] and incubated 1 hr at 4 °C with rotation for extraction. Extracted supernatants were collected by centrifugation at 14000 rpm for 30 min and final nuclear extracts were dialyzed overnight in buffer D [20 mM HEPES pH7.9, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF]. Dialyzed nuclear extracts were stored at -80 °C for future use.

Site-directed Mutagenesis

Site-directed mutations of Cdk8 and IntS12 were generated using QuickChange Kit (Stratagene) protocol. In brief, 25 ng of template plasmids were used for PCR amplification using quickchange oligos listed in Table 2.2 in a total volume of 50 µl reaction, and 1 µl FastDigest DpnI (Fermentas) was used to digest the methylated template DNA at 37 °C for 30 min. 1 µl of DpnI treated PCR product was used for transformation, and the right clones were screened and confirmed by direct sequencing.

S2 Cell Immunofluorescence Microscopy

The method to prepare S2 cells for immunofluorescence microscopy was described by Rogers & Rogers (Rogers and Rogers 2008). Briefly, $1x10^5$ S2 cells in 100 µl Sf-900 II medium were allowed to attach to the Concanavalin A coated round coverslips for 1hr at 28 °C, and followed by two washes with 1xPBS. Cells were fixed by treating with 10% paraformaldehyde diluted in 1xPBS for 10 min at room temperature. Fixed cells were then permeabilized by washing three times with 1xPBS supplemented with 0.1% Triton X-100. Permeabilized cells were then blocked in 1xPBST containing 5% normal goat serum, and

probed by anti-FLAG M2 antibodies (Sigma) with 1:1000 dilutions in the blocking buffer for 1hr at room temperature. Cells were washed using 1xPBST three times for 10 min each, and probed with light sensitive Cy3-conjugated anti-mouse secondary antibodies (1:1000) for 45 min. Cells were washed again twice in 1xPBST and the final wash was supplemented with 4',6-diamidino-2-phenylindole (DAPI) (1:1000). Finally, washed coverslips were mounted facing down on glass slides using fluoro-gel mounting medium (Electron Microscopy Sciences, PA) and then imaged using fluorescence microscopy. FLAG tagged proteins were visualized at excitation wavelength 550 nm, and the nuclear DAPI staining was observed at 350 nm.

GST Recombinant Protein Preparation

pET49 clones expressing GST, GST-tagged full-length IntS12, or various IntS12 PHD fingers were transformed into *E.coli* BL-21(DE3). Bacterial culture were grown at 37 °C in 1liter of LB medium [1% tryptone, 0.5% yeast extract, and 1% NaCl] supplemented with Kanamycin (50 µg/ml) to OD600 of ~0.6. To induce the expression of GST-tagged recombinant protein, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the cultures to at a final concentration of 1mM, and continued to culture for another 2 hrs at 37 °C. Induced cultures were pelleted and resuspended in 30 ml of cold PBS [140 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, supplemented with10 mM DTT, and 1 mM phenylmethylsulfonyl fluoride (PMSF), 1% Triton X-100, pH 8]. Culture suspensions were lysed by using EF-C3 high-pressure homogenizer (Avestin, Germany), and crude lysates were cleared by centrifuging 15 min at 12,000xg at 4 °C. Cleared lysates were incubated with 0.5 ml glutathione agarose beads with rotation at 4 °C for 1 hr. Protein bound beads were pelleted and washed three times for 15 min each with cold PBS. To elute the recombinant GST-tagged proteins, washed beads were resuspended in 1ml Elution buffer [50 mM Tris HCl pH 8, 10 mM reduced glutathione] and incubated at room temperature for 10 min. The supernatants were collected after centrifugation and the concentration of recombinant proteins were determined by Bio-rad protein assay (Bio-rad Laboratory, California). The purified recombinant proteins were supplemented with glycerol to 50% [v/v] and stored at -80 °C for later use.

GST Pulldown Assay

The assay for GST-tagged recombinant protein pulldown of histones was carried out by incubation of 2 µg of purified GST-tagged proteins with 10 µg of global histones (Worthington, NJ) for 3 hrs at 4 °C in 300 µl incubation buffer [50 mM Tris-HCl pH7.5, 500 mM NaCl, 1% NP-40, 0.5 mM EDTA, 1mM PMSF]. Then 20 µl glutathione beads were added and incubated for 1 hr. Beads were collected by centrifugation and washed three times for 5 min each with 1 ml incubation buffer. Washed beads were resuspended in 50 µl 1xSDS loading buffer and boiled

for 3 min at 95 °C. Boiled samples were centrifuged and 20 µl of supernatant were resolved in 15% SDS-PAGE. Western blot analysis was conducted using the standard procedure. Anti-GST serum were generated by injection of Guinea pig with recombinant GST proteins and anti-histone antibodies [H2A (Upstate 07-146), H2B (Upstate 07-371), H3 (Abcam ab1791) and H4 (Activemotif 39269), H1 (Upstate 05-457)] were kindly provided by Dr. Barton's Lab at the University of Texas MD Anderson Cancer Center.

Biotin-labeled Histone Peptide Pulldown Assay

Biotinylated histone peptide pulldown of recombinant GST-tagged IntS12 PHD finger proteins was carried out by incubation of 1 µg GST-tagged proteins with 1 µg of biotinylated histone peptides for 4 hrs at 4 °C in 300 µl binding buffer [50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% NP-40]. Then 15 µl Streptavidin sepharose beads (Amersham) were added to the binding reaction and incubated for another hour. Beads were collected by centrifugation and wash three times for 5 min each with 1 ml binding buffer. Washed beads were resuspended in 60 µl 1xSDS loading buffer and boiled for 3 min at 95 °C. Boiled samples were centrifuged and 10 µl supernatant were resolved in 12.5% SDS-PAGE gel. Biotinylated histone peptides (H3, 1-21; H3, 22-44; H3, 66-88; H4, 1-25; H3K4, K9, K27, K36, K79 and H4K20 mono-, di- or tri-methylated peptide) were kindly provided by Dr. Shi's Lab at the University of Texas MD Anderson Cancer Center.

Chromatin Immunoprecipitation (CHIP)

Cultured S2 cells were crosslinked with 1% formaldehyde for 7 min at room temperature and the crosslinking reaction was quenched by adding glycine to a final concentration of 125 mM for 10 min. Fixed S2 cells were collected and washed twice with ice-cold PBS and resuspended in cold sonication buffer [0.5% SDS, 20 mM Tris-HCl pH 8, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM PMSF] in a ratio 10^8 cells per 1 ml buffer for 10 min on ice. Lysed cells were further sonicated for 12 min (30 sec on/2.5 min off) using the Ultrasonic Converter C5749 and clarified by centrifugation at 14 000 rpm for 10 min. The clarified supernatant was used immediately for immunoprecipitation or stored at -80 °C.

75 µl sonicated chromatin was diluted in 1 ml IP buffer [0.5% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl, 10% glycerol] and pre-cleared by incubating with 30 µl 50% Protein A/G beads slurry for 2 hrs at 4 °C on a rotator before overnight incubation with 2 µg of antigen purified IntS12 antibodies or the M2 resin. The bound materials were washed rigorously once with Low Salt buffer [0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl], three times with High Salt buffer [0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 500 mM NaCl], once with LiCl buffer [2 mM EDTA, 20 mM Tris-

HCl pH 8, 1% Sodium Deoxycholate, 1% NP-40, 250 mM LiCl], twice with Tris-EDTA buffer [10 mM Tris-HCl pH 8, 1 mM EDTA], and eluted 15 min at room temperature with 500 µl Elution buffer [1% SDS, 0.1% $NaHCO_3$]. Eluted fractions were reverse crosslinked by adding 20 µl of 5 M NaCl and incubate at 65 °C in the water bath for at least 4 hrs, and then treated with proteinase K and purified by phenol/chloroform extraction followed by ethanol precipitation. Finally, purified DNA is resuspended in 150 µl ddH_2O, and 2 µl was used for qPCR analysis.

Chapter 3. A Functional RNAi Screen Identifies *Drosophila* Genes Required for snRNA 3' End Formation.

(Partial contents described in this chapter have been published in ***Mol. Cell. Biol. 31, 2011, 328-341, doi:10.1128/MCB.00943-10*** and ***RNA December 2012 18: 2148-2156***, and usage permissions have been granted from the publishers.)

INTRODUCTION

RNA processing is a fundamental biological process required to generate mature RNA species. The U snRNA are non-coding transcripts that play important roles in different RNA processing events. They form the RNA core of spliceosome to help remove introns from mRNA precursors and the U7 snRNA is involved in the 3' end processing of the replication-dependent histone pre-mRNA (Reviewed in (Matera et al. 2007; Marzluff et al. 2008)). Most snRNAs are transcribed by RNA polymerase II, and their 3' end is formed by a single endonucleolytic cleavage at the 3' end cleavage site from the elongating polymerase to allow for release and subsequent export to the cytoplasm (Reviewed in (Egloff et al. 2008)).

Studies in the past decades have established the requirement of three features for snRNA 3' end processing: an snRNA promoter, a 3' box sequence element located 9-19 nt downstream of the mature snRNA 3' end, and the CTD of RNAPII largest subunit Rpb1 (Hernandez 1985; de Vegvar et al. 1986; Hernandez and Weiner 1986; Medlin et al. 2003; Jacobs et al. 2004; Egloff et al. 2007; Egloff et al. 2010)). Promoter-swap experiments reveal that replacement of either human or *Drosophila* snRNA promoters with other RNAPII promoters significantly compromises proper snRNA 3' end formation (de Vegvar et al. 1986; Hernandez and Weiner 1986; Ezzeddine et al. 2011)). The 3' end formation of viral-encoded snRNA-like transcripts in *Herpesvirus saimiri* has recently been reported to be highly dependent on the snRNA promoter as well (Cazalla et al. 2011), further supporting the importance of the snRNA promoter for their 3' end formation. Conserved promoter coupling indicates that the 3' end processing machinery is loaded onto the snRNA promoter early in the transcription cycle. The 3' box sequence element has been determined to be required for proper snRNA 3' end processing though was found to be highly tolerant to mutation (Ach and Weiner 1987; Ezzeddine et al. 2011). It very likely serves as a binding site for the protein complex that carries out the cleavage process for snRNA. Finally, phosphorylation of the Ser2 and Ser7 but not Ser5 in the heptad repeats of the Rpb1 CTD is reported to be important for snRNA 3' end formation (Egloff et al. 2007; Egloff et al. 2010).

Biochemical purification and functional identification of the Integrator complex in snRNA 3' end formation provides new insight into the snRNA 3' end processing events (Baillat et al. 2005; Ezzeddine et al. 2011). Twelve different members were initially identified in the complex associated with the RNAPII CTD and two of them (IntS1 and IntS11) were shown to be functionally required for U1 and U2 snRNA 3' end formation. Among the twelve members, IntS9 and IntS11 are homologous to the poly(A)+ mRNA and poly(A)- histone mRNA 3' end processing factors CPSF100 and CPSF73 (Mandel et al. 2006; Sullivan et al. 2009). IntS9/11 have been shown to form a heterodimer through their unique C-terminal regions and this

interaction is required for the endonucleatytic cleavage activity of the Integrator complex (Albrecht and Wagner 2012). Other than these two proteins, the remainder of the ten founding members in the complex display insignificant similarity with the known factors involved in either poly(A)+ mRNA or histone mRNA 3' end formation, making their role in 3' end formation of snRNA difficult to anticipate. Recent high-throughput mass spectrometry analysis of Integrator immunoprecipitations from mammalian cells has identified a number of additional factors associated with the core Integrator subunits (Malovannaya et al. 2010; Malovannaya et al. 2011), and these include additional RNAPII subunits, a group of phosphatases, OB-fold nucleic acid binding proteins, zinc-finger proteins and many others either uncharacterized and/or not previously known to be involved in snRNA biogenesis. Functional validation of these factors has not been done, making their importance to snRNA 3' end formation unclear.

In this chapter, I describe a functional RNAi screen in *Drosophila* S2 cells to identify genes required for snRNA 3' end formation by utilizing the U7-GFP readthrough reporter assay (Ezzeddine et al. 2011). Our screen scored ten of the twelve known Integrator subunits, and also identified four new non-Integrator subunits necessary for snRNA 3' end formation. Follow-up systematic functional analysis confirmed the functional requirement for 10 of the 12 known Integrator subunits. We also determined that the four novel factors CG4785, Asunder (Asu), cyclin C (CycC), and Cdk8, are required for snRNA 3' end formation. Further analysis demonstrates that Asunder and CG4785 are additional core components of the known Integrator complex, and the Cdk8 kinase activity is required for proper snRNA 3' end processing. We further discovered that CycC/Cdk8 play a conserved role in both fly and human snRNA 3' end formation and it is likely that this cognate CycC/Cdk8 pair exerts it function through a novel pathway independent of the canonical Mediator Cdk8 module. Taken together, this study redefines the *Drosophila* Integrator complex as comprising of fourteen core subunits and demonstrates a novel role of CycC/Cdk8 in snRNA 3' end formation unrelated to its known function in the mediator complex.

RESULTS

(*Excerpts in Results section are from: Chen, J., et al. (2012). "An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation." RNA.)*

Genome-wide RNAi screen for *Drosophila* snRNA 3' end Processing Factors

We previously developed a cell-based reporter that expresses green fluorescent protein (GFP) in response to misprocessing of U7 snRNA, allowing for non-invasive and sensitive detection of misprocessing *in vivo* (Figure 3.1A; (Ezzeddine et al. 2011)). To determine whether this reporter would be suitable for use in a genome-wide RNAi screen, we treated S2 cells with dsRNA targeting either the Polypyrimidine Tract Binding Protein (PTB) (negative control) or IntS12 (positive control) and in both instances found no overt effects on growth. Three days following the dsRNA treatment, cells were transfected with the U7-GFP reporter and fluorescence was assessed 48 hrs later. Robust GFP expression was clearly observed after IntS12 depletion, relative to a low level of background fluorescence in cells treated with PTB dsRNA (Figure 3.1A). Western blot analysis confirmed that only after significant reduction in IntS12 expression was GFP expression observed (Figure 3.1A).

Using the U7-GFP reporter we performed a genome-wide RNAi screen in *Drosophila* S2 cells using a library consisting of 15,881 unique dsRNAs targeting >90% of the annotated fly genome. U7-GFP reporter expression in cells grown in 96 well plates was imaged via automated high throughput microscopy and plotted as a function of fluorescence variance. We observed that in ~7% of the wells the dsRNA targeted essential genes, resulting in no viable cells at the time of image acquisition. The imaging of nearly all plates of the dsRNA screen revealed a relatively homogenous level of background fluorescence in practically all wells. Representative graphs of two plates are shown in Figure 3.1B. A small number of wells displayed strong GFP fluorescence, for example on plate 25, cells treated with dsRNA targeting CG5859 identified the fly orthologue of IntS8 as a 'strong hit'. In addition, a modest number of moderately fluorescent wells were also identified, as evidenced by the GFP signal and cell images presented for plate 54 (Figure 3.1B). The levels of GFP expression after knockdown of either CG10572 or CG10583 were clearly above plate background, yet were below the signal of the positive control dsRNA (targeting *Drosophila* IntS9).

In total, our RNAi screen identified 89 genes that, when subjected to RNAi knockdown, resulted in detectable levels of GFP expression from the U7-GFP reporter (Figure 3.1C and Table 3.1). The screen identified ten of the twelve annotated *Drosophila* Integrator subunits, as well as genes involved in RNA binding, chromatin/DNA function, and UTP biosynthesis. To determine the reproducibility of the data, we regenerated dsRNAs corresponding to each of the 89 target genes identified in the initial screen and performed iterative RNAi followed by transfection of the U7-GFP reporter. From this, the initial list of 89 genes was reduced down to 21 that either scored as high or modest with regards to their GFP expression (Figure 3.1D). Ten of these were Integrator genes that encode proteins previously found in the twelve-membered Integrator complex. Since only IntS1 and IntS11 were functionally validated in the initial study that is required for U1 and U2 snRNA 3' end processing, here we comprehensively investigated the importance of each Integrator subunit in U7 and spliceosomal snRNAs 3' end formation. Importantly, eleven genes scored in our functional screen were not previously known to be involved in snRNA biogenesis. As depletion of four of these genes (CG4785, Asu, cyclin C, and Cdk8) produced high levels of GFP expression comparable to that observed after depletion of known Integrator subunits, we then focused the remainder of our analysis on these genes.

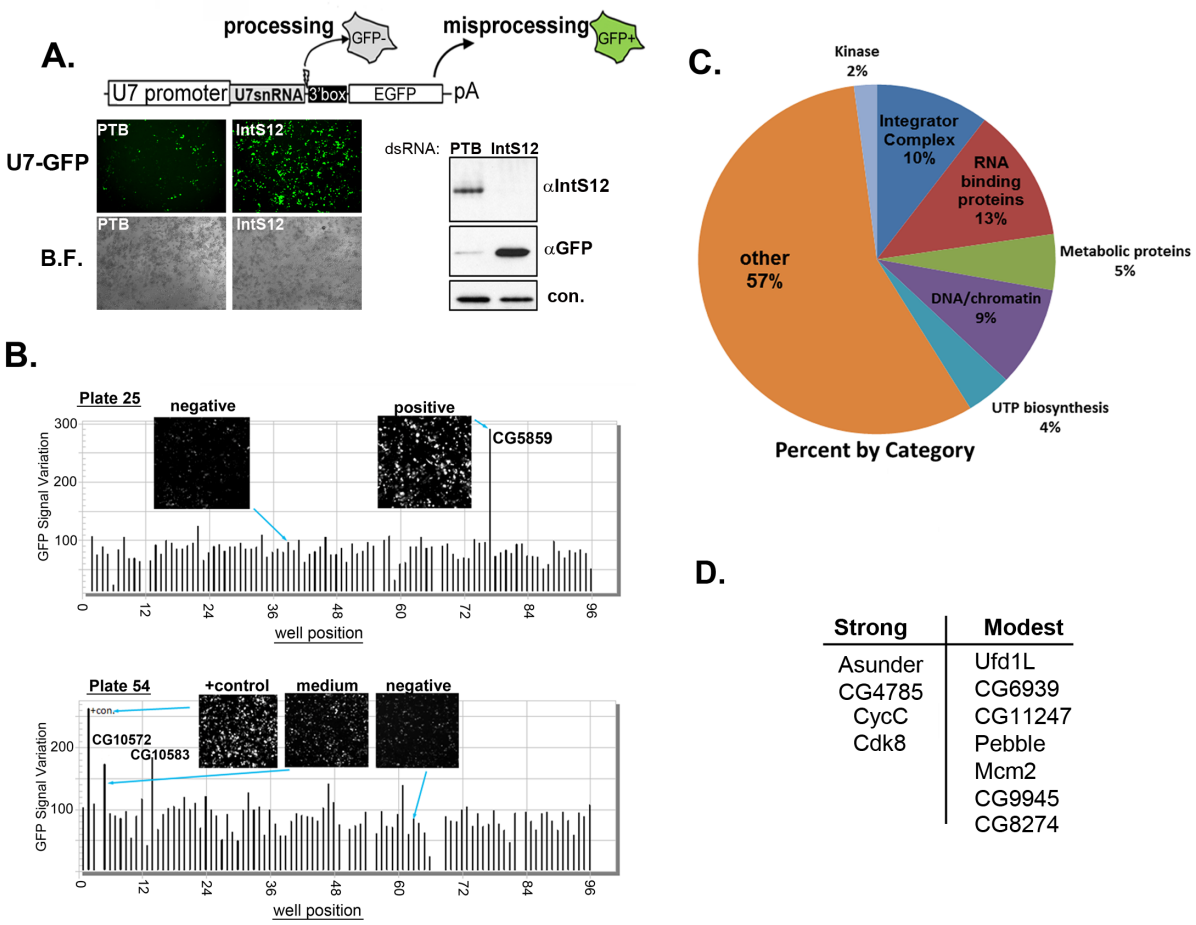


Figure 3.1 Using the U7-GFP reporter to conduct a genome-wide RNAi screen. (**A**). Schematic of U7-GFP reporter and the results from transient transfection of the reporter into S2 cells treated with dsRNA targeting PTB (-control) or IntS12 (+control). Fluorescence and brightfield images are on the left and Western blot analysis on the right. (**B**). Bar graph representing quantification of screen results for plate 25 and 54. The figure insets are taken from the acquired image collection. (**C**). Pie graph representing results of RNAi screen categorically. (**D**). Trimmed down list of non-Integrator genes that reproducibly scored as strong or modest by secondary screening.

Table 3.1 Genes required for snRNA 3' end processing identified from functional RNAi screen			
Gene number	**Fly gene name**	**Human orthologue**	**Function/domains**
CG3173	Integrator 1	Integrator 1	snRNA 3'-end processing
CG8211	Integrator 2	Integrator 2	snRNA 3'-end processing
CG12113	Integrator 4	Integrator 4	snRNA 3'-end processing
CG9591	Integrator 5	Integrator 5	snRNA 3'-end processing
CG3125	Integrator 6	Integrator 6	snRNA 3'-end processing
CG18176	Integrator 7	Integrator 7	snRNA 3'-end processing
CG5859	Integrator 8	Integrator 8	snRNA 3'-end processing
CG5222	Integrator 9	Integrator 9	snRNA 3'-end processing
CG1972	Integrator 11	Integrator 11	snRNA 3'-end processing
CG5491	Integrator 12	Integrator 12	snRNA 3'-end processing
CG6233	UFD1-like	UFD1L	ubiquitin-specific protease activity
CG3593	Rudimentary like	UMPS	UMP biosynthetic process
CG1743	Gs2	GLUL	glutamine synthesis
CG42797	Novel	HECW1	(HECT, C2 and WW domain) protein ubiquitination
CG3307	pr-Set7	SETD8	chromatin modification
CG3931	Rrp4	EXOSC2	deadenylation-dependent decay
CG4482	mol	DUOXA1	protein transport
CG4467	Novel	TRHDE	metallopeptidase
CG6437	GLcT-1	UGCG	Glycosyl transferase
CG6721	Gap1	RASA2	Ras GTPase activator
CG7281	CycC	CCNC	Cyclin
CG34401	novel	KIAA0913	zinc finger protein
CG7610	ATPsyn-gamma	ATP5C1	ATP biosynthetic process
CG7487	RecQ4	NA	ATP-dependent DNA helicase
CG7940	Arp5	ACTR5	DNA repair
CG8114	pebble	ECT2	guanyl-nucleotide exchange factor
CG8274	Megator	TPR	protein import into nucleus
CG8344	RpIII128	POLR3B	RNA polymerase III
CG34389	crossveinless c	STARD13	GTPase activator
CG9428	ZIP1	SLC39A1	metal ion transport
CG10572	Cyclin-dependent kinase 8	CDK8	cyclin-dependent protein kinase
CG10583	Separase	ESPL1	peptidase activity
CG13345	tumbleweed	RACGAP1	Rac GTPase activator
CG18525	Serine protease inhibitor 5	SERPINC1	erine-type endopeptidase inhibitor
CG4260	α-Adaptin	AP2A2	Adaptor protein complex AP-2
CG10911	Novel	NA	NA
CG30184	Novel	NA	phospholipase A2 activity
CG7608	Ecdysone-induced gene 71Ec	NA	defense response to bacterium
CG3181	thymidylate synthase	TYMS	dTMP biosynthetic process
CG42670	pasilla	NOVA1	RNA splicing
CG10134	beat-Va	NA	NA
CG5614	Novel (LisH dimerisation motif)	NA	NA
CG11504	Novel (MADF domain)	NA	NA

Table 3.1 (continued)

Gene number	Fly gene name	Human orthologue	Function/domains
CG43374	Cht6	CHI3L2	Chitinase II
CG1030	Sex combs reduced	HOXB5	sequence-specific DNA binding transcription factor
CG1258	pavarotti	KIF23	microtubule motor
CG3924	Chip	LDB2	LIM domain binding 2
CG4021	Novel	KHDRBS	KH domain containing, RNA binding protein
CG4079	TBP-associated factor 11	TAF11	transcription coactivator
CG4082	Minichromosome maintenance 5	MCM5	DNA helicase
CG4567	iconoclast	GFM1	Translation elongation factor
CG4785	Novel	C15orf44	NA
CG4798	lethal (2) k01209	UCKL1	uridine kinase
CG34420	Novel	DPEP2	peptidase activity
CG5941	Novel	MCTS1	RNA binding
CG6147	Tsc1	TSC1	kinase binding
CG6196	novel	TRAPPC6B	vesicle-mediated transport
CG6814	asunder	ASUN	spermatogenesis regulator
CG6939	SET domain binding factor	SBF2	SET binding factor
CG7108	DNA polymerase α 50kD	PRIM1	DNA primase
CG7109	microtubule star	PPP2CB	protein serine/threonine phosphatase
CG7538	Minichromosome maintenance 2	MCM2	DNA replication
CG8372	Novel	TMEM222	transmembrane protein
CG8388	novel	NA	DNA binding
CG9504	Ecdysone oxidase	CHDH	choline dehydrogenase activity
CG9667	ISY like splicing factor	ISY1	RNA splicing
CG9741	Dihydroorotate dehydrogenase	DHODH	UMP biosynthetic process
CG9755	pumilio	PUM2	RNA binding
CG9945	Novel (WD40 repeat)	DCAF11	protein ubiquitination
CG9943	Surfeit 1	SURF1	cytochrome-c oxidase activity
CG9951	novel	CCDC22	coiled-coil domain containing 22
CG10399	novel	HMGCL	leucine metabolic process
CG10463	novel	DUS3L	tRNA-dihydrouridine synthase
CG11352	jim	NA	regulation of chromatin silencing
CG11207	fascetto	PRC1	cytokinesis
CG11247	novel	NA	Zinc finger protein
CG11291	Novel	PGP	4-nitrophenylphosphatase activity
CG42568	novel	NA	NA
CG32239	Guanine nucleotide exchange factor GEF64C	NET1	guanyl-nucleotide exchange factor
CG13604	novel (SH3 domain)	UBASH3	NA
CG14325	novel (Leucine-rich repeat)	TCTE1	carboxy-lyase activity
CG14105	Novel (Tetratricopeptide repeat)	TTC36	NA
CG10395	Novel	INO80B	DNA repair
CG32845	Novel	PIN1	isomerase activity

Table 3.1 (continued)			
Gene number	**Fly gene name**	**Human orthologue**	**Function/domains**
CG14339	Novel (Tetratricopeptide repeat)	NA	NA
CG32198	Novel	NA	NA
CG14024	Novel	CHST	carbohydrate biosynthetic process
CG10342	neuropeptide F	NA	neuropeptide hormone
CG7837	Novel (Armadillo and BTB domain)	ARMC5	ATP hydrolysis coupled proton transport

Table 3.1 Positive hits from RNAi screen. Color coding describes relative intensity of GFP signal to background within the plate. Red font denotes the strongest scoring hits, blue are medium, and purple are the weakest scoring hits. Column one denotes the CG number present in flybase (www.flybase.net). Column two gives the fly gene name while column three is the putative human orthologue according to ensemble (www.ensembl.org). The fourth column is a short description of known function or putative domains in the absence of known function. Novel means that no characterization of the gene has been found and “NA” means that no orthologue or function is known at this time.

Knockdown of different Integrator components results in differential 3' end processing defects of U7 reporter and endogenous spliceosomal snRNAs.

To systematically assess the importance of individual Integrator subunit for the requirement in snRNA biosynthesis, we measured snRNA misprocessing by using the U7-GFP readthrough reporter in S2 cells depleted of each corresponding Integrator subunit. To accomplish this, we generated dsRNAs targeting all of the known *Drosophila* Integrator subunits using the method described in detail in Chapter 2 (Cell culture and RNAi Section). The dsRNA targeting PTB was used as a negative control. S2 cells were treated with dsRNA for two days before transfection with the U7-GFP reporter, and we observed no overt effect on cell growth in response to the dsRNAs treatment. To assess the knockdown efficiency of endogenous Integrator proteins in S2 cells, antibodies to *Drosophila* Integrator subunits IntS1, IntS9, IntS11 and IntS12 were successfully generated by injection of Guinea pigs with recombinant GST-fusion proteins. Western blot analysis demonstrated that dsRNA targeting each endogenous Integrator mRNA tested significantly reduced the corresponding Integrator protein levels in cells (Figure 3.2A). U7-GFP reporter expression was measured in these cells depleted of each Integrator subunit by both fluorescence microscopy (Figure 3.2B, D) and Western blot analysis (Figure 3.2C, E). Only background GFP fluorescence was observed in control dsRNA (PTB) treated cells while knockdown of nearly every *Drosophila* Integrator subunit gave rise to increased levels of GFP expression from the U7-GFP reporter with the highest level of GFP expression observed from IntS9 depleted cells. However, knockdown of IntS3 and IntS10 only gave rise to low levels of GFP expression equivalent to that observed in control dsRNA treated cells.

To further determine the effect of knockdown of each Integrator subunit on the 3' end processing of endogenous spliceosomal snRNAs, a sensitive quantitative real-time PCR method was developed to measure the accumulation of endogenous premature snRNA species in Integrator knockdown cells. To accomplish this, each snRNA amplicon was designed with a forward primer priming to sequences within the snRNA coding region and reverse primer recognizing ~50 nucleotides downstream of the 3' cleavage site. Therefore, by design, only premature or misprocessed snRNA transcripts would be specifically amplified and detected. PCR primer quality was verified by agarose gel electrophoresis followed by ethidium staining as well as an analysis of the dissociation curves. All amplicons identities were confirmed by sequencing. S2 cells were treated with dsRNA targeting each Integrator subunit or PTB (negative control) for two days, and on the fourth day, total RNA was isolated and subject to reverse transcription (RT). The levels of premature snRNAs were then determined by qRT-PCR analysis using primers specific for premature snRNAs, and data analysis was carried out by using the ΔΔCt method using Rps17 as the reference gene and PTB dsRNA

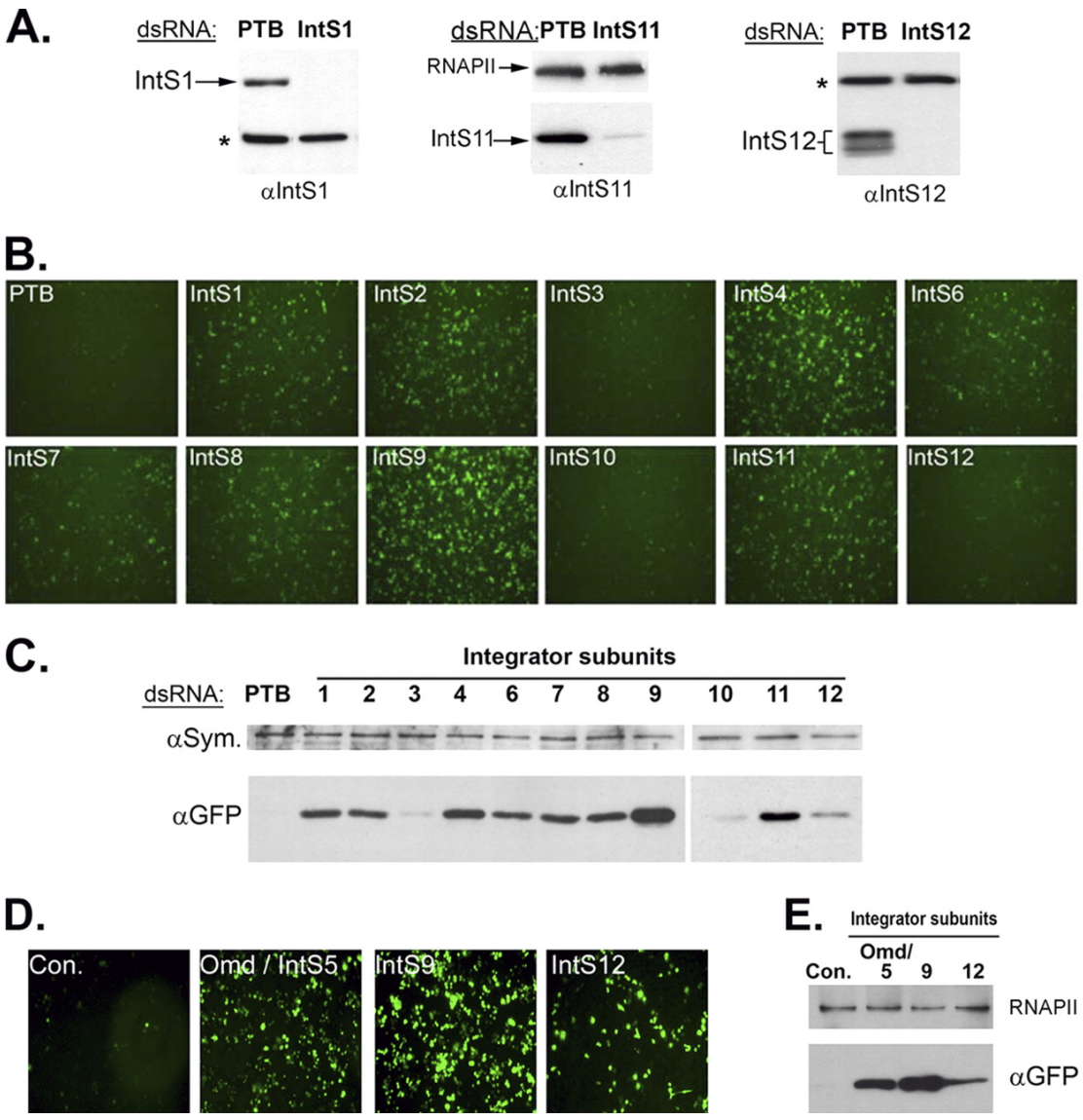


Figure 3.2 Knockdown of the Drosophila Integrator subunits causes misprocessing of the U7-GFP reporter.

(Figure reprinted from Ezzeddine N, Chen J et al. Mol. Cell. Biol. 2011; 31:328-341)

(A) Western blot analysis of lysates from *Drosophila* S2 cells treated with either control dsRNA targeting PTB or a specific dsRNA targeting Integrator 1, 11, or 12. RNAPII (center) or a cross-reacting band (*) serves as a loading control. (**B**) Fluorescence images of cells treated with a specific dsRNA (labeled in white), then transfected with the U7-GFP reporter. (**C**) Western blot analysis of cell lysates from the cells shown in panel B. Lysates were probed with either anti-GFP antibodies or with anti-Symplekin (αSym.) antibodies as a control. (**D**) Fluorescence images of cells treated with dsRNA targeting IntS5, 9 and 12 followed by transfection with the U7-GFP reporter. In this case the control dsRNA targeted LacZ and not PTB. (**E**) Western blot analysis of cell lysates from panel D.

treated cells as a control. Results (Figure 3.3A) were plotted as fold changes in the level of premature snRNA in Integrator knockdown cells versus control treated cells and the average presented was from three independent RNAi experiments. With the exception of Integrator 3 dsRNA treatment, all knockdowns led to increase in the levels of premature spliceosomal snRNAs (Figure 3.3A). Notably, knockdown of Integrator 10 gave rise to the smallest increase, consistent with our observations from the U7-GFP reporter assay (Fig. 3.2B). For each knockdown, data garnered from our qPCR analysis were very consistent with those from our reporter assay where knockdown of Integrators 1, 4, and 9 led to a over 50-fold increase in the levels of premature snRNAs, and knockdowns of Integrators 11 and 7 resulted in a ~10 fold increase. Except for IntS3 and IntS10, knockdown of the remainder of the Integrator subunits led to a 2-10 fold increase in the levels of premature snRNAs. We measured Integrator protein knockdown efficiency in these cells by using antibodies for IntS1, 9, 11 and 12, and measured IntS3 and IntS10 mRNA levels in knockdown cells by qRT-PCR analysis using PTB and IntS9 dsRNAs treated cells as negative and positive controls respectively (Figure 3.3B). We observed a 65%, 75% and ~90% reduction in the levels of IntS3, IntS9 and IntS10 mRNAs respectively, and these results suggest that dsRNAs targeting IntS3 or IntS10 was effective in depleting endogenous Integrator mRNA levels and it is likely that IntS3 and IntS10 are not required for efficient snRNA 3' end formation. Taken together, functional analysis indicates that with the exception of IntS3 and IntS10, all known Integrators are functionally required for snRNA 3' end processing.

Asunder and CG4785 are specifically required for snRNA 3' end formation.

In addition to the known Integrator proteins, our genome-wide RNAi screen also identified four non-Integrator proteins: Asunder, CG4785, cyclin C, and Cdk8.

To confirm that the observed snRNA misprocessing occurs as a consequence of knockdown of CG4785 and Asu, we designed a second set of dsRNA targeting a distinct region of each open reading frame. S2 cells were then treated with dsRNA targeting LacZ or the Stem Loop Binding Protein (SLBP) as negative controls, IntS9 and IntS12 as positive controls, and either dsRNA targeting Asu or CG4785. We observed only background levels of GFP expression in negative control dsRNA treated cells while both dsRNA targeting Asu and CG4785 gave rise to robust and nearly identical levels of GFP expression comparable to those observed after depletion of IntS9 or IntS12 (Figure 3.4A and B). To determine the specificity of the misprocessing phenotype, we utilized two related GFP reporters that we developed previously to determine defects in histone pre-mRNA processing (Wagner et al. 2007; Yang et al. 2009; Ezzeddine et al. 2011). The H3-GFP report is almost identical to the U7-GFP construct, except the histone H3 gene is used. In addition, the Act-H3-GFP reporter replaces the H3 promoter with the actin promoter, which was found to sensitize it to additional factors whose depletion normally extinguishes the histone promoter (Yang et al. 2009). We treated cells with the same series of positive control, negative control,

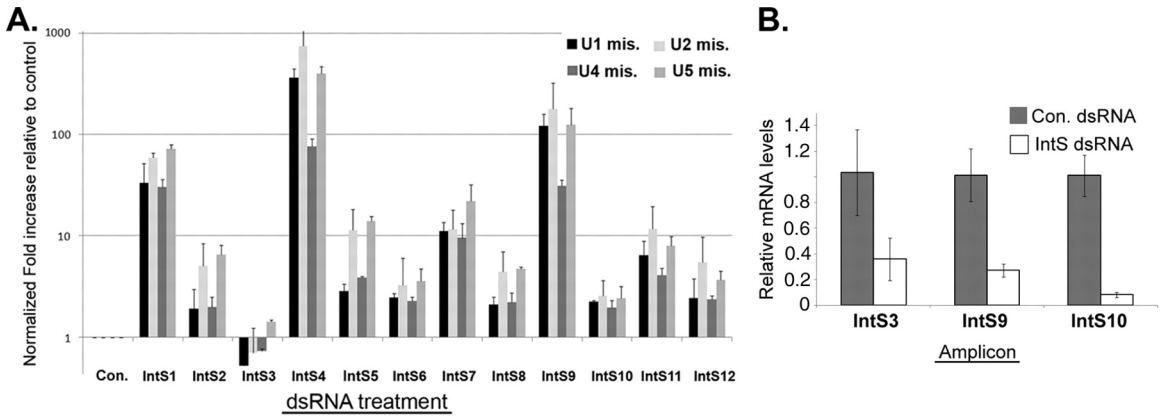


Figure 3.3 Knockdown of Integrator subunits causes various degrees of misprocessing of endogenous spliceosomal snRNAs.

(Figure reprinted from Ezzeddine N, Chen J et al. Mol. Cell. Biol. 2011; 31:328-341)

(**A**) Histogram of real-time PCR experimental data generated using primer pairs designed to detect the presence of misprocessed (mis.) spliceosomal U1, U2, U4, or U5 snRNAs. Results are plotted as fold increases relative to control-treated cells and reflect expression normalized to an internal control gene (RpS17). All results are derived from biological triplicates, with error bars indicative of the standard deviations of the triplicate quantification. (**B**) Histogram of qRT-PCR quantitation of IntS3, IntS9, and IntS10 mRNA following dsRNA treatment. Levels represented are the averages of triplicate experiments normalized to an internal control (RpS17) and then normalized to control-treated cells.

and test dsRNA as described above followed by the transfection with either H3-GFP or Act-H3-GFP reporter. Using these other two constructs, we only observed GFP expression after depletion of SLBP, confirming that both Asu andCG4785 are specifically required for snRNA 3' end processing. To characterize the role of Asunder and CG4785 in the processing of endogenous snRNA, we used a sensitive qRT-PCR assay designed to measure misprocessed U1 and U5 snRNA following RNAi depletion. Total cell RNA was isolated following several days of incubation with dsRNA targeting LacZ (negative control), IntS9 or IntS12 (positive controls) or dsRNAs targeting Asu and CG4785 (two separate dsRNAs each gene). We subjected total RNA to reverse transcription followed by real-time PCR analysis using primers specific for the misprocessed U1 and U5 snRNA (Ezzeddine et al. 2011). Relative to control treated cells, we observed a 5-10 fold increase in the level of misprocessed snRNA (Figure 3.4C). The levels of misprocessed snRNA were consistent between the two individual dsRNAs for each target and comparable to that observed after depletion of IntS12. Taken together, these results demonstrate that Asu and CG4785 are required for both reporter and endogenous snRNA 3' end formation.

Cyclin C and Cdk8 are required for snRNA 3' end formation.

The observation that both Cdk8 and CycC were independently isolated in our screen strongly suggests that they are involved in snRNA 3' end formation, as they form a cognate cyclin/cdk pair (Leclerc et al. 1996).To address the possibility of any potential off-target effects and characterize their specificity towards snRNA 3' end formation we created a second set of dsRNA targeting distinct regions of each ORF. Treatment of S2 cells with dsRNAs targeting Cdk8 or targeting CycC resulted in robust expression of GFP from the U7-GFP reporter (Figure 3.4D). We also tested the effect of CycC or Cdk8 depletion on both the histone H3-GFP and Actin-H3-GFP reporters. Neither the H3 nor Actin H3 reporters were found to produce GFP in response to Cdk8 or CycC knockdown. In contrast, we observed robust expression of GFP from both reporters in response to treating cells with dsRNA targeting SLBP. These data implicate that CycC and Cdk8 are specifically involved in the process of snRNA 3' end formation.

To test whether the kinase activity of Cdk8 is required for snRNA 3' end formation, we cloned wild-type *Drosophila* Cdk8 and generated a kinase-dead (D173A) form analogous to a mutant human CDK8 shown previously to behave as a dominant-negative (DN) (Akoulitchev et al. 2000). Both wild-type and DN cDNAs were expressed as N-terminal myc-tagged proteins in S2 cells co-transfected with the U7-GFP reporter. Despite the reduced accumulation of the catalytically inactive Cdk8 relative to the wild-type protein, we observed potent misprocessing of U7-GFP reporter (Figure 3.4E). These data demonstrate that the kinase activity of the CycC/Cdk8 complex is essential for correct snRNA 3' end processing.

One possible explanation for the observed phenotype is that CycC/Cdk8 fulfills a general requirement in the expression of one or more Integrator subunits and therefore CycC/Cdk8 plays an indirect role in snRNA 3' end formation. To test this, we analyzed the expression of all twelve Integrator subunits as well as Asunder and CG4785 after depletion of CycC or Cdk8 using both Western blot analysis (Figure 3.5A) and qRT-PCR (Figure 3.5B/C). We found no observable differences in the levels of Integrator subunit expression following CycC or Cdk8 knockdown. Therefore, we conclude that these two proteins are likely functioning directly in the 3' end formation of *Drosophila* snRNA.

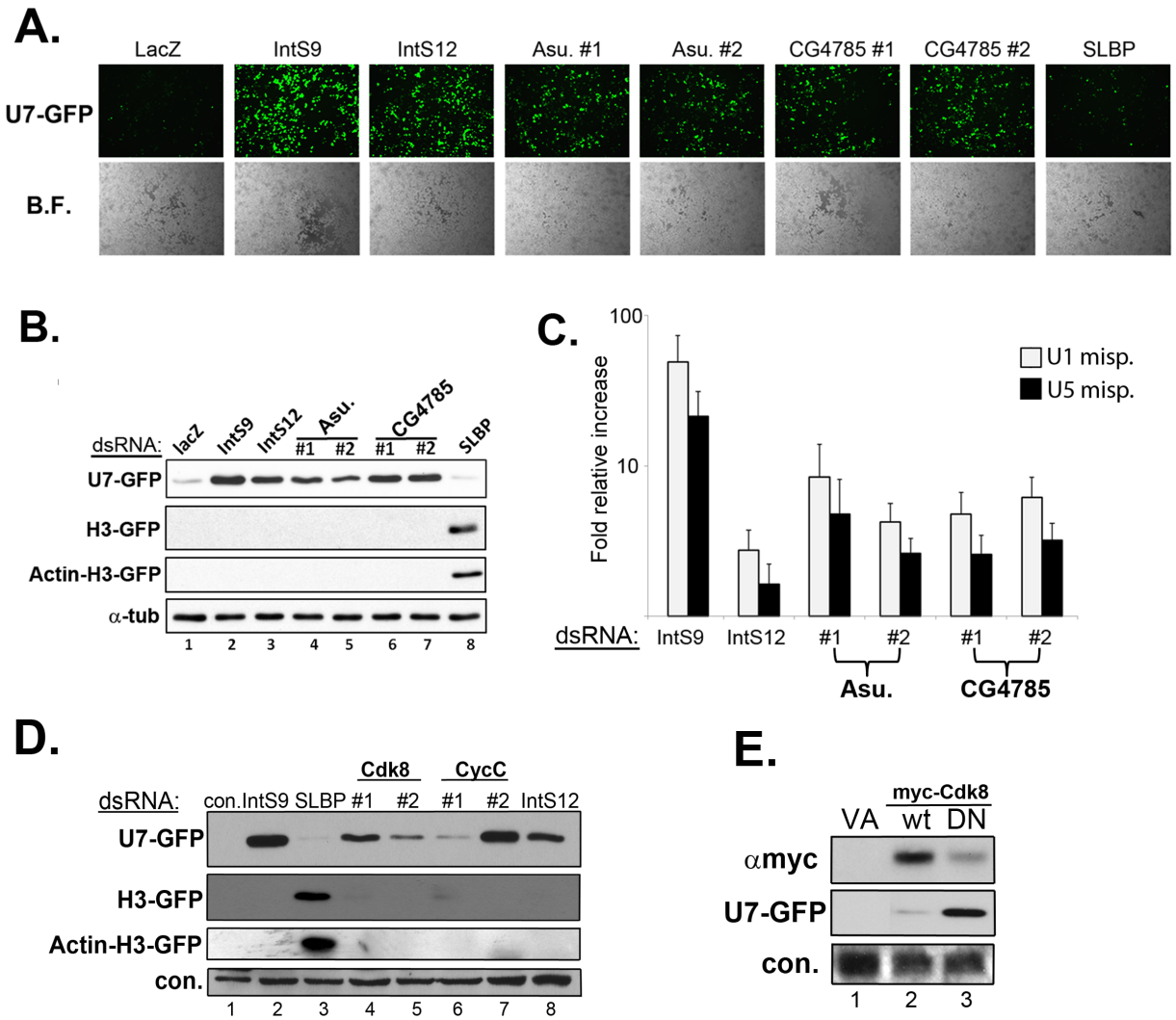


Figure 3.4 Validation of Asunder, CG4785, CycC, and Cdk8 as required for snRNA 3' end formation. (Figure reprinted from Chen J et al. RNA. 2012. doi:10.1261/rna.035725.112) (**A**). Images from S2 cells transfected with U7-GFP reporter after dsRNA treatment targeting Asu or CG4785. (**B**). Western blot analysis of cell lysates from Panel A. (**C**). qRT-PCR analysis specific for misprocessed endogenous snRNA from knockdown cells. (**D**). Western blot analysis of S2 cells treated with dsRNA targeting CycC or Cdk8 followed by transient transfection with reporters measuring snRNA or histone mRNA 3' end formation. (**E**). Western blot analysis of cells cotransfected withU7-GFP reporter and myc-tagged Cdk8 that is wild type or catalytically inactive.

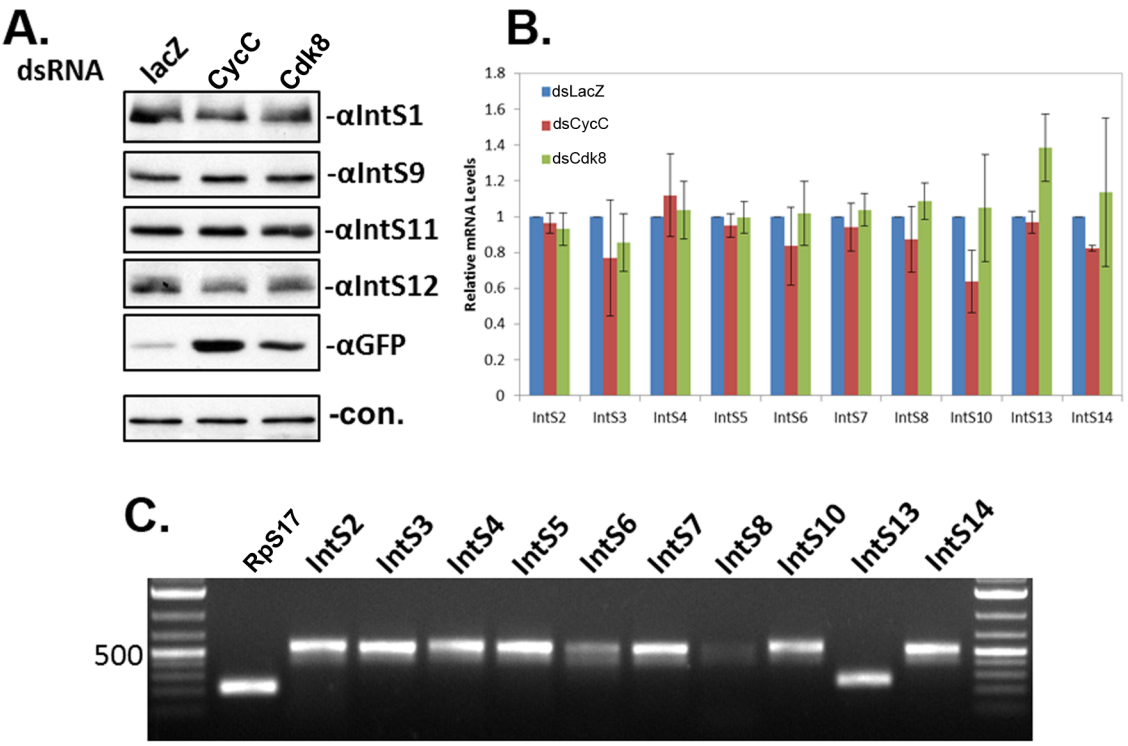


Figure 3.5 Functional involvement of Cyclin C and Cdk8 in snRNA 3' end formation is not through regulation of Integrator expression.

(Figure reprinted from Chen J et al. RNA. 2012. doi:10.1261/rna.035725.112)

(**A**). Western blot analysis of protein levels of Integrator subunits after treating cells with dsRNA targeting LacZ, CycC, or Cdk8. Cells were also transfected with the U7-GFP reporter to monitor snRNA misprocessing as a functional readout of knockdown of CycC/Cdk8. (**B**). Quantitative RT-PCR analysis using SYBR green staining of PCR products. RNA was isolated from S2 cells, treated with the three dsRNAs described in panel A and subjected to real-time PCR analysis using amplicons specific to Integrator subunits not tested by Western blot analysis in panel A. All measurements were normalized to RpS17 signals as a housekeeping internal control. (**C**). Ethidium bromide staining of agarose gel electrophoresis of the amplicons amplified in panel B.

Asunder and CG4785 associate with Integrator Subunits.

To determine if either CG4785 or Asunder biochemically associate with the other Integrator subunits, we cloned full-length cDNAs and generated stable S2 cell lines expressing either FLAG-tagged Asunder or FLAG-tagged CG4785. Western blot analysis of lysates from these cell lines using anti-FLAG antibodies demonstrates the expression of specific bands of the predicted molecular weight for both proteins (Figure 3.6A). We analyzed the subcellular localization of both of the tagged proteins and observed a nuclear localization of CG4785 and a bimodal localization of Asu in both the nucleus and the cytoplasm (Figure 3.6B). Nuclear extracts were prepared from both stable cell lines from which epitope tagged proteins were immunoprecipitated using anti-FLAG agarose. A large majority of FLAG-tagged protein was recovered in each precipitate from both cell lines and we determined the levels of interacting endogenous Integrator subunits using antibodies to *Drosophila* IntS1, IntS9, IntS11, and IntS12 proteins (Figure 3.6C). A large proportion of all four Integrator subunits were detected specifically associating with Asu and CG4785, relative to control pull downs (Figure 3.6C). We did not detect SLBP in either immunoprecipitate, consistent with the observation that depletion of these two proteins does not affect histone mRNA 3' end formation. To confirm potential direct interactions with *Drosophila* Integrator subunits, we performed binary interaction tests using pairwise, directed yeast two-hybrid analysis. This is an ideal assay to perform this analysis given the lack of endogenous yeast Integrator subunits. Under low and high stringency conditions, we did not observe any interaction between Asu and the known Integrator subunits (not shown). In contrast, we did observe a robust interaction between CG4785 and IntS10 by Y2H (Figure 3.6D). Collectively, these data demonstrate that, in addition to their functional requirement in snRNA 3' end formation, both Asu and CG4785 biochemically associate with the previously described members of the *Drosophila* Integrator complex (Baillat et al. 2005). While the interaction of CG4785 is most likely mediated through IntS10, the binding partner(s) of Asunder remain to be identified.

Requirement for Cyclin C and Cdk8 in snRNA 3' end formation is conserved and independent of its function in Mediator.

There has been extensive previous characterization of CycC and CDK8 as members of the Mediator CDK8 module in conjunction with Mediator 12 and 13 (reviewed in (Galbraith et al. 2010)). Knockdown of these two Mediator subunits did not produce GFP expression in our screen, suggesting that either they are not required for snRNA processing or were not depleted effectively. To address the latter possibility, we created independent dsRNAs specific to Med12 and Med13 and observed that they elicited significant reduction in the levels of endogenous Med12 and Med13 yet failed to trigger misprocessing in the U7-GFP reporter (Figure 3.7A). These results imply that CycC and Cdk8 are required for snRNA 3' end formation independent of Med12 and Med13. To determine if CycC or CDK8 play a conserved role in the 3' end formation of human snRNA, we treated HeLa cells with two siRNAs targeting both transcripts and then transfected the human version of the U7-GFP reporter that we showed previously to express GFP in response to Integrator activity disruption (Albrecht and Wagner 2012). We observed that both siRNAs resulted in efficient reduction of the levels of endogenous CycC and siRNA #2 resulted in significant depletion of CDK8 and gave rise to GFP expression comparable to those after depletion of IntS12 (Figure 3.7B). We noted that CDK8 siRNA#1 did not give rise to as much GFP expression as siRNA#2 but the knockdown was not as effective and also not as effective as co-depleting CycC. These results demonstrate the requirement for CycC and CDK8 in snRNA 3' end formation in metazoans.

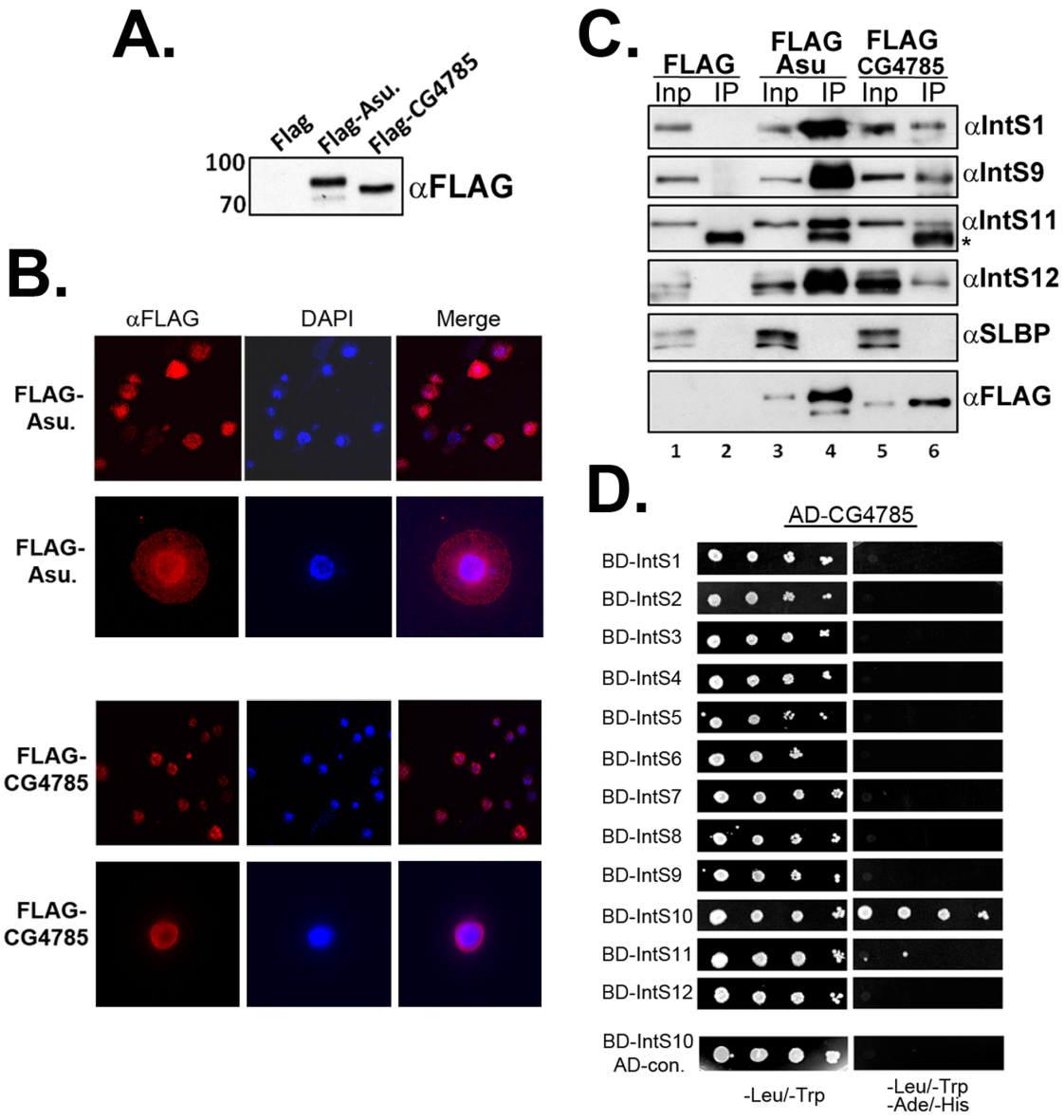


Figure 3.6 Asunder and CG4785 biochemically associate with the fly Integrator Complex.

(Figure reprinted from Chen J et al. RNA. 2012. doi:10.1261/rna.035725.112)

(**A**). Western blot analysis of FLAG-tagged Asunder and CG4785 expression in S2 stable cells. (**B**). IF analysis of stable S2 cell lines expressing FLAG- tagged Asunder or CG4785. (**C**). Western blot analysis of FLAG immunoprecipitates from cell lines described in panel A. Input lanes represent 5% of input and IP represents 50% of the immunoprecipitate. (**D**). Directed yeast two-hybrid using fly Integrator subunits and CG4785 demonstrating an interaction with IntS10.

The functional requirement for CycC/Cdk8 in fly snRNA 3' end formation generates the question of whether these two proteins are associated with Integrator subunits. To test this possibility, we generated stable S2 cell lines expressing FLAG-mcherry (negative control), FLAG-Cdk8, or FLAG-CycC and performed immunoprecipitations from nuclear extracts prepared from these cell lines. We observed substantial amounts of endogenous Med12 and Med13 associating with both FLAG-CycC and FLAG-Cdk8 as predicted, yet we also observed a reproducible interaction with the Integrator subunits (Figure 3.7C). The reduced levels of Integrator proteins in the co-IP suggest that the interaction is relatively weak or transient and that only a small pool of CycC/Cdk8 not associated with Med12 and Med13 binds to the Integrator subunits. To test these predictions, we used IntS12 antibodies to immunoprecipitate the Integrator complex from nuclear extract derived from S2 cells stably expressing FLAG-Cdk8. We observed significant amount of IntS11 and a portion of FLAG-Cdk8 associating with IntS12 (Figure 3.7D). Importantly, we did not detect Med13 interacting with IntS12. This reciprocal pulldown confirms the presence of a small amount of Cdk8 associating with the fly Integrator complex in the absence of Mediator.

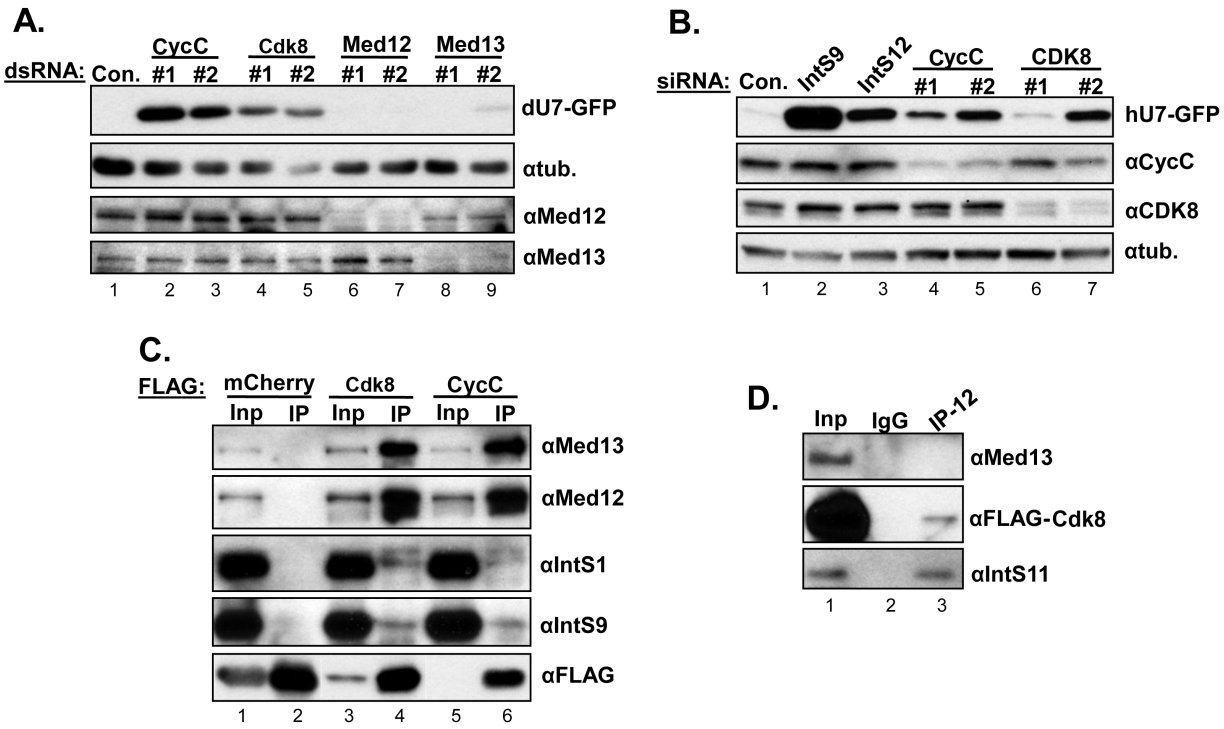


Figure 3.7 CycC/Cdk8 function in snRNA 3' end formation independent of Mediators 12/13 and are associated with Integrator subunits.

(Figure reprinted from Chen J et al. RNA. 2012. doi:10.1261/rna.035725.112)

(**A**). Western blot analysis of lysates from S2 cells treated with dsRNA targeting members of the *Drosophila* Mediator Cdk8 module and transfected with the U7-GFP reporter. (**B**). Western blot analysis of lysates from HeLa cells treated with siRNA to CycC or CDK8 followed by transfection with a human version of the U7-GFP reporter. (**C**). Immunoprecipitations using αFLAG-agarose to detect interactions between CycC/Cdk8 and Integrator 1 and 9. Input lanes represent 2% of input and IP lanes represent 50% of immunoprecipitate. (**D**). Western blot analysis of immunoprecipitations using anti-IntS12 antibodies.

DISCUSSION

The functional genome-wide RNAi screen presented here has defined the *Drosophila* Integrator complex to comprise of fourteen core subunits. It is likely that the core members of the Integrator complex are conserved as the human orthologues of Asu and CG4785 were reported to be present in the human Integrator immunoprecipitations with the levels comparable to the known twelve Integrator members (Malovannaya et al. 2010). The conserved requirement for CycC/Cdk8 in the formation of both fly and human snRNA 3' end suggests that besides the conserved core complex, the regulation of snRNA 3' end formation is likely retained in metazoan species.

U7-GFP reporter as a sensor for snRNA misprocessing

The snRNA-GFP reporter allows us to visually monitor snRNA 3' end formation with great sensitivity. However, one limitation of the U7-GFP reporter is that for an RNAi-knockdown to be scored as a positive "hit", the protein depleted must only disrupt snRNA 3' end formation but not snRNA gene transcription. In this case, any factor that extinguishes snRNA transcription will be eliminated. This likely explains why members of the snRNA Activating Protein Complex (SNAPc) and the recently defined Little Elongation Complex were not scored in our RNAi screen (Hung and Stumph 2011; Smith et al. 2011). Depletion of any of these factors would compromise U7-GFP transcription and lead to no GFP expression. Developing an snRNA reporter that is able to visualize both snRNA transcription and 3' end formation will significantly improve our understanding to the screen results by discriminating the effect resulted from transcription from 3' end formation. Recently, RNA aptamers have been reported to lighten up modified RNAs in cells by forming the RNA aptamer/fluorophore complex, which produce a wide spectrum of fluorescence(Paige et al. 2011). A modified snRNA reporter that fuses these newly developed RNA aptamers to the snRNA genes will enable us to monitor snRNA transcription and 3' end formation simultaneously by measuring the different spectrum of fluorescence, such as GFP and mCherry. Interestingly, the fly orthologue of RPAP2 (CG34183), a protein factor recently proposed to mediate the interaction between Integrator complex and the RNAPII CTD (Egloff et al. 2012), was not scored in our screen either. One possible explanation of this is that RPAP2 may be required for snRNA gene transcription as well as 3' end processing. Alternatively it may not be required in *Drosophila*. Considering the markedly distinct structure of fly RNAP II CTD (very few Ser7 within the heptad repeat), and the significantly smaller fly RPAP2 (143a.a.) versus its human counterpart (612a.a.), it is possible that RPAP2 may not be involved in *Drosophila* snRNA 3' end formation.

Novel components identified to be required for snRNA 3' end processing

Our functional screen identified Asu and CG4785 as two additional core components of the Integrator complex. Asu was initially named Mat89b, due to its maternal deposition in the *Drosophila* embryo (Stebbings et al. 1998). Later, it was characterized to be a substrate for PanGu kinase, and the reduction of Asu expression in the spermatocytes was reported to cause aberrant centrosome and spindle assembly (Anderson et al. 2009). Asu has a predicted molecular weight of 76kDa with a conserved yet uncharacterized DUF2151 domain. The CG4785 gene encodes a 65kDa protein with a less conserved (Pfam: E-value 0.022) von Willebrand Factor (vWF) type A (VWA) domain, which was observed in IntS6 as well. Previous studies show that the VWA domains in extracellular eukaryotic proteins are involved in cell surface adhesion via metal ion-dependent adhesion sites (MIDAS), and intracellular VWA domains containing proteins have been found to be involved in various functions such as transcription, DNA repair, ribosomal and membrane transport and the proteasome(Whittaker and Hynes 2002). The human orthologues of Asu and CG4785 have been found to be present in the human Integrator protein immunoprecipitations by large-scale mass spectrometry at the levels comparable to known Integrator proteins, suggesting that these two proteins are likely to play conserved roles in snRNA biogenesis (Malovannaya et al. 2010; Malovannaya et al. 2011). Based on the existing evidence and our results, we suggest that Asu be known as Asu/IntS13 and CG4785 as IntS14.

This study also identified the functional requirement of CycC/Cdk8 in snRNA 3' end formation, and suggests that CycC/Cdk8 are likely to function in a pathway independent of Mediator 12 and 13. Our results are consistent with previous observations that sex comb phenotypes of CycC– and Cdk8– clones are indistinguishable from each other but clearly distinct from the phenotypes of Med12– and Med13– clones (Loncle et al. 2007). Here we propose two potential functions for Cyclin C/Cdk8 in snRNA 3' end formation: phosphorylation of the RNAPII CTD, or as an Integrator subunit kinase. Phosphorylation of Ser2 and Ser7 in the RNAPII CTD is reported to be specifically required for Integrator complex recruitment and activation of snRNA 3' end processing, but the kinase responsible for Ser7 phosphorylation is not known yet. Cdk8 has been shown to phosphorylate the CTD of Rpb1 *in vitro* (Rickert et al. 1999), so it is possible that Cdk8 may act as a RNAPII CTD Ser7 kinase to mediate snRNA 3' end processing. Alternatively, CycC/Cdk8 may regulate an Integrator subunit through direct phosphorylation similar to the role it plays in the mediator complex where CycC/CDK8 phosphorylates Mediator 13 (Knuesel et al. 2009).

In conclusion, the most remarkable feature of the Integrator complex is that an essential network of interactions is present making it intolerant to perturbation. In most cases, depletion of even a single subunit leads to inefficient snRNA 3' end processing. Thus, characterization of

the functional interactions and identification of the role of each Integrator subunit in snRNA 3' end formation represent the most compelling challenges in the field.

Chapter 4. Functional Analysis Identifies an Autonomous IntS12 microdomain Mediating Activation of the Drosophila Integrator Complex.

INTRODUCTION

The accurate production of snRNAs is an important bioprocess needed for efficient downstream RNA processing events including intron removal and histone mRNA 3' end formation (reviewed in (Matera et al. 2007; Egloff et al. 2008)). With the exception of the U6 snRNA, spliceosomal snRNAs are transcribed by RNA Polymerase II (RNAPII) and their 3' ends are processed by the Integrator complex. Both the snRNA promoter and the 3' box sequence element located downstream of the cleavage site have been established as features required for snRNA 3' end formation in metazoans (Hernandez 1985; de Vegvar et al. 1986; Hernandez and Weiner 1986). Unlike the rigid requirements of the poly(A) signal (PAS) in protein coding genes or the histone downstream element (HDE) in histone mRNA processing, the snRNA 3' box can be removed with only minor perturbation to snRNA biosynthesis (Ach and Weiner 1987; Ezzeddine et al. 2011). This demonstrates that the snRNA processing machinery is moderately tolerant to mutations within demarcating *cis* elements and that specificity of the RNA cleavage event is in part brought about through additional means.

Precise snRNA 3' end cleavage is predicated on transcription being initiated from an snRNA promoter, with distinct RNAPII C-terminal domain (CTD) protein modifications found on transcription complexes active at snRNA loci. Phosphorylations within the CTD heptad repeat at serines 2 and 7 have been shown to be essential for Integrator recruitment and subsequent snRNA 3' end formation (Egloff et al. 2007; Egloff et al. 2010). Replacement of native snRNA promoters with RNAPII promoters from protein-coding genes prevents snRNA 3' end formation, consistent with Integrator being assembled onto the RNAPII complex early in the transcription cycle (de Vegvar et al. 1986; Hernandez and Weiner 1986; Ezzeddine et al. 2011).

The initial biochemical identification of the Integrator complex, subsequent analyses of immunoprecipitates, and a recent genome-wide RNAi screen have identified fourteen members of the complex to date (IntS1 through IntS12, Asu/IntS13 and IntS14) (Baillat et al. 2005; Malovannaya et al. 2010; Malovannaya et al. 2011; Chen et al. 2012). Each member of the human Integrator complex is conserved in *Drosophila* and RNAi-mediated depletion of nearly any Integrator subunit in S2 cells causes snRNA misprocessing (Ezzeddine et al. 2011). This latter result suggests the existence of a network of interactions within the Integrator complex that is highly sensitive to disruption.

The only well-established protein-protein interaction among Integrator subunits is between IntS9 and IntS11 (Dominski et al. 2005b; Albrecht and Wagner 2012). These two proteins contain highly conserved metallo-β-lactamase and β-CASP domains and likely represent the catalytic core of the complex (reviewed in (Dominski 2007)). They form a highly stable

heterodimer *in vivo* and their association is mediated through conserved C-terminal domains on both proteins. Formation of this heterodimer is required for snRNA 3' end formation and likely is important to activate the endonuclease activity of IntS11 (Albrecht and Wagner 2012). The role of the remaining subunits in snRNA 3' end formation has yet to be determined and functional domains within other subunits have yet to be experimentally identified.

There are several evolutionarily conserved motifs identifiable within Integrator subunits in addition to the conserved β-CASP/β-lactamase domains of IntS9/11 (reviewed in (Chen and Wagner 2010)). These include the HEAT repeats within IntS4, a von-Willebrand factor type A (VWA) motif within IntS6, and a Plant Homeodomain (PHD) finger in IntS12. PHD fingers are typically ~60 amino acid motifs comprising a C4HC3 signature that coordinate two zinc ions (reviewed in (Bienz 2006; Musselman and Kutateladze 2011)). Proteins containing PHD fingers are almost exclusively found in the nucleus and are commonly present in protein complexes that govern transcriptional regulation. The PHD finger itself robustly interacts with N-terminal tails of histones, most commonly histone H3. Typically, PHD fingers exhibit preference toward unique chemical modifications of amino acids in histone H3 with a particular preference toward lysine methylation. These attributes make analysis of the role of the IntS12 PHD finger in snRNA 3' end formation an attractive entry point to further our understanding of Integrator subunit function.

Here, we investigate the role of *Drosophila* IntS12 in snRNA 3' end formation using snRNA-specific GFP reporters to assess Integrator complex activity. To identify regions of IntS12 required for Integrator activity, we devised an RNAi-rescue strategy to re-express RNAi-resistant forms of *IntS12* mRNA in cells depleted for endogenous IntS12 protein. Unexpectedly, we determined the PHD finger to be dispensable for IntS12 activity and instead identified a small microdomain at the N-terminus of IntS12 that is essential for activity. Surprisingly, the IntS12 microdomain by itself is sufficient to rescue the snRNA misprocessing defected from our reporters as well as endogenous snRNAs observed in IntS12 depleted cells. Moreover, the microdomain is sufficient to mediate interaction between Integrator subunits and a heterologous protein. We also show that the IntS12 microdomain interacts with IntS1 in the absence of other Integrator subunits and the stability of these two subunits requires their interaction. Collectively, these results suggest a critical regulatory function for IntS1 and IntS12 required for Integrator activity in snRNA 3' end formation.

RESULTS

Development of an RNAi-rescue assay to study IntS12 function.

Previously, we developed a GFP-based reporter system for *in vivo* monitoring of U7snRNA 3' end formation in *Drosophila* cells (Ezzeddine et al. 2011; Chen et al. 2012). The advantage of the U7-GFP reporter as a method of monitoring Integrator complex activity is that loss of activity can be readily detected *in vivo* with high sensitivity via GFP fluorescence. A potential limitation of this reporter is that it uses the U7 snRNA gene, which is somewhat atypical from the spliceosomal snRNA genes due to unique variations within its core promoter elements (Dominski et al. 2003; Hernandez et al. 2007). To control for potential experimental bias, we created a second analogous spliceosomal snRNA reporter based instead upon the *Drosophila* U4:39B snRNA gene. To test the functionality of this new reporter (U4-GFP), we treated S2 cells with separate non-overlapping dsRNAs (IntS12#1 or IntS12#2) to induce RNAi-mediated depletion of IntS12. Cells were then transfected with either the U7-GFP or U4-GFP reporter to assess Integrator complex functionality (Figure 4.1A). Depletion of IntS12 resulted in robust levels of GFP expression from both reporters relative to control dsRNA (*LacZ*) treated cells. Western blot analysis of lysates prepared from treated cells confirmed both the loss of endogenous IntS12 expression and a consequential increase in GFP production from both reporters (Figure 4.1B). While the overall sensitivity of the U4-GFP reporter is similar to that of the U7-GFP reporter, the U4-GFP reporter displayed greater specificity, as seen by the lower background GFP fluorescence in the control treated cells. Finally, to confirm that dsRNA-mediated depletion of IntS12 led to misprocessing of endogenous snRNAs, we isolated total RNA and performed qRT-PCR analysis using primers that specifically detect the misprocessed forms of the U1, U2, U4 and U5 snRNAs (Figure 4.1C). Both dsRNAs directed against *IntS12* were found to increase the levels of misprocessed snRNA 2-4 fold, consistent with our previous observations (Ezzeddine et al. 2011).

To validate that the misprocessing of the U7 and U4-GFP reporters we observed is induced by depletion of the endogenous IntS12 protein, we set to rescue the phenotype by introducing back an *IntS12* plasmid refractory to RNAi. As initial attempts to target the endogenous *IntS12* mRNA 5′ or 3′ UTRs using dsRNA failed to generate adequate depletion, we developed a RNAi-resistant *IntS12* expression construct containing 188 silent mutations within the region targeted by the IntS12#2 dsRNA (cDNA bases 1-418) (Figure 4.2A). Expression of this RNAi-resistant *IntS12* cDNA (*IntS12**) from a ubiquitin promoter yielded ~50 kDa doublet of IntS12 protein, albeit at slightly reduced levels relative to the wild-type *IntS12* cDNA expressed under the same conditions. Next we tested the ability of *IntS12** to restore snRNA 3' end processing in cells when endogenous IntS12 has been depleted.

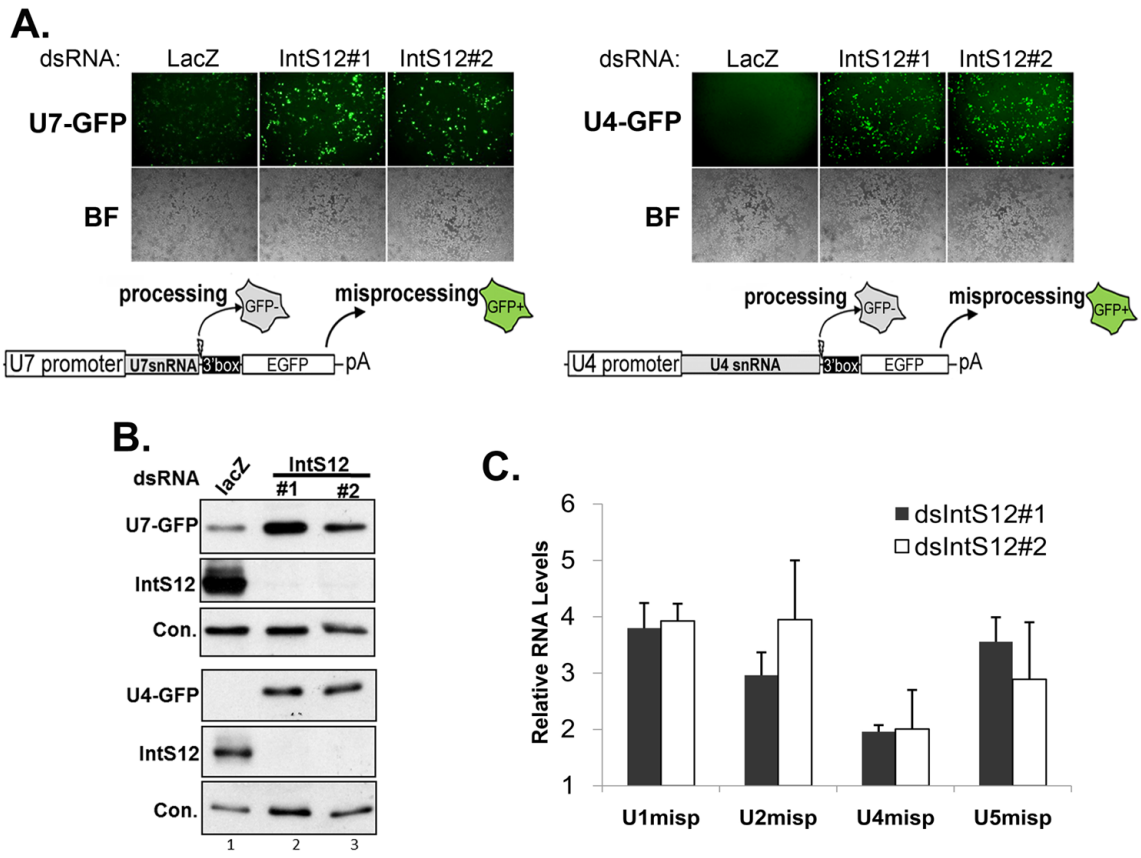


Figure 4.1 Dual GFP reporters reveal snRNA misprocessing following IntS12 knockdown in *Drosophila* S2 cells. S2 cells treated with dsRNA corresponding to bases 418-719 (IntS12#1) and 1-418 (IntS12#2) of IntS12 mRNA, or control dsRNA (lacZ) were subsequently transiently transfected with either U7-GFP or U4-GFP 3' cleavage reporter constructs. (**A**) Fluorescence and brightfield images of treated S2 cells and. (**B**) Western blot analysis of cell lysates using anti-GFP or anti-IntS12 antibodies. A nonspecific band that cross-reacts with the IntS12 antibody is shown as a loading control (**C**) Graphical representation of qRT-PCR quantification of snRNA misprocessing. Amplicons specific for non-processed snRNAs isolated from S2 cells treated with either IntS12#1 or IntS12#2 dsRNAs. Data represent the average of triplicate independent experiments normalized to Rps17 mRNA and plotted as fold increase relative to LacZ-treated control cells.

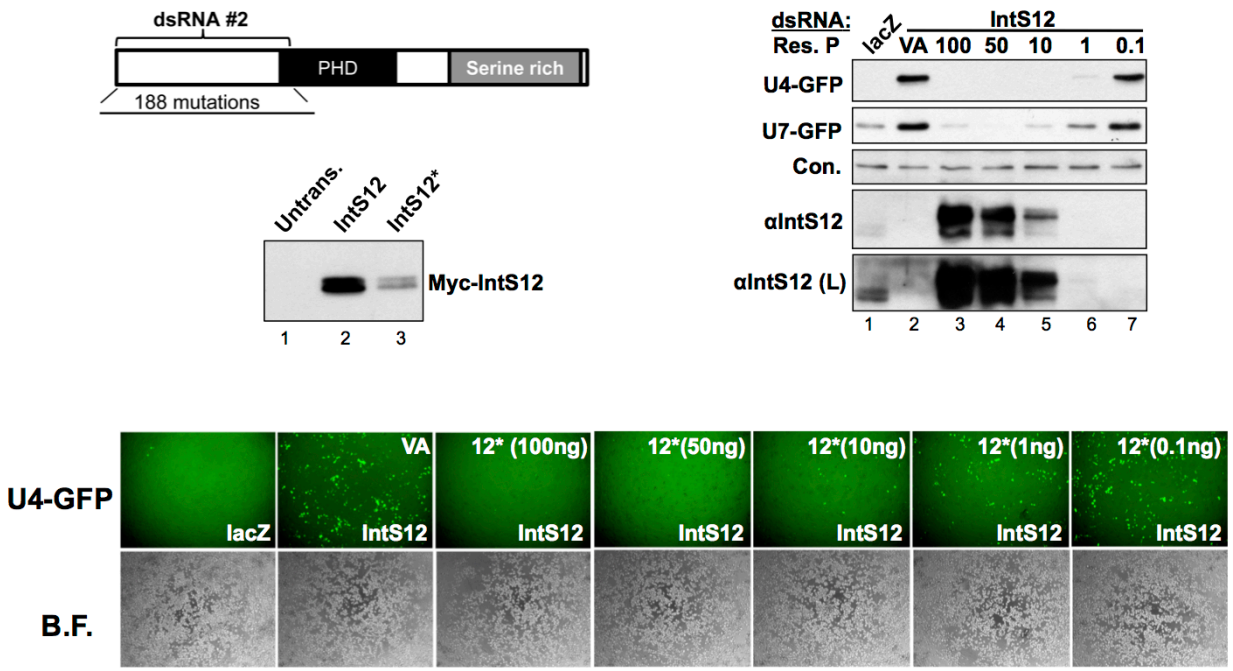


Figure 4.2 RNAi-resistant *IntS12 Rescues dsRNA-induced U4 and U7 snRNA Misprocessing.** (**A**) Schematic of the features encoded in the *Drosophila* (CG5491) *IntS12* gene, showing the relative location of dsRNA#2 and the 188 silent site changes used to generate the *IntS12** cDNA. (**B**) Western blot analysis of lysates from cells transfected with myc-tagged *IntS12* (wild-type) cDNA or *IntS12** cDNA using anti-myc antibodies. (**C**) Western blot analysis demonstrating dose-dependent rescue of the U7-GFP and U4-GFP misprocessing phenotype using the *IntS12** cDNA following RNAi-mediated depletion. Lanes 2-7 are from S2 cells treated with *IntS12* dsRNA#2 subsequently cotransfected with reporter plasmid and rescue plasmid DNAs. Doses of rescue plasmid (Res. P) are indicated in nanograms. (**D**) Representative fluorescence images of S2 cells treated as described in panel (C). In all cases "VA" is the abbreviation for transfecting cells with vector alone.

To this end, we cotransfected either the U4-GFP or U7-GFP reporters with decreasing amounts of plasmid DNA encoding *IntS12** into S2 cells depleted of endogenous IntS12 by treatment with dsRNA#2. Both fluorescence microscopy and Western blot analysis of lysates from transfected cells revealed a clear dose-dependent response between the amount of transfected *IntS12** and the level of snRNA 3' end processing as measured by both reporters (Figure 4.2C/D). As 10 ng. of transfected *IntS12** cDNA per 96-well was found to be the minimal amount required to achieve full rescue of IntS12 knockdown, this amount was used in all further experiments. Collectively, these results demonstrate that depletion of endogenous IntS12 leads to measureable and reproducible GFP expression from both U4-GFP and U7-GFP reporters and that dsRNA mediated knockdown of endogenous IntS12 can be fully rescued through the expression of an RNAi resistant *IntS12** cDNA.

The N-terminus is both Necessary and Sufficient for IntS12 Function.

Alignment of IntS12 protein sequences from multiple species identified two regions with significant homology (Figure 4.3). The largest region includes the amino acids within the PHD finger known to be important for zinc coordination as well as several other residues known to be essential for maintaining PHD structure (Bienz 2006). A smaller conserved region within the first 50 amino acids of the N-terminus bears no resemblance to any known motifs. Lastly, the C-terminal region consists of a poorly conserved serine rich region. To investigate the functional contribution of these conserved features to snRNA 3' end formation, we utilized the dual U7 or U4-GFP reporter together with the RNAi-rescue strategy we developed. To do this, we generated a series of RNAi-resistant deletion constructs derived from the *IntS12** cDNA and tested their ability to rescue snRNA misprocessing in cells depleted of endogenous IntS12 (Figure 4.4A). Constructs lacking various combinations of the N-terminus, centrally located PHD finger domain, or serine-rich C-terminus were myc-tagged and transfected into S2 cells. Western blot analysis using anti-myc antibodies confirmed protein expression from all constructs while some variability was observed (Figure 4.4B). To determine the relative ability of these deletion mutants to restore snRNA processing after knockdown of endogenous IntS12, we co-transfected each mutant construct with either the U7-GFP or the U4-GFP reporter into cells pre-treated with *IntS12* dsRNA#2. Data obtained using GFP fluorescence imaging (Figure 4.4C) and Western blot analysis using anti-GFP antibodies were congruent (Figure 4.4D). The results gathered using both GFP reporters were remarkably consistent and, to our surprise, demonstrated that the conserved PHD finger is not required for IntS12 to mediate snRNA 3' end processing. We found that constructs containing N-terminal amino acids (ΔC, NP, N, and ΔP) were as active as full-length (FL) protein in restoring reporter 3' end processing after endogenous IntS12 knockdown.

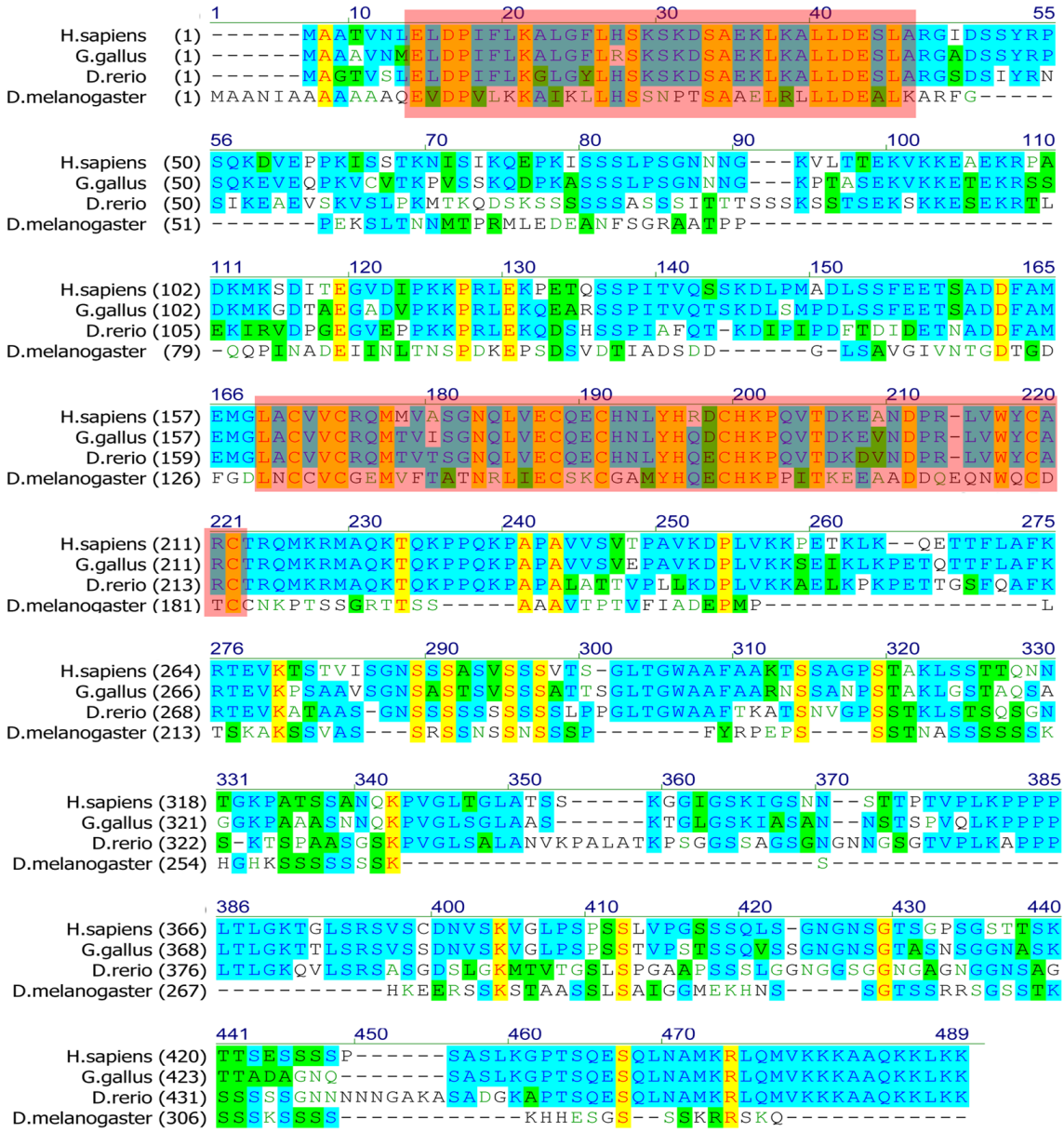


Figure 4.3 Protein sequence alignment of IntS12 from four metazoan species. Vector NTI Advance (Invitrogen) is used to generate the sequence alignment graph. Four different species are selected: human (*H.sapiens*), chicken (*G.gallus*), zebrafish (*D.rerio*) and fly (*D.melanogaster*). Blue highlighted residues represent similar amino acids and yellow highlights represent identical residues. The highlighted red boxes denote the identified functional microdomain and the defined PHD finger.

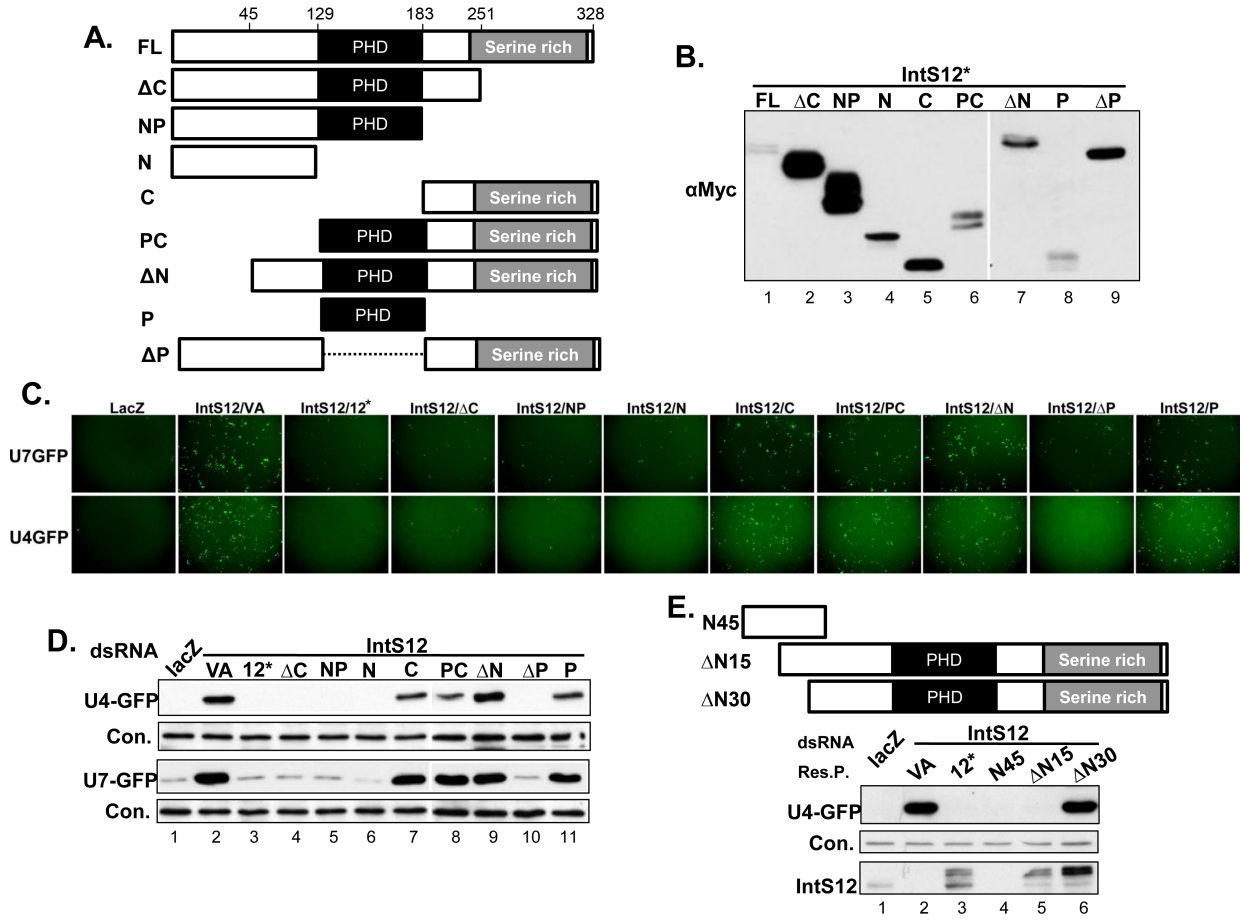


Figure 4.4 The N-terminus of *Drosophila* IntS12 is required for snRNA 3' end formation. (**A**) Schematic of IntS12* truncation and deletion constructs, which were designed based upon predicted domains. Relevant amino acid sequences are numbered. (**B**) Western blot analysis of cell lysates isolated from S2 cells transiently transfected with plasmids encoding myc-tagged IntS12* proteins. (**C**) Representative fluorescence images of S2 cells treated with either control dsRNA or IntS12 dsRNA#2 followed by cotransfection of either the U4-GFP or U7-GFP reporters with the myc-tagged IntS12* cDNAs. (**D**) Western blot analysis of cell lysates from panel (C). (**E**) Schematic and Western blot analysis of three additional IntS12* deletion mutants cotransfected with the U4-GFP reporter. In all panels "VA" stands for empty vector transfection as a control.

Conversely, all mutants lacking the N-terminus (C, PC, ΔN, and P) were incapable of restoring Integrator activity. These results indicate that the N-terminal 45 amino acids are required for 3' end processing activity while the first 129 amino acids are sufficient, whereas the PHD domain was dispensable.

Mutations within an N-terminal microdomain disrupt IntS12 Activity.

To further characterize the N-terminus of IntS12, we generated three more IntS12 constructs derived from *IntS12**. Two encode proteins harboring deletions of either the first 15 or first 30 amino acids, whereas a third generated a truncated protein consisting of only the first 45 amino acids (Figure 4.4E, upper panel). The ΔN15 mutant was observed to function as well as full-length IntS12; however, deletion of the first 30 amino acids (ΔN30) abolished IntS12's ability to rescue misprocessing of the U4-GFP reporter (Figure 4.4E, blot) despite being expressed at levels comparable to the FL protein. In addition, expression of the first 45 amino acids of IntS12 (N45) was sufficient to restore snRNA processing as effectively as the FL expression construct. These data demonstrate that amino acids 16-45 are required for IntS12 activity and that the first 45 amino acids are sufficient to restore snRNA 3' end processing after endogenous IntS12 knockdown.

Detailed examination of the evolutionary similarities within the N-terminal region of IntS12 (Figure 4.5A) shows high conservation of the residues located within the region identified as required for IntS12 function. The combination of evolutionary conservation and requirement in snRNA 3' end processing implicates this region as forming a functional microdomain critical for IntS12 activity. To achieve better resolution of the relative contribution of amino acids within this microdomain, we created a series of 6 mutant constructs (Mt1 through Mt6) spanning amino acids 16-45, each comprising 5 contiguous alanine substitutions in the context of the full-length *IntS12** (Figure 4.5A). In instances where alanine was encoded in the wild-type protein at a specific residue, no change was introduced. Therefore two of the mutants contained 4 amino acid changes (Mt2 and Mt6) and one (Mt4) contained only three substitutions. Western analysis confirmed each of these mutants to be expressed to similar levels at the wild-type IntS12 protein (Figure 4.5B), yet they displayed strikingly different abilities to promote snRNA 3' end processing. Mt1, Mt2, Mt4, and Mt6 were as capable as the wild-type IntS12 in restoring snRNA processing; however, both Mt3 and Mt5 were completely ineffective in abrogating GFP reporter gene expression (Figure 4.5B). When this same mutation series was retested in the context of the "N" construct (Figure 4.3A) comprising the first 129 amino acids only, virtually identical results were obtained (data not shown). These data define a highly conserved ~15 amino acid 'core' of a microdomain within the N-terminus

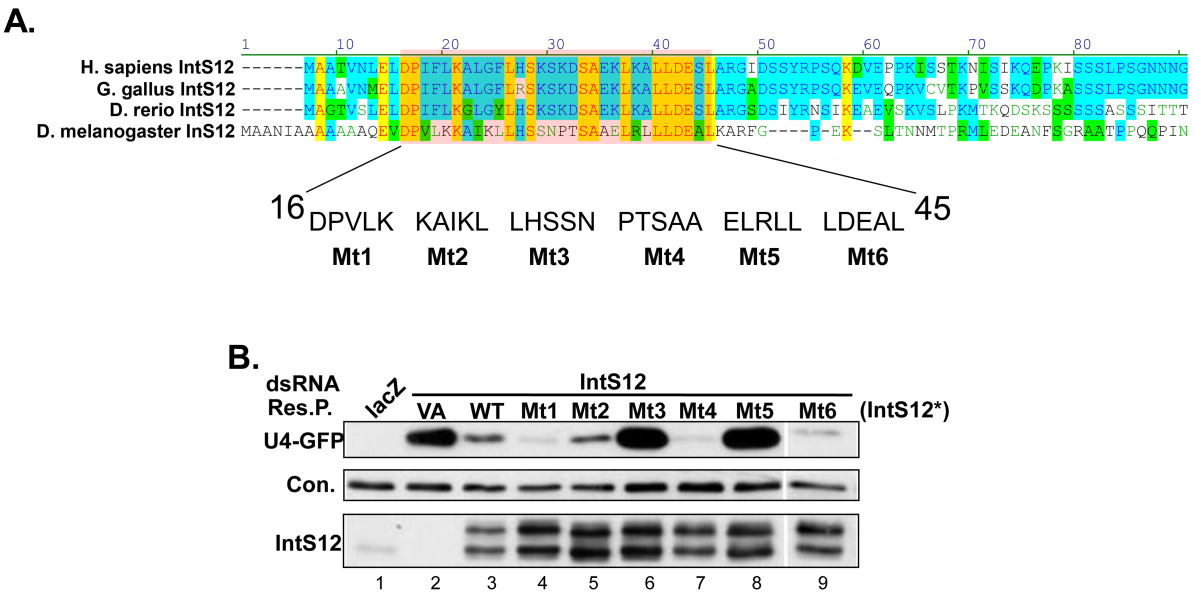


Figure 4.5 Mapping critical residues within the N-terminal IntS12 microdomain required for snRNA 3' end formation. (**A**) Upper panel, alignment of several species' IntS12 N-termini. Blue highlighted residues represent similar amino acids and yellow highlights represent identical residues. The highlighted red box denotes the identified functional microdomain; the labeled amino acids are the subject of six different alanine-scanning mutants. (**B**) Western blot analysis of cell lysates treated with either control dsRNA or IntS12 dsRNA#2 that were then cotransfected with U4-GFP reporter and IntS12* plasmids containing mutations as described in panel (A).

of IntS12 that is required for restoring snRNA 3' end processing after endogenous IntS12 knockdown.

The IntS12 Microdomain is Necessary and Sufficient for Incorporation of IntS12 into the Integrator Complex.

Given the discrete and conserved nature of the IntS12 microdomain, we hypothesize that it may function as a protein-protein interaction motif required for association with other member(s) of the Integrator complex. To test this possibility, we cloned full-length RNAi-resistant *IntS12** cDNAs encoding either full-length (FL), Mt3 or Mt5 proteins in frame with an N-terminal FLAG to facilitate immunoprecipitation studies. Analogous constructs comprising a deletion of the first 45 amino acids of IntS12* (ΔN) or only the first 45 amino acids at the N-terminus (N45) were also generated. We observed that different FLAG-tagged proteins restored 3' end processing of both the U7-GFP and U4-GFP reporters in an identical fashion to the myc-tagged proteins described above (Figure 4.6 v.s. Figures 4.4 and 4.5). We further determined the effect of these FLAG-tagged proteins on endogenous snRNA processing. We generated stable cell lines expressing FLAG-tagged IntS12* constructs, treated them with dsRNA targeting either control LacZ or IntS12 and measured snRNA misprocessing using the qRT-PCR assay described in Figure 4.1C. We observed that in the control stable line expressing the FLAG only, depletion of IntS12 led to a 4-5 folds increase in the levels of misprocessed endogenous U2 or U5 snRNA. Consistent with the reporter observations, stably expressing FLAG-IntS12* or FLAG-N45 in IntS12 depleted cells rescued the misprocessed endogenous U2 and U5 snRNAs to similar basal levels. In contrast, depletion of IntS12 in stable lines expressing FLAG-tagged Mt3, Mt5 or ΔN led to a 3-5 folds increase in misprocessed levels of endogenous U2 or U5 snRNA (Figure 4.7A). These data demonstrate that functionally the IntS12 microdomain is necessary and sufficient to mediate the 3' end formation of both reporter and endogenous snRNAs.

To test interactions of other Integrator subunits with the IntS12 microdomain, we generated nuclear extracts from these same stable cells lines and immunoprecipitated IntS12-associated proteins using anti-FLAG agarose beads (Figure 4.7B). While we observed rescue of snRNA processing using the N45 Int12 cDNA, levels of this protein were markedly reduced relative to the other stably expressed Int12 proteins obscuring the ability to directly compare IP efficiencies. Therefore, we initially compared the binding of endogenous Integrator subunits to the full length IntS12, Mt3, Mt5, and the ΔN. We observed robust and significant levels of endogenous IntS1 and IntS9 associating with FLAG-tagged full length (FL) IntS12 relative to control pulldowns (Figure 4.7B. lane 4 vs. 2). IntS12 protein expressed from Mt3- or Mt5-containing constructs poorly associated with endogenous IntS1 or IntS9 and this association

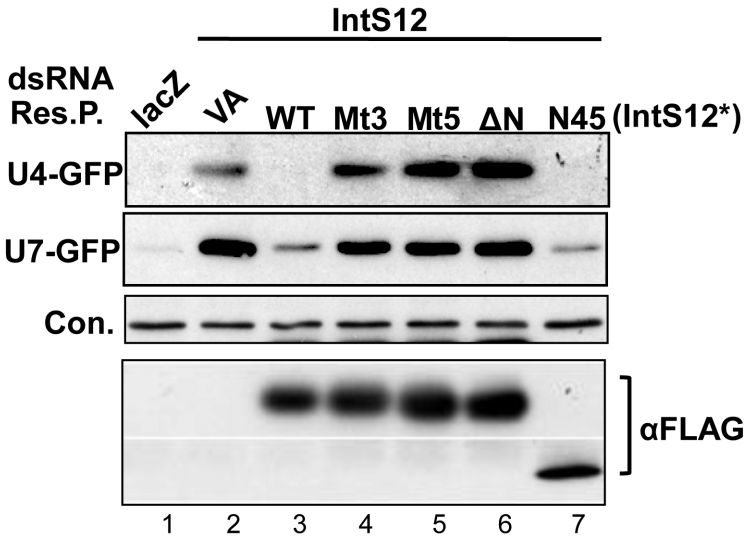


Figure 4.6 FLAG-tagged IntS12 microdomain restores processing of U4 and U7 snRNA-GFP reporters. Western blot analysis of cell lysates from S2 cells treated with either control dsRNA or IntS12 dsRNA#2 followed by cotransfection of FLAG-tagged rescue plasmids and either the U4- GFP or U7-GFP reporter.

was completely absent in cells expressing the ΔN IntS12 protein truncation (Figure 4.7B. lanes 6, 8, & 10 versus lane 4). These data demonstrate that mutations introduced into the IntS12 N-terminal microdomain compromise Integrator activity in snRNA 3' end processing by reducing the ability of IntS12 to interact with other Integrator subunits.

As the data presented in Figure 4.7A shows expression of only the first 45 amino acids of IntS12 is sufficient to rescue depletion of the endogenous protein, we next asked if the IntS12 microdomain alone (N45) is sufficient to mediate interaction with endogenous Integrator subunits. To circumvent the issue of low expression of the N45 peptide we generated stable cell lines expressing either FLAG-mCherry or FLAG-N45mCherry, where the first 45 amino acids of IntS12 was fused to the N-terminus of mCherry. To confirm that fusion to mCherry did not generate any unintended effects, we transfected FLAG-mCherry, FLAG-N45mCherry, or full-length FLAG-IntS12* into cells pretreated with IntS12 dsRNA#2. Western blot analysis determined that both FLAG-mCherry and FLAG-N45mCherry were expressed at comparable levels, however, only the FLAG-N45mCherry was capable of rescuing the snRNA processing defects associated with depletion of endogenous IntS12 (Figure 4.7C, lane 2 vs. 3). FLAG-tagged proteins from cell lines stably expressing these constructs were then immunoprecipitated utilizing anti-FLAG agarose and probed for their ability to pull-down endogenous IntS1 and IntS9. We found robust levels of both these subunits interacting with the FLAG-N45mCherry relative to FLAG-mCherry control (Figure 4.7D, lane 2 vs. 4). We also noticed that the FLAG-N45mCherry was not as efficient as FLAG-tagged full-length IntS12 in interacting with either the IntS1 or IntS9 subunit suggesting that residues beyond the first 45 amino acids are likely to contribute to the interactions between IntS12 and other members of the complex as well. Nevertheless, the N-terminal 45 amino acid IntS12 microdomain is sufficient for binding to IntS1 and IntS9 and to restore snRNA 3' end cleavage, either as a peptide (FLAG-N45) or in the context of a heterologous protein (FLAG-N45mCherry).

IntS1 Binding to the IntS12 Microdomain is Essential for Maintaining IntS1 levels.

As co-depletion is commonly observed between interacting proteins and has been found to occur between several members of the histone pre-mRNA processing complex, including Symplekin and CPSF73 (Sullivan et al. 2009), we next asked whether expression of endogenous IntS12 is dependent on any other Integrator subunit. To do that, S2 cells were treated in duplicate with non-overlapping dsRNAs for each of the Integrator genes for which we had antibodies (IntS1, IntS9, IntS11, IntS12), or with a control dsRNA (LacZ). Depletion of Integrator subunits was equally effective with either dsRNAs relative to control dsRNA treated cells (Figure 4.8A). In addition, both dsRNAs targeting IntS12 were as effective in depleting IntS1 protein level as were dsRNAs directly targeted IntS1,

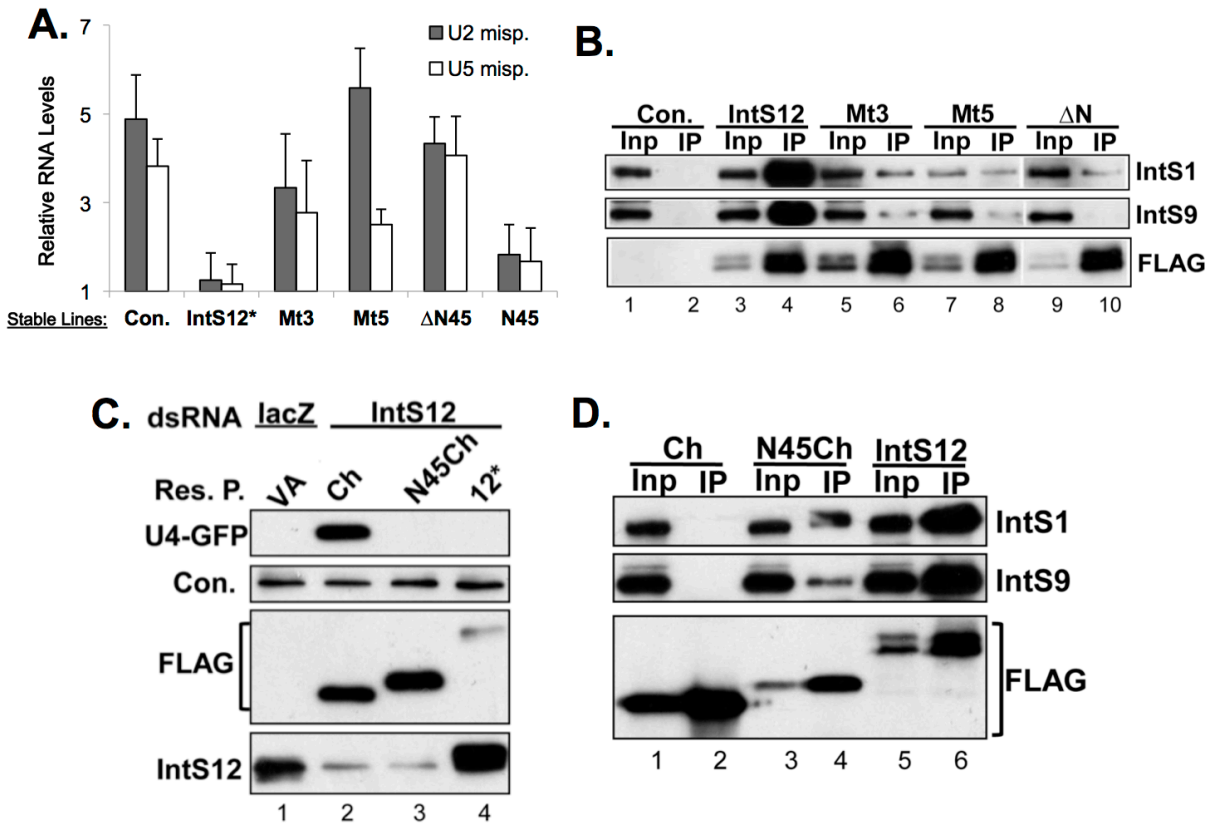


Figure 4.7 Mutations in the IntS12 Microdomain causing loss of function disrupt IntS12 interaction with endogenous Integrator subunits. (**A**) Quantitative real-time PCR measuring levels of misprocessed endogenous U2 or U5 snRNA in cells treated with dsRNA targeting IntS12. Control represents S2 cells expressing FLAG only while the cell lines treated with IntS12 dsRNA that are also stably expressing FLAG-tagged IntS12 proteins are labeled on the x-axis. All results are plotted as fold increase relative to LacZ dsRNA treatment and normalized to RpS17 mRNA levels. (**B**) Western blot analysis of immunoprecipitations using anti-FLAG agarose from nuclear extracts prepared from cell lines stably expressing FLAG-tagged IntS12* proteins. (**C**) Western blot analysis of cell lysates from S2 cells treated with either control dsRNA or IntS12 dsRNA#2 followed by cotransfection with U4-GFP and FLAG-mCherry plasmids with or without the N-terminal 45 amino acids of IntS12. (**D**) Western blot analysis of immunoprecipitations using anti-FLAG agarose from nuclear extracts purified from cell lines stably expressing FLAG-tagged IntS12* proteins and FLAG-mCherry proteins with and without the IntS12 Microdomain (N45). The upper panels are probed for endogenous IntS1/9 and the bottom panel is probed with anti-FLAG antibody to confirm pull down.

and reciprocally, dsRNAs targeting IntS1 were capable of co-depleting IntS12. In contrast, IntS9 and IntS11 levels were unaffected by depletion of IntS1 or IntS12 and targeting either IntS9 or IntS11 did not impact on the levels of IntS1 or IntS12. We repeated these experiments in the cell lines stably expressing RNAi-resistant full-length IntS12*, Mt3-containing or Mt5-containing full-length IntS12*, or the FLAG-N45mCherry proteins. We observed that expression of the resistant full-length IntS12* blocked the codepletion of IntS1 in response to treatment of cells with IntS12#2 dsRNA whereas neither Mt3- or Mt5-containing constructs were effective in maintaining IntS1 stability (Figure 4.8B, lane 2 vs lanes 6, 7). Consistent with the microdomain binding studies (Figure 4.7), stable expression of the first 45 amino acids of IntS12 was sufficient to stabilize IntS1 in cells when endogenous IntS12 had been depleted. Collectively, these data demonstrate that there is an interdependency of protein stability between IntS1 and IntS12 mediated by the IntS12 N-terminal microdomain and that amino acid substitutions in this microdomain that disrupt IntS12 activity and its association with endogenous Integrator subunits are also required to maintain endogenous IntS1 protein stability.

Finally, to examine the role of direct protein-protein interactions in Integrator complex composition and stability we utilized a directed yeast two-hybrid assay, since *Saccharomyces cerevisiae* does not encode any orthologous Integrator proteins. We expressed full-length *Drosophila* IntS12 fused to the Gal4 transcriptional activation domain (AD-IntS12) and all other members of the fly Integrator complex as individual fusions to the Gal4 DNA binding domain (BD-IntS1 through -IntS12). When individual BD fusions were coexpressed in cells along with AD-IntS12, only those expressing IntS1 and IntS10 supported growth on nutritional selection plates lacking histidine (Figure 4.9A). However, in this system we were unable to conclusively determine the existence of an IntS10/12 association, as expression of the BD-IntS10 alone supported growth on media lacking histidine. To determine if the yeast two-hybrid interaction between IntS1 and IntS12 was sensitive to mutations within the N-terminal microdomain of IntS12, we expressed AD-fusions of full-length, ΔN, N45, Mt3- or Mt5-containing full length IntS12 in the presence of BD-Ints1. We observed a total absence of growth on selective media lacking histidine for all IntS12 protein fusions that lacked or contained inactivating mutations within the N-terminus previously identified as unable to complement the effects of endogenous IntS12 knockdown (Figure 4.9B). Importantly, Western blotting confirmed loss of growth was indicative of loss of interaction rather than loss of expression (Figure 4.9C). Lastly, we examined the binding of Gal4 fused to the N-terminal 45 amino acids of IntS12 and found that this small peptide, when fused to Gal4 was nearly as capable as the full-length protein in supporting growth on selective media. These results demonstrate that the IntS12

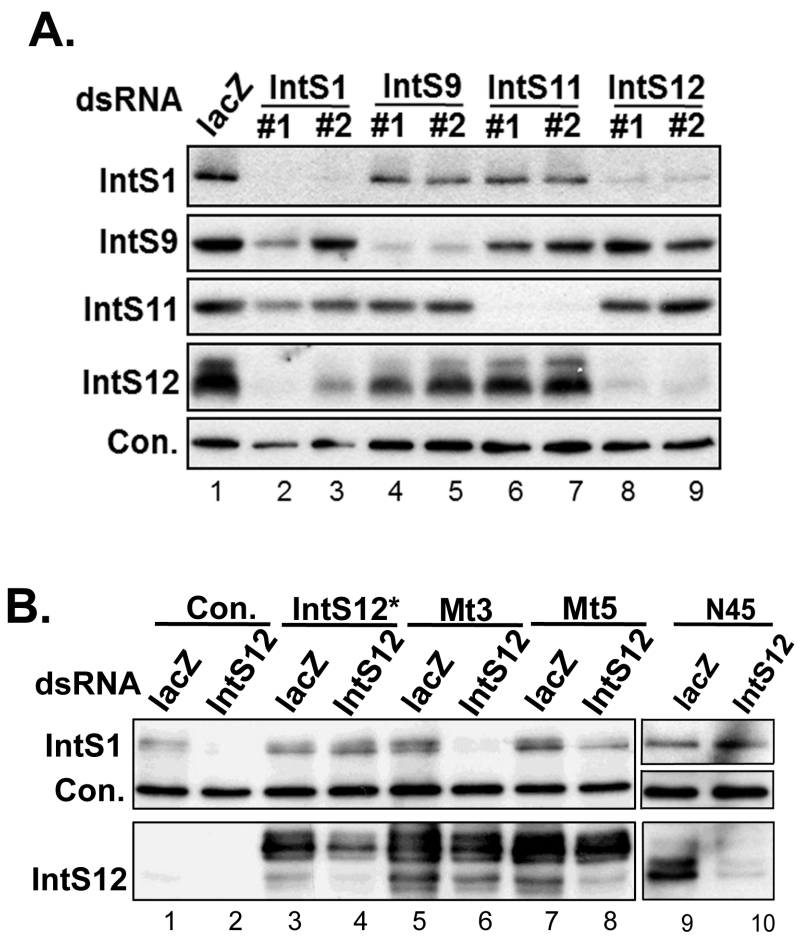


Figure 4.8 Expression of *Drosophila* IntS1 and IntS12 is interdependent and stability of IntS1 requires an intact IntS12 Microdomain. (**A**) Western blot analysis of endogenous Integrator subunit expression from S2 cells treated with various dsRNA targeting IntSs. (**B**) Western blot analysis of endogenous IntS1 expression in S2 cells stably expressing FLAG-tagged IntS12* proteins containing mutations within the IntS12 microdomain.

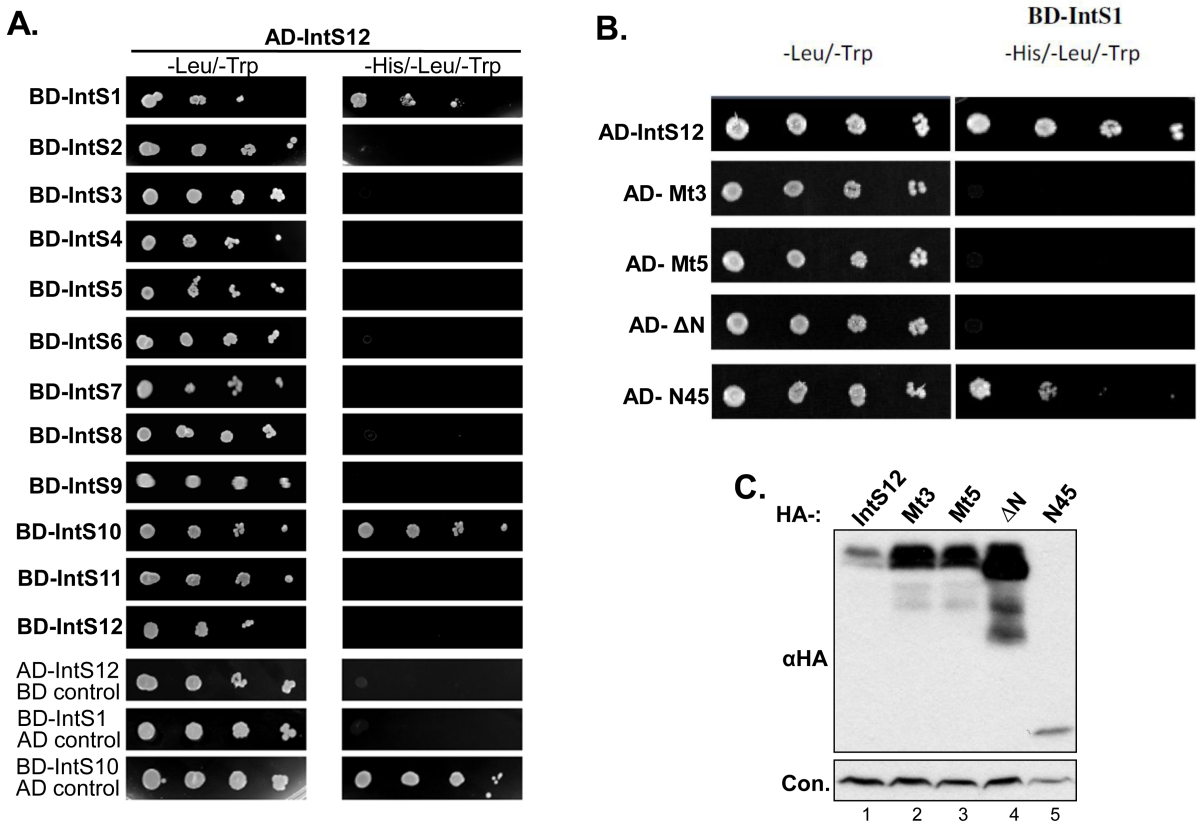


Figure 4.9 The IntS12 microdomain is required and sufficient to mediate interaction with IntS1 in the absence of other Integrator subunits. (**A**) S. cerevisiae (AH109) were cotransformed with plasmids encoding hybrid proteins containing IntS12 fused to the Gal4 activation domain (AD-IntS12) and each of the other Integrator subunits fused to the Gal4 DNA binding domain (BD-IntSs). A dilution series of overnight cultures was spotted on either SD/-Leu/-Trp plates or the same media without histidine to test interaction. All BD constructs were tested for auto-activation. (**B**) Similar to panel A, except AH109 yeast were transformed with full-length IntS12 containing mutations 3, 5, or a deletion of the 45 N-terminal amino acids (DN), or the first 45 amino acids of IntS12 (N45). (**C**) Western Blot analysis confirming the expression of HA-tagged IntS12 constructs in yeast strain AH109.
(*The work for yeast-two-hybrid assay shown here is collaborating with Dr. Bill Warren's group at James Cook University Australia, and I performed Western blot analysis to determine the expression of HA-tagged IntS12 constructs in yeast AH109.*)

microdomain binds to IntS1 in the absence of other Integrators and that the first 45 amino acids are sufficient to support a stable and autonomous interaction with IntS1.

DISCUSSION

Here we present a detailed functional analysis of the *Drosophila* IntS12 protein through characterizing its role in snRNA 3' end formation. Unexpectedly, our investigations determined that the highly conserved PHD finger is dispensable for IntS12 to promote nascent snRNA processing (Figure 4.4). Rather, we identified a small highly conserved ~30 amino acid microdomain near the N-terminus that is both necessary and sufficient for the restoration of Integrator function in IntS12 depleted cells (Figures 4.4-4.6). Residues within the IntS12 microdomain are also required for interaction with other endogenous Integrator complex subunits, and when mutated disrupt their stable interaction (Figures 4.7-4.9). These data establish a critical role for IntS12 in regulating the activity of the Integrator complex, mediated through binding to and stabilization of IntS1, the largest Integrator subunit. Additionally, our approach using dual snRNA GFP reporters in combination with functional rescue through expression of RNAi-resistant cDNAs represents a powerful tool going forward to elucidate function of the other Integrator subunits.

Functions of IntS12 microdomain.

Our results demonstrate a strong correlation between IntS12 microdomain binding to IntS1 and its ability to promote snRNA 3' end formation. This argues that this small interaction motif is critical for Integrator complex function. Moreover, results shown in Figures 4.7A and 4.7C reveal that the microdomain is nearly as effective as full-length IntS12 in restoring the processing of the reporter and endogenous snRNAs, thereby excluding a simple localization sequence function. However, as the FLAG-N45mCherry protein was not as effective at pulling down endogenous Integrator subunits as the full-length protein, other residues of IntS12 may play a minor role in achieving maximal stability of the IntS1/IntS12 complex (Figure 4.7D).

PSIPRED secondary structure prediction posits the IntS12 microdomain may form a helix-coil-helix structure (Jones 1999; McGuffin et al. 2000; Buchan et al. 2010) similar to the hepatocyte nuclear factor 1-alpha dimerization domain and the dimerization domain required for the heterodimeric association of the SinR-SinI anti-repressor complex (Johnen and Kaufman 1997; Colledge et al. 2011). In both of these cases, small helix-coil-helix domains act as crucial protein-protein interaction motifs that elicit dimerization. Based upon these examples, we propose that the microdomain of IntS12 binds to an as-yet unidentified complementary domain within IntS1. This IntS1/IntS12 interaction is not only essential for

IntS1 function but also is required for the stability of both IntS1 and IntS12 (Figure 4.8). In this respect, the IntS12 microdomain is behaving similar to the RNAse E microdomain that interacts with the glycolytic enzyme Enolase (Py et al. 1996; Chandran and Luisi 2006; Nurmohamed et al. 2010). RNAse E is the essential component of the *Escherichia coli* RNA degradosome and contains a N-terminal catalytic domain as well as a long C-terminal scaffolding domain (reviewed in (Carpousis 2007)). The scaffold domain binds to the RhlB helicase, the phosphorolytic exoribonuclease polynucleotide phosphorylase, and Enolase. The RNAse E microdomain resides within the scaffolding section and consists of a ~28 amino acid conserved region that binds to a dimeric interface between two Enolase proteins to promote the decay of specific mRNAs in *E. coli* (Bernstein et al. 2002; Bernstein et al. 2004; Chandran and Luisi 2006; Nurmohamed et al. 2010). Given the large size of IntS1, it may simultaneously interact with multiple other members of the Integrator complex in an analogous way to the scaffolding domain of RNAse E. Our data support a model where binding of IntS12 to IntS1 alters IntS1 confirmation allows further interaction(s) with other proteins and activation of the complex.

Chapter 5. Biochemical Analysis Identifies IntS12 as a Phosphoprotein and its Conserved PHD Finger Plays Roles in Histone H3 Interaction in vitro and Integrator Subunits Interaction in vivo

INTRODUCTION

The serendipitous biochemical purification of the twelve-member Integrator complex and the establishment of its role in snRNA 3' end processing has expanded the existing repertoire of the 3' end processing machinery for RNAPII transcripts (Baillat et al. 2005). The polyadenylated mRNA and poly(A)- histone mRNA use the same endonuclease CPSF73 for their 3' end endonucleolytic cleavage (Reviewed in (Dominski 2007; Dominski 2010)). Unlike these two classes of RNAPII transcripts, a distinct endonuclease called IntS11 catalyzes the 3' end processing of snRNA. IntS11 is homologous to CPSF73, and both proteins belong to the metallo-β-lactamase superfamily with a signature β-lactamase/ β-CASP domain (Weiner 2005) (See Figure 1.5 for a schematic representation of protein domain). In the Integrator complex, another Integrator subunit, IntS9, also has the MBL/β-CASP domain but with alterations in critical residues rendering it catalytically inactive. IntS9 shows significant homology with the mRNA 3' -end processing factor CPSF100. Recent work from our lab shows that IntS9 and IntS11 form a heterodimer through a unique C-terminal domain and that the interaction between these two subunits is required for the cleavage activity of the Integrator complex (Albrecht and Wagner 2012). Besides these two subunits, no significant homology has been identified between subunits of the complex and other known proteins, making it difficult to predict their roles in snRNA 3' end processing.

Most of Integrator subunits are conserved in metazoans and plants but absent in *Saccharomyces cerevisiae* (Baillat et al. 2005), which is consistent with the observation that yeast use a distinct Nrd1/Nab3/Sen1 complex to carry out the 3' end processing of snRNAs. A careful protein domain analysis of the remaining ten Integrator subunits reveals the presence of HEAT repeats in IntS4, ARM repeats in IntS4 and IntS7, a von Willebrand factor type A (VWA) domain in IntS6 and a plant homeodomain (PHD) finger in IntS12. The HEAT and ARM repeats are from a common phylogenetic origin and both mediate protein-protein interactions (Andrade et al. 2001). The VWA domain is found in both intracellular and extracellular proteins and in both cases likely functions in protein-protein interactions (Whittaker and Hynes 2002). The PHD finger domain is a reader domain of chromatin modifications (Bienz 2006; Musselman and Kutateladze 2011), and the IntS12 PHD finger is the second most conserved domain after the MBL/β-CASP domain of IntS9/11.

Drosophila IntS12 is a 328-amino acid protein, consisting of a well-conserved central PHD finger, a conserved N-terminal motif and a less conserved C-terminal serine-rich region (See Figure 4.3). The PHD finger is a well-defined signature chromatin-associated motif that uses the Cys4–His–Cys3 motif to coordinate two zinc ions to form an intertwined topology (Sanchez and Zhou 2011). At the secondary structure level, it consists of a two-strand anti-parallel β-

sheet and a C terminal α-helix (present in many PHDs). PHD fingers are commonly found in proteins that either possess catalytic activities or act as scaffolding proteins bridging multisubunit enzymatic complexes to specific genomic loci (Musselman and Kutateladze 2011; Sanchez and Zhou 2011). The biological outcome of an interaction for a particular PHD finger is usually determined by the function of the complex in which the PHD finger resides. It is widely accepted that PHD fingers are epigenetic effectors that recognize unique histone modifications present in the histone N -terminal tail, preferably the tail of histone H3 (Reviewed in (Musselman and Kutateladze 2011)). Recently, PHD fingers are also implicated in recognition of non-histone proteins where PHD finger acts as a protein-protein interaction motif to bind cofactors and further facilitate histone recognition (Hom et al. 2010; Miller et al. 2010). The Drosophila IntS12 PHD finger contains ~55 amino acids and is phylogenetically conserved cross metazoan species. Its mouse homologue has been experimentally determined to adopt a typical PHD topology in solution (Figure 5.1A). Moreover, sequence alignment of IntS12 PHD finger to other known PHD fingers indicates that it is most closely related to a subgroup of PHD fingers that recognize unmodified histone H3 tail (Figure 5.2A/B). Based on these evidence and the complex context that IntS12 PHD finger resides in, it is possible that IntS12 acts as an adaptor to bring the Integrator complex to snRNA genomic loci by recognition of unique chromatin modifications through its conserved PHD finger domain. Alternatively, given the emerging evidence that some PHD fingers have the ability to bind protein co-factors, it is also possible that IntS12 PHD finger works as a protein interacting motif to mediate Integrator complex assembly.

In this study, to test these possibilities mentioned above, we characterized several biochemical properties of IntS12 PHD finger and investigated their functional involvement in snRNA 3' end processing. We also mapped a residue threonine 76 important for IntS12 phosphorylation, and investigated the biochemical and cellular roles of this phosphorylation in Integrator interaction and snRNA 3' end processing. Our data reveal that IntS12 PHD is able to bind histone H3 in vitro and enhance association with other Integrator subunits in vivo. However, functional analysis indicated that association with histone H3 or enhanced interaction with other Integrator subunits through the PHD finger is not essential for reporter snRNA 3' end processing. We also demonstrate that IntS12 phosphorylation through T76 is functionally not required for snRNA 3' end processing but may play a negative role in mediating protein-protein interaction between IntS12 and the other Integrator subunits. Collectively, these data characterized several features of IntS12 protein and demonstrate their function in Integrator protein interaction. We envision that a yet-to-be identified cellular/biological function for the Integrator complex besides snRNA biosynthesis may involve in the biochemical properties of IntS12 that we described here.

RESULTS

The PHD finger of Integrator subunit 12 binds histone H3 *in vitro*

Recent studies show that PHD fingers have a general preference toward recognizing unmodified or modified N-terminal tail of histone H3 thereby involving them in fundamental cellular processes such as transcriptional regulation, nucleosome remodeling, and DNA recombination (Reviewed in (Musselman and Kutateladze 2011; Sanchez and Zhou 2011)). IntS12 has a central PHD finger that is conserved across different metazoan species (Figure 5.1A, left panel). The Nuclear Magnetic Resonance (NMR) structure of the mouse IntS12 PHD finger (PDB id: 1wev) shows that this PHD finger adopts a typical PHD zinc finger structure consisting of an antiparallel beta-sheet and a α-helix in the C-terminus (Figure 5.1A, right panel). To determine whether the conserved IntS12 PHD finger is able to recognize specific histones, an *in vitro* glutathione S-transferase (GST) pull-down assay was performed by incubation of recombinant GST-tagged human IntS12 PHD finger, *Drosophila* IntS12 full length protein, or *Drosophila* IntS12 PHD finger with a purified mixture of histones (Worthington, NJ). Bound histones were co-precipitated with GST-tagged proteins by the glutathione agarose beads, and detected by Western blot analysis using histone antibodies. Both fly and human IntS12 PHD fingers as well as the full-length fly IntS12 protein were found to specifically bind histones even in the presence of 500mM NaCl in the incubation buffer (Figure 5.1B, left panel). Further analysis of the co-precipitates using individual histone antibodies revealed that histone H3 is specifically co-precipitated with IntS12 PHD finger (Figure 5.1B, right panel).

To further confirm the binding specificity of IntS12 PHD finger toward histone H3, ten IntS12 PHD finger point mutants were generated and used for the GST pull-down assay. These mutants were constructed by substituting highly conserved residues in IntS12 PHD fingers (Figure 5.1A, left panel) to alanine using site-directed mutagenesis. Two of these IntS12 PHD mutations (L145 and E147) were found to significantly reduce the levels of co-precipitated histone H3 (Figure 5.1C), indicating their important roles in histone H3 interaction. Both residues were found to reside in the first β-strand of IntS12 PHD finger, which has been shown to play essential roles in recognition of histone H3 tails (Pena et al. 2006; Shi et al. 2006). These results are highly similar to two previously well-characterized cases of recognition of tri-methyl Lysine 4 of histone H3 by the Inhibitor of Growth protein 2 (ING2) PHD finger and recognition of unmodified lysine 4 of histone H3 by the autoimmune regulator (AIRE) PHD finger (Pena et al. 2006; Org et al. 2008). In these instances, the first β-strand of PHD finger pairs with another β-strand from the histone H3 tail to form an antiparallel β-sheet to facilitate histone H3 tail binding (Figure 5.1D). Superimposition of the mIntS12 PHD finger

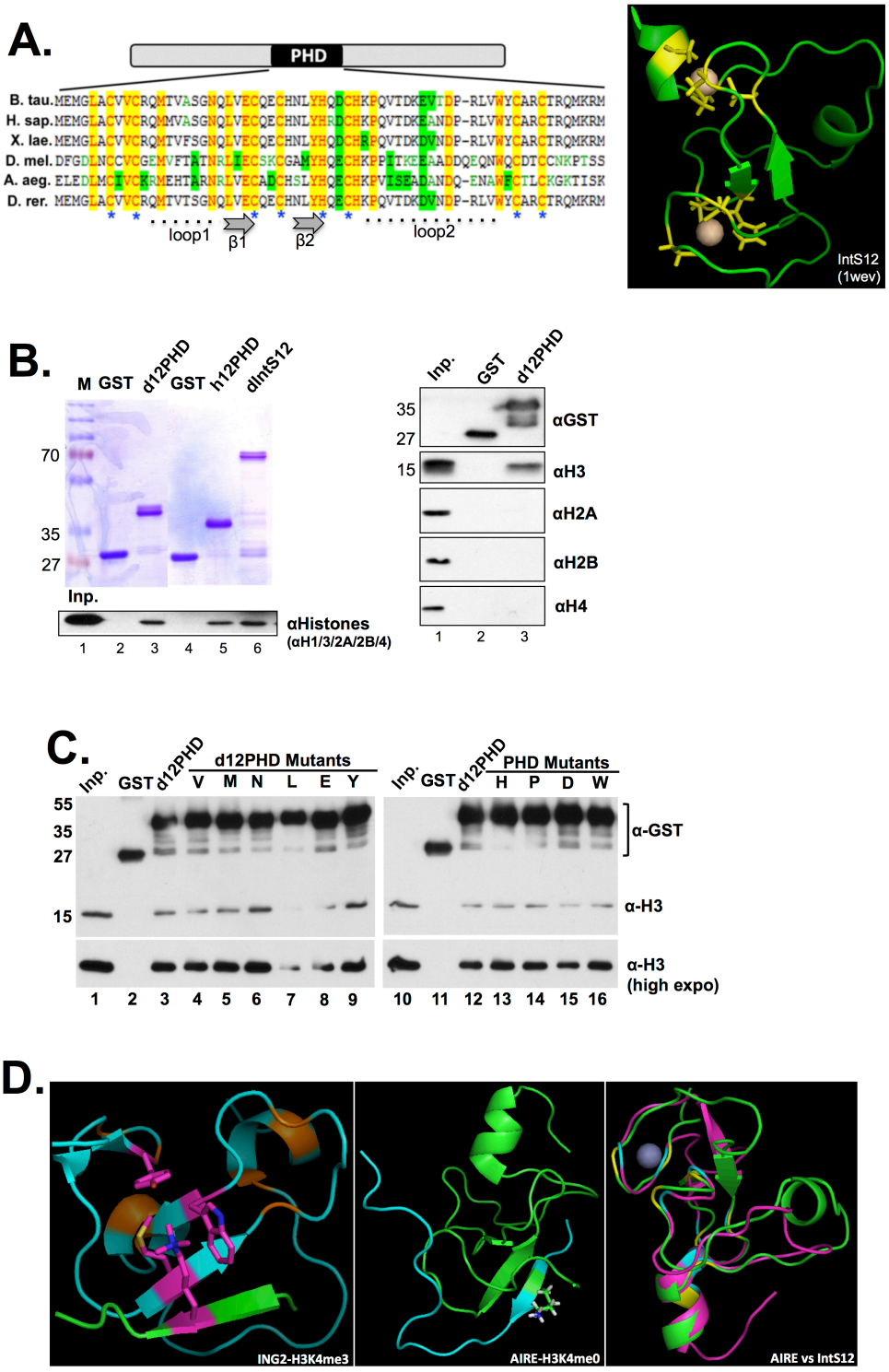


Figure 5.1 The plant homeodomain (PHD) finger of Integrator subunit 12 binds histone H3 in vitro.

Figure 5.1 The plant homeodomain (PHD) finger of Integrator subunit 12 binds histone H3 in vitro.

(**A**)(left). Amino acid sequence alignment of IntS12 PHD finger across several different metazoan species. Green highlighted residues represent similar amino acids and yellow highlights represent identical residues. Blue stars on the bottom highlight the eight residues coordinating zinc atoms, and the dashed lines and solid arrows indicate residues forming the loops and β-strands secondary structures. (**A)**(right). NMR structure of the mouse IntS12 PHD finger. Two zinc atoms are shown as balls, and zinc chelating cysteines and histidine are highlighted in yellow with side chain shown. (**B)**(left). GST pull-down assay was carried out using recombinant fly or human IntS12 PHD co-incubating with purified histone mixture. Glutathione beads bound GST-tagged proteins were visualized by Coomassie staining, and histones were detected by western blot analysis using a mixture of histone antibodies. (right). Co-precipitated histone proteins were detected by western blot using individual histone antibodies. (**C)**. GST pull-down assay was performed using recombinant IntS12 PHD finger point mutants. Co-precipitated histone H3 were shown as two different exposures. (**D)**. 3D-structures of ING2 PHD finger recognition of H3K4me3 or AIRE PHD finger recognition of H3K4 (left, middle), Pymol 3D-superimposition of AIRE PHD finger with mIntS12 PHD finger (right).

on mING2 PHD finger or human AIRE PHD finger by using the molecular graphics tool PyMOL demonstrates that IntS12 PHD finger adopts a very similar tertiary structure to the PHD fingers tested with minor difference where mIntS12 PHD finger contains a significantly longer loop2 region (Figure 5.1D). Collectively, all data indicate that the IntS12 PHD finger is structurally similar to those for histone H3 recognition, and is able to bind histone H3 in vitro.

IntS12 PHD finger does not recognize canonical histone modifications

To further determine the specificity of IntS12 PHD finger toward unique histone modifications, the amino acid sequence of IntS12 PHD finger was first aligned to various subgroups of known PHD fingers using the homology-extended alignment and a predicted secondary structure algorithm. IntS12 PHD finger is most homologous and structurally similar to PHD fingers (AIRE-PHD1, CHD4-PHD2, TRIM24) that bind to unmodified histone H3 lysine 4 (H3K4me0) rather than those (DPF3-PHD1) that bind acetylated histone H3 lysine (H3K14ac) or those that bind (Pygo1, MLL1-PHD3, BPTF, ING2) tri-methylated histone H3 lysine 4 (H3K4me3) as shown by both sequence alignment (Figure 5.2A) and tree representation (Figure 5.2B). However, a distinct characteristic of IntS12 PHD finger compared to the H3K4me0-recognition PHD fingers is the presence of two much longer loop regions, a feature that is also observed in two PHD fingers (Pygo1, MLL1-PHD3) capable of binding non-histone proteins (Figure 5.2A). Overall, in silico data analysis indicates that IntS12 PHD finger is likely to recognize either an unmodified histone H3K4me0 marker or bind to a non-histone protein.

To experimentally determine the potential unique histone modification recognized by the IntS12 PHD finger, a biotinylated histone peptide pull-down assay was performed. In brief, GST-tagged IntS12 PHD finger recombinant proteins were incubated with an array of biotin-labeled, 21-23 amino acid-long peptide corresponding to different regions of the N-terminal histone tails with each containing a unique lysine methylation (H3, 1-21; H3, 22-44; H3, 66-88; H4, 1-25; H3K4, K9, K27, K36, K79 and H4K20 mono-, di- or tri- methylated peptide). Bound IntS12 PHD finger proteins were co-precipitated with biotin-labeled histone peptides by using streptavidin agarose beads, and were detected by Western blot analysis using anti-GST antibodies. The PHD finger of lysine demethylase PHF2 was used as a positive control in this study, which recognizes di- and tri-methylated histone H3K4. As a proof-of-principle, PHF2 PHD finger was shown to specifically bind di-methylated H3K4 under both salt conditions tested in our biotinylated pull-down assay (Figure 5.2C, top panel). We followed this with systematic biotinylated pull-down analysis, however, this revealed that neither the human nor *Drosophila* IntS12 PHD finger was able to be co-precipitated with any of the biotin-labeled histone peptide used for this study (Figure 5.2C, bottom panel), suggesting that either IntS12

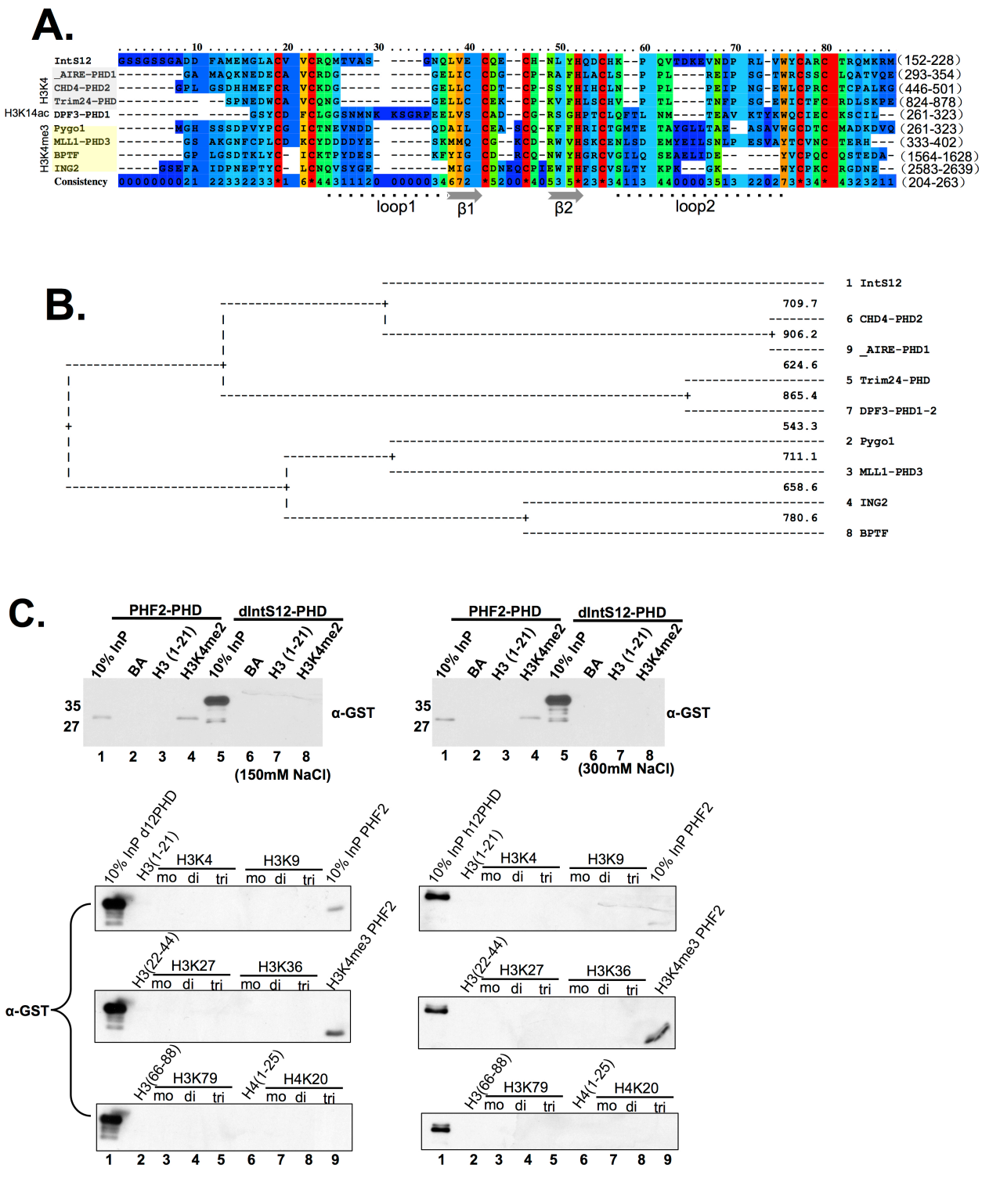


Figure 5.2 IntS12 PHD finger does not recognize canonical histone H3 modifications.

Figure 5.2 IntS12 PHD finger does not recognize canonical histone H3 modifications. **(A)**. Representation of conserved amino acids from structure-based alignments of PHD finger sequences using PRALINE. The absolutely conserved Zinc-coordinating residues are shown in red, and the two core β-strands are shown in gray. Residues are colored according to sequence conservation. The residue numbers corresponding to the PHD finger in the full-length protein are shown in parentheses on the right. (**B**). Tree representation of structure-based alignments of representative PHD finger sequences using PRALINE. The score indicates the similarity between two close PHD fingers (the higher, the more similar), and the distance from the branch (+) to the right side indicates similarity as well. Two PHD fingers in the same branch indicate a similar structure, and the closer the branch on the right, the more similarity present in the two groups of PHD fingers within the same branch. (**C**). Biotinylated histone peptides were unable to pulldown GST-tagged recombinant IntS12 PHD finger. Biotin-labeled histone peptides each containing a single unique modification were incubated with recombinant IntS12 PHD finger or PHF2 PHD finger (positive control), and were precipitated by streptavidin agarose beads. Co-precipitated recombinant PHD fingers were detected by western blot analysis using anti-GST antibodies. PHF2 input and pulldown are separated on two blots.

PHD finger recognizes a histone modification marker that is not tested here or it may require other co-factors for histone recognition similar to the property observed in Pygo1 and MLL1.

IntS12 is a phosphoprotein and Threonine 76 residue is responsible for IntS12 phosphorylated.

IntS12 migrates as a doublet in SDS-PAGE gel detected by Western blot analysis, suggesting that this protein may be phosphorylated (Figure 5.3A, left panel). To formally test that hypothesis, phosphatase treatment and site-directed mutagenesis analysis of IntS12 were performed. Briefly, cell lysates from untransfected S2 cells or cells transfected with Myc-tagged IntS12 were treated with alkaline phosphatase (AP) or lambda protein phosphatase (λPP), and the migration shift of IntS12 proteins was detected by Western blot analysis using anti-IntS12 or anti-myc antibodies. Both endogenous and myc-tagged exogenous IntS12 were found to migrate faster (lower band) after phosphatase treatment (Figure 5.3A, Middle and right panels), indicating that IntS12 is a phosphoprotein. To narrow down the phosphorylated region, six myc-tagged IntS12 deletion mutants (N, NP, C, ΔP, PC, P) were generated and tested for migration profile after phosphatase treatment. IntS12 N-terminal containing deletion mutants (N, NP, ΔP) were found to both migrate as doublets and faster after phosphatase treatment (Figure 5.3B, lane 2, 4), indicating that the phosphorylation is present in the N-terminus. To further map the position of the phosphorylated residue, potential phosphorylation sites within IntS12 N-terminus were predicted using the Kinase-specific Phosphorylation Sites Prediction Tool GPS 2.0 (Xue et al. 2008). Eight potential phosphorylation sites (S2829, S33, T56, T60, T76, T91, S93) were predicted, and substituting of these sites with alanine did not affect myc-tagged IntS12 N-terminus migration with the exception of the Threonine 76 (Figure 5.3 C). The T76A mutation lost its doublet migration pattern and instead ran as a clear single faster migration (Figure 5.3C, let panel lane 5).The single band was completely inert to phosphatase treatment (Figure 5.3C, right panel lane 5, 6), strongly indicating that the T76 is an important residue responsible for IntS12 phosphorylation. A close examination of the amino acid sequence context of the phosphorylation shows the presence of a consensus cyclin-binding motif (RXL) and a distinct threonine-proline (TP) phosphorylation motif for mitogen-activated protein kinase (MAPK) or CDK, suggesting that it is a potential substrate either of these two protein kinases. Our previous genome-wide RNAi screen identified CycC/Cdk8 as an important factor required for snRNA 3' end formation. It is possible that Cdk8 is involved in snRNA 3' end formation through post-translational modification of IntS12 activity. However, mutation of the potential consensus cyclin-binding motif or depletion of CycC or Cdk8 protein in cells did not affect IntS12 phosphorylation pattern (data not shown), suggesting other Cdk

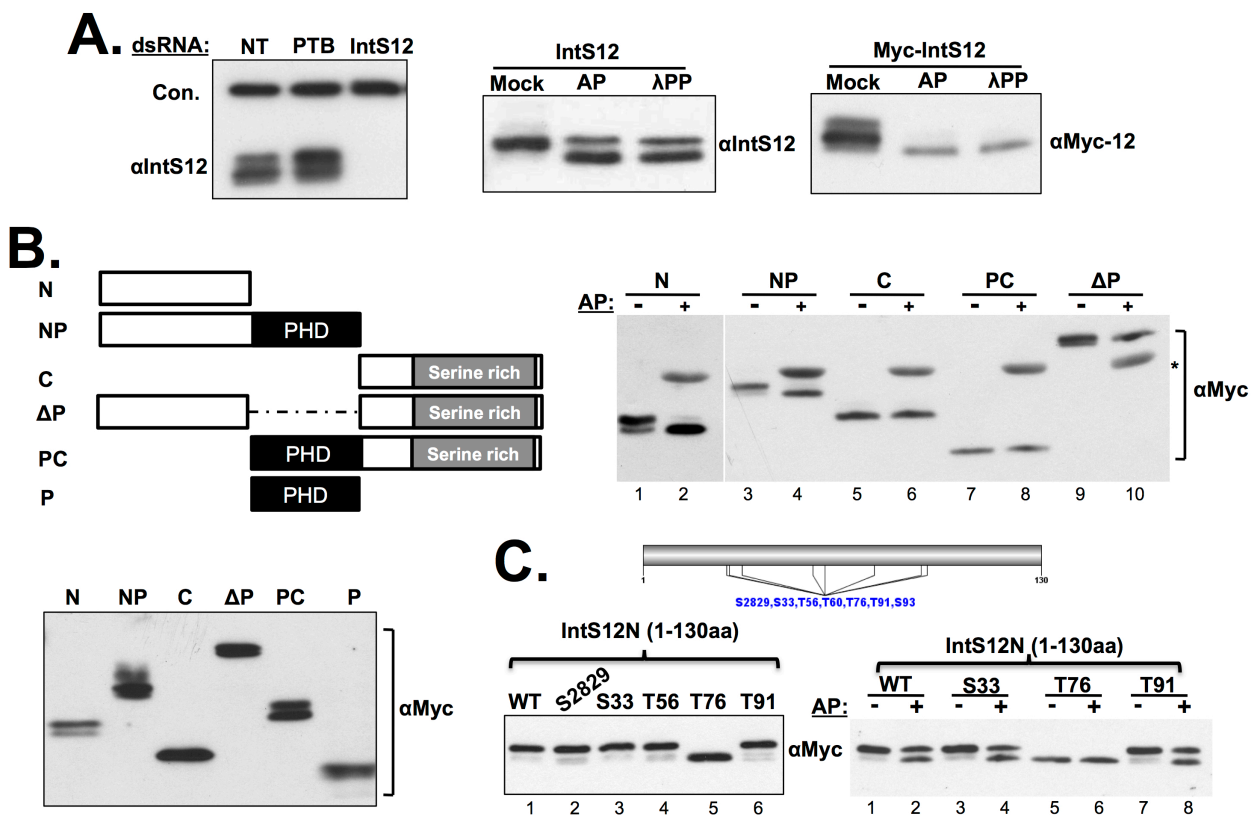


Figure 5.3 IntS12 is a phosphoprotein and the residue threonine 76 is important for IntS12 phosphorylation.

(**A**)(left). IntS12 is observed to migrate as doublets by Western blot detection. (Middle, right). Alkaline phosphatase or lambda protein phosphatase treatment of IntS12 resulted in a downshift for both endogenous and myc-tagged exogenous IntS12 as detected by Western blot. (**B**)(left top). Schematic representation of IntS12 deletion constructs for mapping phosphorylation region. (Left bottom, right top). Myc-tagged IntS12 deletion mutants expressed in S2 cells were detected by Western blot before and after alkaline phosphatase treatment. The star labeled band represents cross-reactant from the alkaline phosphatase to the anti-myc antibodies. (**C**)(top). Representation of potential phosphorylation sites in IntS12 N-terminus predicted by GPS 2.0. (bottom). Western blot analysis of cell lysates from cells transfected with potential phosphorylation site point mutants before and after alkaline phosphatase treatment.

kinases may be responsible for IntS12 phosphorylation. Taken together, the data presented here show the threonine 76 is responsible for *Drosophila* IntS12 phosphorylation in cells.

IntS12 PHD finger enhances interaction between IntS12 microdomain and other Integrator subunits.

We described in last chapter that the 45 amino acid IntS12 microdomain is fully functional for reporter snRNA 3' end formation, however, the interactions between IntS12 microdomain and IntS1, IntS9 were not as strong as that observed for full-length Integrator IntS12. This prompted us to test whether other parts of IntS12, especially the conserved PHD finger, is involved in biochemical association between IntS12 and other Integrator subunits. To test that possibility, S2 stable cell lines expressing FLAG-tagged mCherry (Ch, negative control), full-length IntS12 (IntS12), C-terminal fused mCherry of IntS12 N-terminal 1-130 a.a. (12NCh), C-terminal fused mCherry of IntS12 NP 1-185 a.a. (12NPCh) and N-terminal fused mCherry of IntS12 C-terminus 185-328 a.a. (Ch12C) were generated (Figure 5.4A). Then, anti-FLAG immunoprecipitation assays using nuclear extracts prepared from these stable cells were conducted. Consistent with previous pull-down results (Figure 4.7B/D), significant amount of endogenous IntS1 and IntS9 were found to be co-precipitated with full-length IntS12 (Figure 5.4B, lane 4 v.s. lane 2). We surprisingly found that the IntS12 C-terminal truncation mutant (12NPCh) was able to pulldown the endogenous IntS1 and IntS9 at the levels equivalent to full-length IntS12 (Figure 5.4B, lane 10 v.s. lane 4), and significantly better than N-terminus only (12NCh) or C-terminus only (Ch12C) (Figure 5.4B, lane 10 v.s. lane 6,14). These results suggest that IntS12 PHD finger plays an important role in biochemical interaction between IntS12 and other Integrator subunits, though it was found dispensable for snRNA 3' end formation in our snRNA-GFP reporter assay.

Since we characterized an important residue T76 for the IntS12 phosphorylation and specific interaction of IntS12 and histone H3 in vitro, here we wanted to biochemically test the involvement of these features of IntS12 with other Integrators. To accomplish this, two S2 stable cell lines expressing FLAG-tagged C-terminal fused mCherry of IntS12 N-terminal phosphorylation mutant 1-130 a.a. (12NT76ACh) and a C-terminal fused mCherry of IntS12 NP histone binding-deficient mutant 1-185 a.a. (12NPLACh) were generated, and were tested for interactions by anti-FLAG immunoprecipitation assays. I found that the N-terminal phospho-mutant (12NT76ACh) was able to significantly increase the co-precipitated endogenous IntS1 and IntS9 to the levels similar to full-length IntS12 compared to the deletion mutant containing IntS12 N-terminus (Figure 5.4B, lane 8 v.s. lane 4, 6), suggesting that the T76-mediated phosphorylation is likely to negatively regulate IntS12 association with other Integrators. In contrast, the L145 mutation in the IntS12 C-terminal deletion mutant

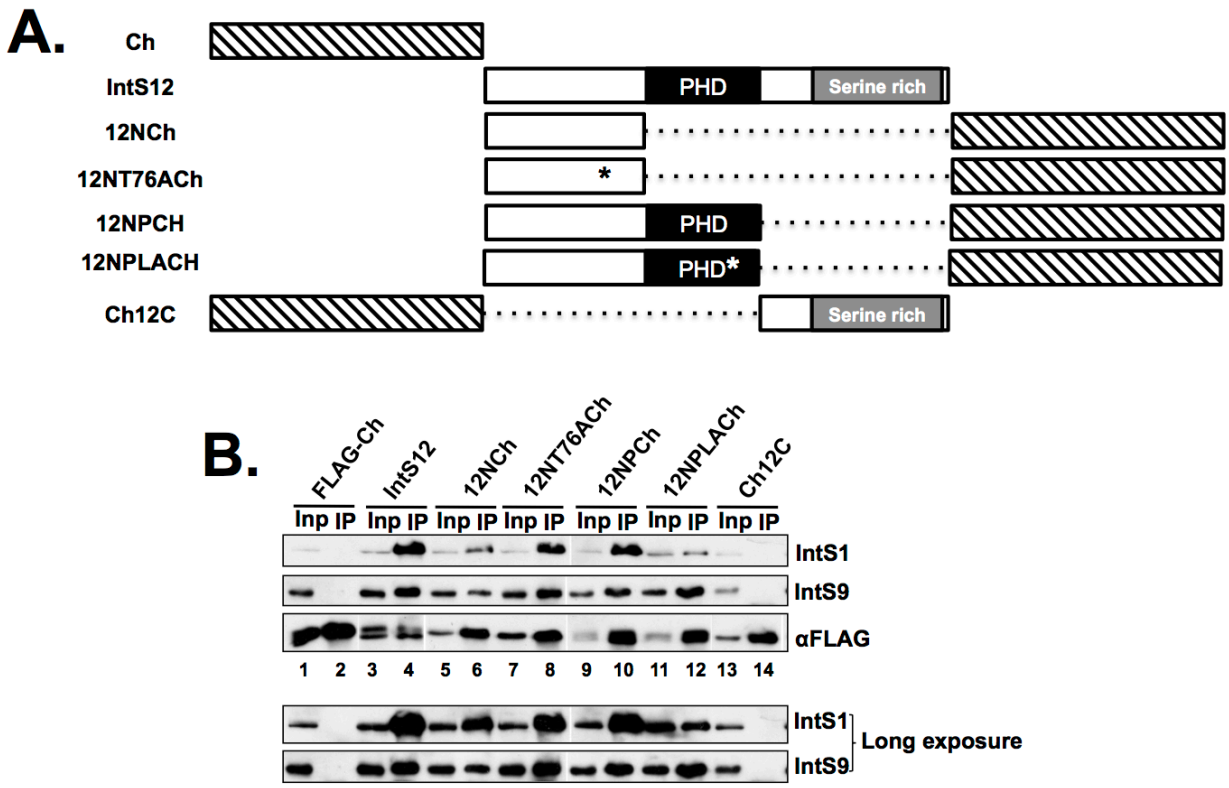


Figure 5.4 IntS12 PHD finger enhances association between IntS12 microdomian and other Integrator subunits.

(**A**). Schematic representation of IntS12 mCherry fusion constructs for study of IntS12 domains important for Integrator interaction. The star within two of the constructs represents a point mutation being introduced. (**B**). Western blot analysis of immunoprecipitates using anti-FLAG agarose from nuclear extracts purified from cell lines stably expressing FLAG-tagged mCherry-fused IntS12* proteins. The upper panels are probed for endogenous IntS1/9 and the bottom panel is probed with anti-FLAG antibody to confirm pull down. Two different exposures were presented.

(12NPLACh) was found to significantly reduce the pull-down levels of IntS1 and IntS9 compared to that observed for IntS12 C-terminal deletion mutant (12NPCh) (Figure 5.4B, lane 12 v.s. lane 10), indicating that the L145 residue, which has been determined to be important for *in vitro* histone H3 binding, is important for *in vivo* Integrator interaction as well. Taken together, these data presented here suggest that besides interacting with histone H3 *in vitro*, the IntS12 PHD finger is also likely to play an important role in enhancing interaction between IntS12 microdomain and other Integrators in vivo and that this interaction might be modulated by the T76-related phosphorylation.

IntS12 PHD finger and T76-related phosphorylation are not required for reporter snRNA 3' end processing.

To further test the functional requirement of the IntS12 T76-related phosphorylation and histone or Integrator interaction mediated through the PHD finger in snRNA 3' end formation, a functional RNAi rescue assay was performed as described previously (See Chapter 2 RNAi-rescue/snRNAGFP reporter for details). S2 cells were treated with either control dsRNA (LacZ) or IntS12 dsRNA and co-transfected with U4-GFP reporter (See Chapter 2 Materials and Methods section for details) and IntS12 constructs described in the pulldown assay. Cells were imaged and subsequently lysed for Western blot analysis for GFP expression. In IntS12 depleted cells, the mCherry vector (mChe) and mCherry fused IntS12 C-terminus only (mChe12C) completely failed to restore proper 3' end processing of U4-GFP reporter as significant GFP fluorescence was observed (Figure 5.5A/B). For the remaining constructs, we observed reduced GFP fluorescence to background level in IntS12 depleted cells (Figure 5.5A/B), indicating that all of these constructs were fully functional in restoring proper reporter snRNA 3' end processing. The phosphorylation mutant (12NT76AmChe) and the histone H3 binding mutants (12NP(EA)mChe, 12NP(LA)mChe) fully restored reporter snRNA 3' end formation, implying that IntS12 T76-mediated phosphorylation and binding to histone H3 are dispensable events for processing of the reporter snRNA 3' end.

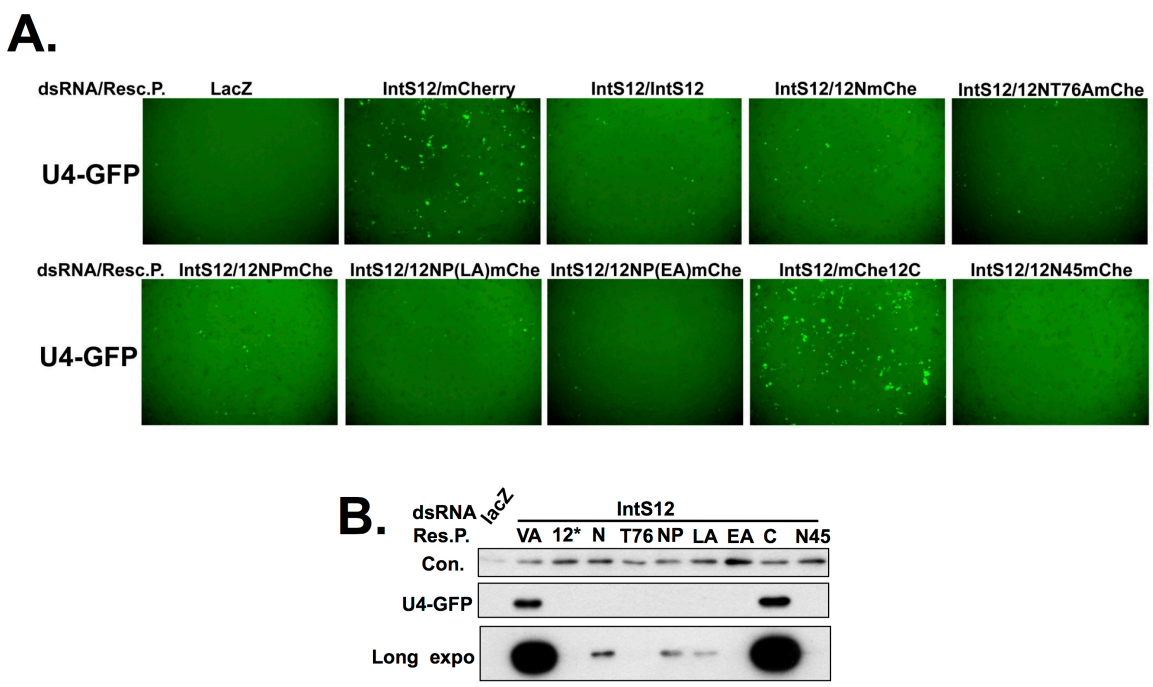


Figure 5.5 IntS12 PHD finger and a phosphorylation are not required for reporter snRNA 3' end processing.

(**A**). Representative fluorescence images of S2 cells treated with either control dsRNA (LacZ) or IntS12 dsRNA#2 followed by cotransfection of the U4-GFP reporter with the FLAG-tagged mCherry-fused IntS12* cDNAs. (**B**). Western blot analysis of cell lysates from panel A.

Discussion

In this chapter we presented the biochemical characterization of IntS12 and investigated the functional involvement of these features in snRNA 3' end formation. The IntS12 PHD finger was determined to bind histone H3 *in vitro* (Figure 5.1) and enhance Integrator protein interaction *in vivo* (Figure 5.4). The former result is consistent with the prevailing findings that PHD fingers have the ability to recognize a variety of histone modification markers (Musselman and Kutateladze 2011; Sanchez and Zhou 2011). However, the latter result indicates that IntS12 PHD may also contribute to protein-protein interactions, which is supported by two recently reported findings demonstrating that the PHD finger from Pygo1/2 or MLL1 is able to recognize non-histone proteins (Hom et al. 2010; Miller et al. 2010). Our data presented here support both non-exclusive models for IntS12 PHD action: act as a histone reader to bring the Integrator complex to specific genomic loci or to work as a protein-protein interacting domain for Integrator complex assembly. Further experiments need to be conducted to better understand these potential working mechanisms. We also determined that T76 is responsible for IntS12 phosphorylation, and our data suggest the phosphorylation is likely a negative regulator for Integrator protein interaction. Nevertheless, snRNA-GFP reporter combined with RNAi-rescue assay indicates that all of these biochemical features of IntS12 protein are dispensable for snRNA 3' end processing, making the roles of these characterized features unknown. It is very likely that these features we identified here contribute to a yet-to-be characterized cellular process.

Why has evolution selected for the central IntS12 PHD finger?

The central PHD finger region of IntS12 is one of the few domains found among the Integrator subunits whose function can be readily inferred from sequence conservation aside from the conserved MBL/β-CASP domains of IntS9/11. Members of the PHD finger family of proteins are involved in a diverse range of biological functions; yet typically bind to specific histone H3 posttranslational modifications (reviewed in (Bienz 2006; Musselman and Kutateladze 2011)). IntS12 PHD finger forming a canonical PHD finger topology is supported by NMR structure analysis of the mouse IntS12 PHD finger (pdbid:1WEV; He F, Muto Y, Inoue M, Kigawa T, Shirouzu M, Terada T, Yokoyama S, unpublished). Here we have also experimentally confirmed that recombinant *Drosophila* and Human IntS12 PHD finger proteins bind histone H3 *in vitro*. While nascent transcription from an snRNA promoter is essential for Integrator function, a central role for chromatin in 3' end formation has not been established (de Vegvar et al. 1986; Hernandez and Weiner 1986; Cazalla et al. 2011; Ezzeddine et al. 2011). Although the presence of a stable nucleosome located between the DSE and PSE in

human cells has been reported by several groups (Stunkel et al. 1997; Boyd et al. 2000; Zhao et al. 2001; Pavelitz et al. 2008), the preponderance of evidence argues for a general deficiency of histones within snRNA genes, with the snRNA promoter establishing a perpetual "open" transcription state (Pavelitz et al. 2008; Egloff et al. 2009). In *Drosophila*, the lack of a DSE (Hernandez 2001) and our own observations that the Proximal Sequence Element A (PSEA) from the snRNA gene promoter alone is sufficient to impart Integrator sensitivity to our actin 5C promoter driven reporters (JC, EJW, unpublished), suggests there is not likely a stable nucleosome present at fly snRNA promoters. These data together with our observations that the PHD domain is dispensable for snRNA 3' end processing argues against the IntS12 PHD finger coupling the Integrator complex to the snRNA promoter via histone binding.

We cannot exclude the possibility that, the IntS12 PHD finger is involved in recruiting the Integrator complex to chromatin at non-snRNA genes, though there is presently no indication that the Integrator complex functions elsewhere in the genome. Whether the endogenous target of the IntS12 PHD finger is indeed histone H3 in vivo has also yet to be determined. There are numerous examples of PHD fingers that associate with non-histone substrates (Musselman and Kutateladze 2011) including the MLL1 methyltransferase PHD3 finger bound to the RNA recognition motif (RRM) of nuclear cyclophilin Cyp33 (Fair et al. 2001). The IntS12 PHD finger may behave similarly and interact with a non-histone partner.

Chapter 6. Conclusions and Future Directions

OVERVIEW, SUMMARY AND FUTURE DIRECTIONS

As the functional snRNPs are well known for their role in removal of introns from pre-mRNA, it is predictable that any event disturbing the snRNP biogenesis pathway is likely to affect downstream mRNA maturation, which in turn may have a significant impact on the biological outcomes. One such well-documented case is SMA, which is caused by reduced expression of the SMN due to mutations in the gene (Burghes and Beattie 2009; Coady and Lorson 2011; Workman et al. 2012). The SMN complex plays critical roles in snRNP biogenesis in that it is responsible for loading the Sm core onto the snRNA Sm site and facilitates subsequent snRNP nuclear import (Figure 1.3A). Though it is still unclear how exactly a mutation in SMN causes motor neuron death, it has been well documented that snRNP maturation and mRNA splicing are both significantly affected in SMA patients and mouse models (Gabanella et al. 2007; Zhang et al. 2008). Currently, the most tantalizing model for the cause of motor neuron dysfunction and death in SMA patients is that inadequate expression of SMN protein leads to insufficient snRNP assembly, which in turn affects critical pre-mRNA splicing events that are essential for motor neurons. The results from SMA research vividly demonstrate the importance of the snRNP biogenesis pathway in human health.

The biogenesis of RNAPII-transcribed snRNA encounters a sophisticated nuclear and cytoplasmic life cycle where they begin as snRNA precursors through a transcription-coupled 3' -end processing event. Therefore, the nuclear 3' end cleavage of snRNAs is an integral part for snRNP maturation, and this event is brought about by the poorly understood Integrator complex (Baillat et al. 2005). Twelve different polypeptides were initially thought to be present in the Integrator complex, which associates with the C-terminal domain of RNAPII largest subunit Rpb1. Analogous to the SMN complex, the Integrator complex is also essential for snRNP biogenesis, and disruption of the Integrator complex is expected to affect downstream cellular processes and biological outcomes where a functional snRNP is required, such as the pre-mRNA splicing. Indeed, analysis of the RNA-sequencing data from cells depleted of the catalytic subunit of the Integrator complex (IntS11) revealed a global defect in pre-mRNA splicing, manifested by prevalent exon skipping and intron retention (Wagner lab, unpublished data). It was also reported in Zebrafish that disruption of Integrator subunit 5 leads to a specific red blood cell differentiation defect that is caused by missplicing of genes essential for the hematopoiesis signaling pathway (Tao et al. 2009). Moreover, studies from several different groups also show a requirement for the Integrator complex in mouse and fly development, where it is likely to function through regulation of the snRNP biogenesis (Rutkowski and Warren 2009; Ezzeddine et al. 2011). While these data contribute to understanding of the

significance of the Integrator Complex, the molecular details of the 3' end processing of the RNAPII transcribed snRNAs are far from understood.

The existing working model for transcription-coupled 3' end processing of snRNA is described in Figure 6.1. The key questions to be addressed are how the Integrator Complex is recruited to snRNA genes and also how it recognizes nascent transcripts to cleave them at the cleavage site. Studies of the other two RNAPII transcripts (polyadenylated mRNA and nonpolyadenylated histone mRNA) have established the requirements of two different RNA-binding proteins/protein complexes through recognition of two conserved cis-regulatory sequence elements present in each transcript. These binding events serve to recruit the catalytic core to the cleavage site and position it for proper 3' end processing (Figure 1.4). Interestingly, two functionally important 3' end elements, a conserved 3' box element and a less conserved 3' stem loop element have been identified in snRNAs (Hernandez 1985; Ezzeddine et al. 2011), however, the factors responsible for recognition of these elements are not known yet. An added challenge to identify these factors is that no identifiable RNA-binding domains are found in the known Integrator complex (Figure1.5). Moreover, analogous to 3' end formation of polyadenylated mRNA, the snRNA 3' end formation is a transcription-coupled process. The RNAPII CTD serine 5 phosphatase Ssu72 has been reported to serve as an adaptor to bridge the RNAPII CTD and the cleavage/polyadenylation factor Symplekin (Pta1 in yeast), thus coupling the mRNA transcription and subsequent 3' end cleavage and polyadenylation (He et al. 2003). During the period of this work, another RNAPII CTD serine 5 phosphatase RNA Pol II-associated protein 2 (RPAP2) has been proposed to play an analogous role of Ssu72 to bridge the RNAPII CTD to the Integrator complex by specific recognition of the pSer7 marker on the RNAPII CTD. RPAP2 is then thought to be important for the recruitment of Integrator complex to RNAPII CTD (Egloff et al. 2012). Confounding this result, however, is the subsequent observation that the RPAP2 yeast homolog Rtr1, lacked a phosphatase catalytic domain and was devoid of dephosphorylation activity toward RNAPII CTD *in vitro* (Xiang et al. 2012). These results do not rule out a model where RPAP2 serves as an adaptor between RNAPII CTD and the Integrator complex but reduce the possibility that RPAP2 exhibits any catalytic activity. Finally, the identity of the Integrator subunit that would behave analogously to the mammalian cleavage/polyadenylation factor Symplekin to couple transcription and 3' end formation is still not determined yet.

The studies described in this dissertation are aiming to probe these biologically important questions. To address these questions, I first identified missing protein factors that are functionally required for snRNA 3' end processing through a functional RNAi screen in *Drosophila* S2 cells in Chapter 3. This was followed by a detailed structural and functional analysis of one Integrator subunit that has the potential to lay the foundation to address the

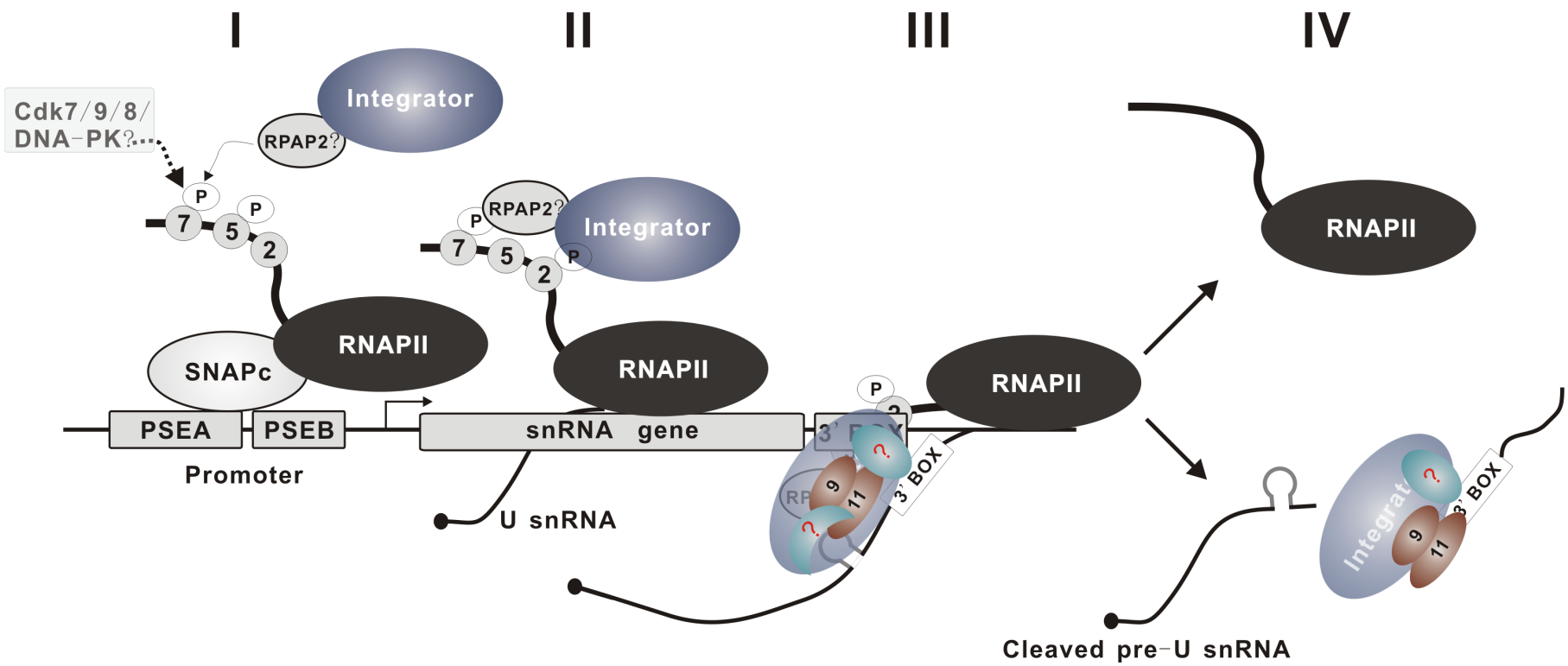


Figure 6.1 Schematic model for transcription-coupled metazoan snRNA 3' end processing through the Integrator complex.
(I) The process is initiated by recognition of the RNAPII CTD pSer7 marker by the RNA Pol II-associated protein 2 (RPAP2), which may recruit the Integrator complex to the RNAPII at the snRNA gene promoter. The kinase responsible for Ser7 phosphorylation is not fully characterized yet. (II) The phosphorylation of the Ser2 and dephosphorylation of the Ser5 kicks off the elongation process governing by the newly formed elongation complex. (III) Once the RNAPII surpasses the 3' box element, the Integrator proteins recognize the snRNA 3' end processing signals (3' box and a potential 3' stem-loop) and cleaves the nascent snRNA transcripts. (IV) The cleaved precursor snRNA is released and gets to the nuclear export pathway and the Integrator complex may also function in transcription termination.

above-mentioned questions. Below, I reflect on the significance of these two efforts and expand on what I believe to be their meaning and future implications.

Functional RNAi screen identifies protein factors required for snRNA 3' end formation.

In Chapter 3, I described a genome-wide RNA interference (RNAi) screen in *Drosophila* S2 cells to identify important genes required for snRNA 3' end formation by utilization of an snRNA-GFP reporter. The U7-GFP is a sensor of the U7 snRNA 3' end formation, and only knockdown of a gene causing U7 snRNA 3' end cleavage defect will result in expression of GFP in cells. Using this reporter, the screen identified 21 genes that when depleted leads to strong or moderate levels of GFP expression in cells. This screen determined that 10 out of the 12 known Integrator proteins (not IntS3 or IntS10) are functionally required for snRNA 3' end formation. Interestingly, 11 novel factors were identified in our screen as well. These genes encoding protein factors are involved in a variety of cellular functions, including protein ubiquitination, phosphorylation, transport, DNA replication and others. More importantly, I also revealed that Asunder and CG4785 are two additional core Integrator proteins in the Integrator complex as both are functionally required for snRNA 3' end formation and biochemically associate with the known Integrator complex. Moreover, the CycC/Cdk8 kinase was determined to be required for proper snRNA 3' end formation where the kinase activity of CdK8 is involved but is likely working through a pathway independent of the Mediator complex. At this stage, it is still unclear how exactly these four factors carry out their function in snRNA 3' end formation.

The Asu/IntS13 was originally identified to be a regulator of mitotic cell cycle and *Drosophila* development (Stebbings et al. 1998), and recently was found to regulate the perinuclear dynein localization during *Drosophila* spermatogenesis (Anderson et al. 2009). Asu/IntS13 was also identified as a regulator of small RNA biogenesis in a S2 cell RNAi screen (Zhou et al. 2008). Here we reported a novel function for Asu/IntS13 in snRNA 3' end formation. Whether the reported different function of Asu/IntS13 could be ascribed to the same pathway in cells is not clear, but it is possible that Asu/IntS13 may function in multiple pathways as the reported Asu/IntS13 associated processes take place in different cellular compartments consistent with its bimodal localization observed in our study (Chen et al. 2012). The Asu/IntS13 is a conserved protein across metazoan species and belongs to the cell cycle regulator Mat89Bb family that contains a well-conserved but functionally unknown DUF2151 domain (Figure 6.2). It would be interesting to determine the functional domain(s) mediating the differential roles of Asu/IntS13, especially in snRNA 3'end processing by using the powerful RNAi/rescue/snRNA-GFP reporter system described in Chapter 4. An effort to characterize the potential direct Integrator binding partner of Asu by directed Y2H assay failed

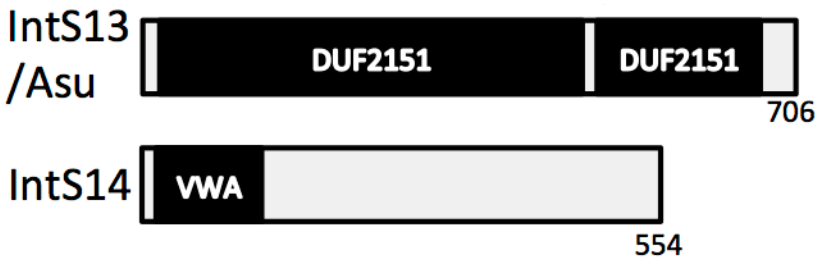


Figure 6.2 Schematic representation of protein domains of Asu/IntS13 and IntS14/CG4785.

Protein domain search is conducted by using Pfam analysis, and protein domains are labeled in black. The species used for conservation analysis are human, cow, chicken, *Drosophila* and zebrafish. Abbreviations: DUF, domain of unknown function; VWA, von Willebrand factor type A domain; IntS, Integrator subunit.

to score any known Integrator protein, leaving the functioning details of Asu in snRNA 3' end formation an open question to be addressed.

IntS14/CG4785 is a protein with much less known and there is a weakly defined von Willebrand factor type A (VWA) domain in its N-terminal (Figure 6.2). A previous RNAi screen in S2 cells for protein factors involved in small RNA (miRNA/siRNA) pathways also scored IntS14 as a positive regulator (Zhou et al. 2008), implying that IntS14 may function in sRNA biogenesis in *Drosophila*.

The identification of CycC/Cdk8 kinase in our screen was unexpected as it is predominantly present in the Mediator Complex of the RNAPII, which is likely separated from the Integrator Complex (Baillat et al. 2005; Galbraith et al. 2010). In eukaryotic cells, Cdk8 is a cell-cycle dependent protein kinase but has not been observed to oscillate throughout nor affect the cell cycle. CycC/Cdk8 is most characterized as part of the kinase module that interacts with the Mediator Complex and consists of CycC, Cdk8, Mediator 12 (Med12) and Med13. This module plays an important role in transcriptional regulation of gene expression as a means of cellular adaptation to different environmental cues. Our results (Figure 3.4D, Figure 6.3) demonstrate that the CycC/Cdk8 but not other Cdks are specifically required for U7 reporter snRNA 3' end formation, and its kinase activity in particular is involved in this process. This is the first report that implicates the function of CycC/Cdk8 in snRNA 3' end formation, and the underlying mechanism of this reaction is not known yet. Our results also rule out the possibility that CycC/Cdk8 contributes to snRNA 3' end processing through transcriptional control of Integrator expression as depletion of CycC or Cdk8 in cells did not affect expression of the Integrator subunit mRNA or protein (Figure 3.5). We believe that the function of these two proteins may be analogous to their role in the RNAPII-associated Mediator Complex, where the Mediator Cdk8 module can phosphorylate the Rpb1 CTD as well as a specific Mediator subunit (Knuesel et al. 2009). In the case of the Integrator Complex, CycC/Cdk8 may phosphorylate the Rpb1 CTD or Integrator subunit. Indeed, our results showed that CycC/Cdk8 functions independent of the Mediator Cdk8 module as depletion of the other two essential components of the Mediator Cdk8 module Med12 and Med13 did not affect the snRNA 3' end processing (Figure 3.7A/B). Furthermore, a small fraction of CycC/Cdk8 was found to associate with Integrator subunits but neither Med12 nor Med13 was found to associate with the Integrator Complex(Figure 3.7C/D). Taken together, all of the existing data support the presence of an Integrator Cdk8 module where the CycC/Cdk8 phosphorylates either the RNAPII CTD or the Integrator subunit to regulate the transcription-coupled snRNA 3' end formation. Since phosphorylation of the Ser2 and Ser7 in the Rpb1 CTD has been determined to be essential events for the Integrator complex recruitment as shown that mutation in these two residues abolished integrator complex association with RNAPII in vitro

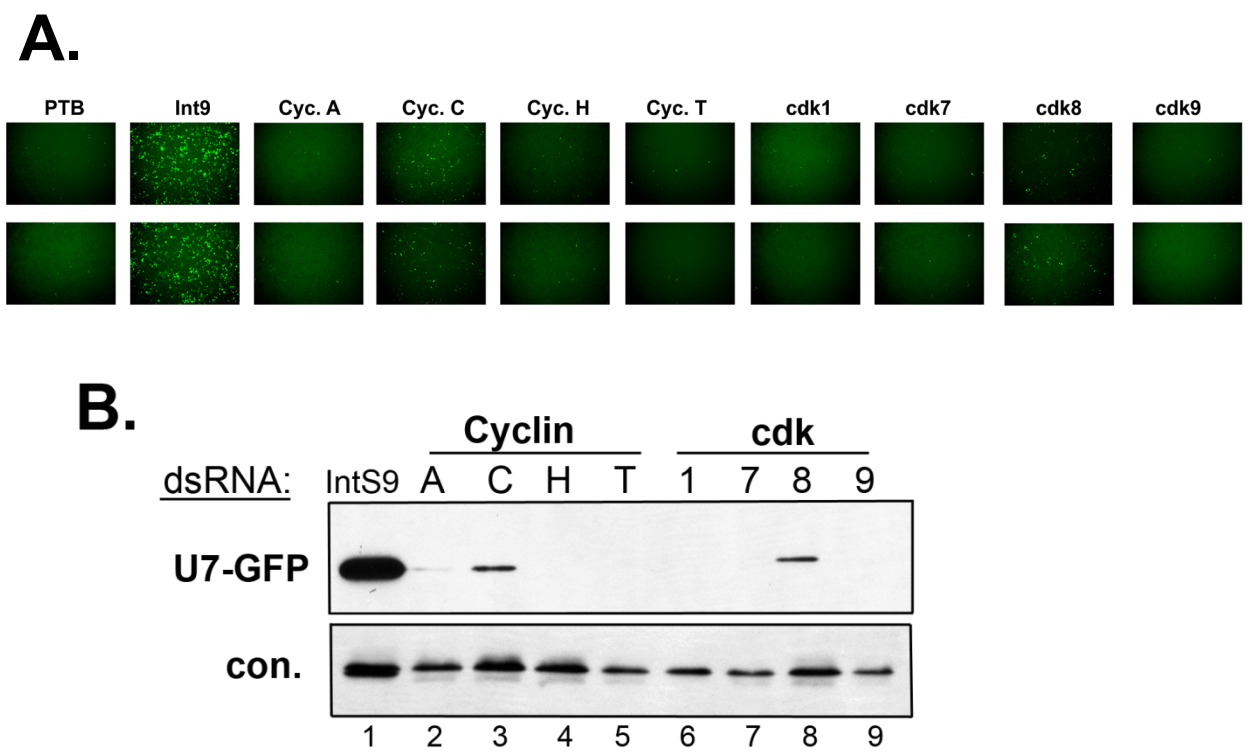


Figure 6.3 Specific involvement of Cdk8 and CycC in snRNA 3' end formation.
(**A**). Fluorescence image of S2 cells treated with cyclin or Cdk dsRNA followed by transient transfection with U7-GFP reporters measuring snRNA 3' end formation. Two representative images were shown for each dsRNA treatment. (**B**). Western blot analysis of cell lysates from panel (A).

as well as occupancy on snRNA gene in vivo, which in turn affected the snRNA 3' end formation (Egloff et al. 2007; Egloff et al. 2010), it would be tempting to test whether CycC/Cdk8 can serve as a Ser7 kinase or a redundant Ser2 kinase besides Cdk9. Indeed, previous study has shown that *Drosophila* Cdk8 is able to phosphorylate RNAPII CTD in vitro (Leclerc et al. 1996), a more specific in vitro kinase assay using recombinant proteins or an in vivo assay to determine the phosphorylation alternations using RNAPII CTD phosphorylation-specific antibodies in CycC/Cdk8 depleted cells will provide more details on CycC/Cdk8 function mechanism in snRNA 3' end formation.

Finally, in addition to the "strong hits" analyzed above, our screen also identified 7 "medium hits", including a WD40 repeat-containing uncharacterized protein CG9945 and a zinc-finger containing protein CG11247. Recent study of the Gemin5, a component of the SMN complex, determined that the WD repeat is a previously undescribed RNA-binding domain that specifically binds snRNA sequence (Lau et al. 2009). Results from recent studies of zinc-finger proteins also showed that they are able to bind single-stranded RNA as well expanding the canonical roles in DNA recognition and protein-protein interaction (Burdach et al. 2012). Therefore, to functionally validate these two proteins and to determine their potential for snRNA recognition and Integrator association may lead to the discovery of the elusive RNA binding proteins in the snRNA 3' end processing machinery.

Structural and functional analysis of the PHD finger-containing Integrator subunit IntS12

It has been well established that chromatin states control gene transcription, and emerging evidence from recent studies suggest that chromatin states also play important roles in exon definition as well as pre-mRNA alternative splicing (Luco et al. 2010; Shukla et al. 2011). In both cases, the epigenetic effectors are critical readers that transmit the chromatin pattern to biological outcomes. While no study has been reported to implicate the chromatin architecture in the 3' end processing of pre-RNA, it is possible that such chromatin reader mediated recognition may be involved. The plant homeodomain (PHD) finger is a reader domain of the chromatin that preferentially recognizes modified or unmodified histone lysine markers (Reviewed in (Musselman and Kutateladze 2011; Sanchez and Zhou 2011)), and it was found to be the most conserved domain with a well-defined function in the Integrator proteins besides the MBL/β-CASP domain present in IntS9/11 (Figure 1.5). It is possible that the Integrator complex is recruited through recognition of specific chromatin markers present at the 5' or 3' end of snRNA genes via the PHD finger present in Integrator subunit 12 (IntS12). This provides the basis for our structural/functional study of IntS12.

Chapter 4 and 5 described our biochemical and functional characterization of IntS12 to understand how this PHD finger-containing Integrator protein functions in snRNA 3' end formation. Since IntS12 contains a chromatin reader domain, we tested the chromatin recruitment of Integrator complex model by determination of the interaction between the IntS12 PHD finger and histones. As expected, GST-pulldown assay established a specific interaction between IntS12 PHD finger and the histone H3 (Figure 5.1B/C). However, subsequent functional analysis clearly demonstrates that neither the conserved IntS12 PHD nor the characterized interaction with histone H3 is required for mediating the U7 or U4 reporter snRNA 3' –end formation (Figure 4.4C/D, Figure 5.5). To our surprise, systemic analysis of IntS12 functional domain identified an N-terminal 30 amino acid long microdomain that resides in a conserved region that is predicted to form a helix-coil-helix fold (Figure 4.4C/D/E, Figure 4.5A, Figure 6.4). This IntS12 microdomain functions autonomously as shown functionally the microdomain by itself is able to mediate both reporter and endogenous snRNA 3' end formation (Figure 4.6, 4.7A). Further, biochemically it is sufficient to stabilize the largest Integrator subunit IntS1 through a potential direct binding (Figure 4.7, 4.8, 4.9). This study provided the first detailed functional analysis of one of the Integrator subunits and established a working model for IntS12 in snRNA 3' end formation where the N-terminal region of IntS12 binds IntS1 to stabilize the putative scaffold protein to facilitate further complex assembly (Figure 6.5). IntS1 has been determined to play important roles in snRNA 3' end formation and animal development. Depletion of IntS1 in S2 cells causes accumulation of significant amount of premature snRNAs (Figure 3.3A) and disruption of IntS1 in mouse causes misprocessing of snRNAs and eventually leads to embryonic lethality (Hata and Nakayama 2007). The expression of IntS1 was found to be dependent on the expression of several other Integrator subunits as shown in our co-depletion assay (Figure 6.6), further supporting a scaffolding role of IntS1 in the Integrator complex. The functional microdomain of IntS12 is predicted to adopt a helix-coil-helix fold that has been shown to serve as a protein-protein dimerization surface (Johnen and Kaufman 1997; Colledge et al. 2011). However, an effort to determine the structure of the IntS12 microdomain by nuclear magnetic resonance (NMR) spectroscopy revealed that solution structure of the recombinant microdomain is likely to be intrinsically disordered (data not shown). One possible explanation for the disparity is that the correct folding for IntS12 microdomain requires the presence of its binding partner (potentially IntS1). It would be interesting to determine which region of IntS1 is recognized by IntS12 microdomain that will further advance our understanding on the function of the IntS12 microdomain in snRNA 3' end formation.

Our results from the study of the IntS12 functional domain disfavors our initial model for chromatin recruitment of Integrator complex to snRNA genes through the conserved PHD

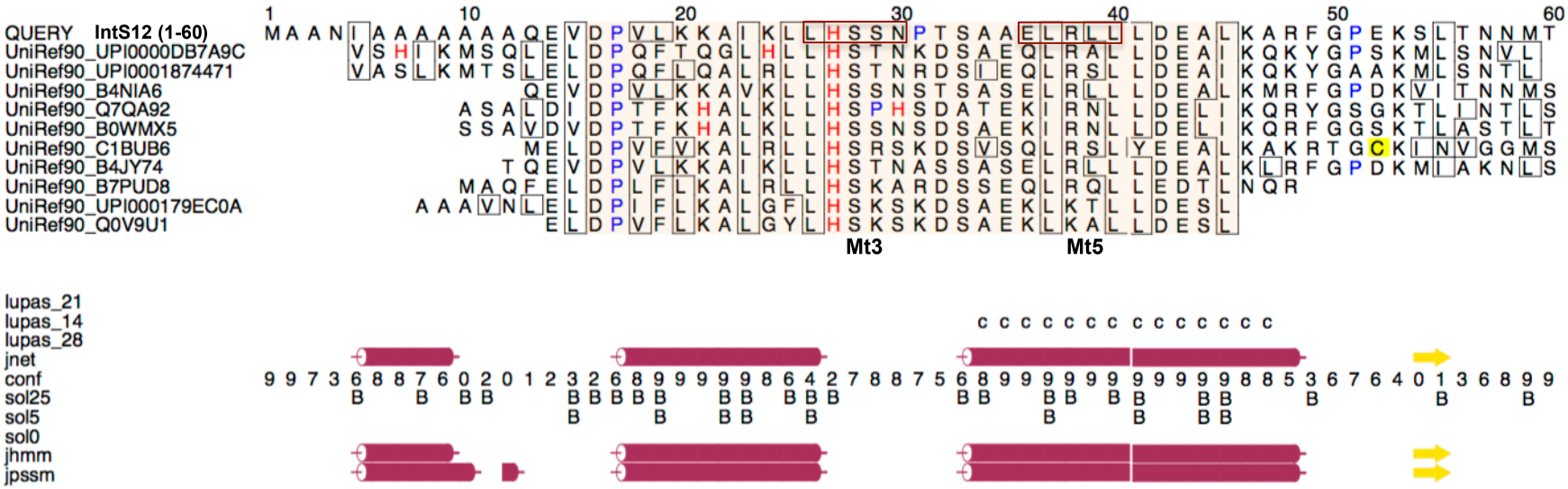


Figure 6.4 Secondary structure prediction of IntS12 microdomain by Jpred3 predicts a helix-coil-helix fold.

Drosophila IntS12 N-terminal 1-60 amino acid sequence was used for secondary structure prediction using the Jpred3 secondary structure prediction server (http://www.compbio.dundee.ac.uk/www-jpred/). The QUERY is the IntS12 (1-60) sequence and UniRef90 items are sequences predicted with closest secondary structure to QUERY sequence. The predicted α-Helix is shown with read stick, and the prediction confidence is right below (0-9, the larger the number, the more confidence). The red shade box represents the identified functional microdomain, and the functional important residues within the microdomain are red-boxed.

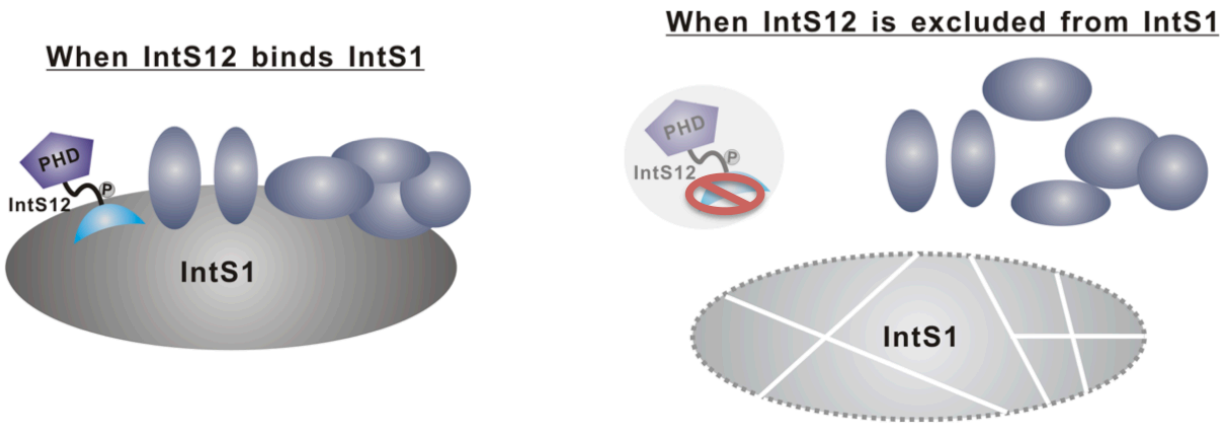


Figure 6.5 Schematic working model of IntS12-IntS1 interaction in Integrator complex assembly.

When IntS12 interacts with putative scaffolding Integrator subunit IntS1, it stabilizes IntS1 protein in cells, which serves as a platform for other Integrator binding and the complex assembly. If this interaction is disrupted by either sequestering IntS12 or promoting its degradation, IntS1 in cells will be destabilized and subject to degradation. Loss of the putative scaffolding protein eventually will disassemble the Integrator complex. IntS12 is schematically shown to have three features: an N-terminal microdomain (crescent), a PHD finger (pentagon) and a phosphorylation at the Thr 76 residue.

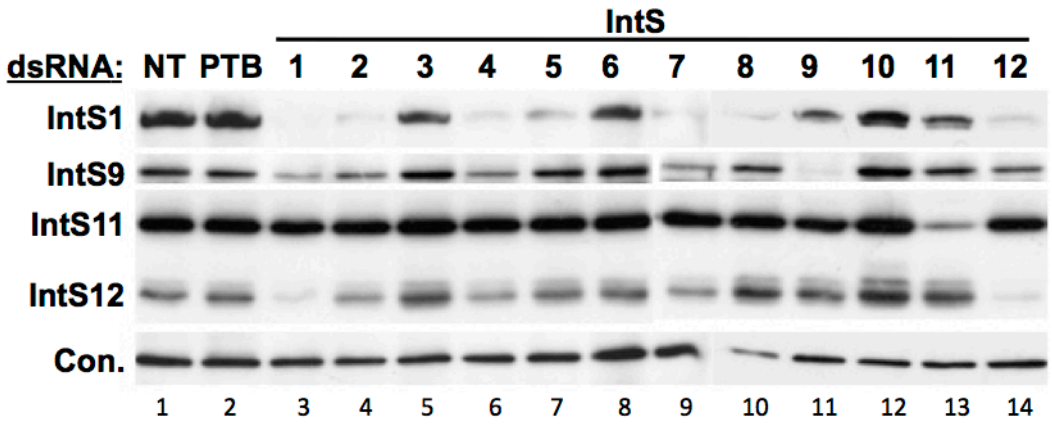


Figure 6.6 The stability of IntS1 in cells is dependent upon the expression of many other Integrator subunits.

S2 cells were treated with dsRNA targeting either control *PTB* mRNA or different *Integrator* mRNA for three days, and cell lysates were prepared from these cells on the fifth day. Lysates were then subjected to Western blot analysis using antibodies specific for IntS1, IntS9, IntS11 and IntS12. NT, cells not treated with any dsRNA; IntS, Integrator subunit; a cross-reacting band with IntS1 antibodies was used as a loading control.

finger containing protein IntS12 to mediate proper 3' end formation. This then generates the question of what is the function of the IntS12 PHD finger? While we did observe that the IntS12 microdomain can function autonomously, it exhibited lower efficiency in interaction with other Integrators (Figure 4.7A/D). So there is a possibility that IntS12 PHD may contribute to full interaction between IntS12 and other Integrators. Our results from the co-IP support this idea as the PHD finger retaining IntS12 N-terminal fragment exhibited significantly stronger interaction with other Integrators to a level equivalent to the full-length IntS12 protein (Figure 5.4B, lane 10 v.s. 2, 4). These results await follow-up experiments to determine whether the observed enhanced Integrator interaction contributes to endogenous snRNA 3' end formation.

Though here we focus our research interest on the Integrator Complex in snRNA 3' end formation, there are also sporadic reports that suggest Integrator subunits function in processes beyond snRNA 3' end formation (Reviewed in (Chen and Wagner 2010)). Recently genome-wide CHIP-Seq studies of the SNAP protein and endonuclease IntS11 suggest that the Integrator complex is potentially recruited to the genomic loci distinct from snRNA genes ((Baillat et al. 2012); Baillat & Shiekhattar, unpublished data). More interestingly, a statistical analysis of *Drosophila* Integrator proteins scored in the public screens also suggests that the Integrator complex may function in a variety of cellular and biological pathways. These pathways include: notch signaling (Mourikis et al. 2010), mitogen-activated protein kinase (MAPK) signaling, protein aggregate formation (Zhang et al. 2010) and cell morphology control (Kiger et al. 2003) (Table 6.1). Thus, we can envision that the conserved IntS12 PHD may contribute to these yet-to-be identified/confirmed cellular and biological processes.

Finally, this thesis project provides a powerful tool for characterization of functional domains present in other Integrator subunits that are required for snRNA 3' end formation. The unique *Drosophila* RNAi rescue system we presented here combined with our snRNA-GFP reporters enables us to ultimately validate the candidate genes and dissect any functional domains present in the validated factors. The Drosophila RNAi system uses ~500bp dsRNAs to elicit depletion of endogenous proteins, which makes it difficult to generate an RNAi-resistant cDNA. This, in turn, makes any RNAi-rescue structure/function analysis intrinsically more challenging. Two common methods to do RNAi rescue in fly cells are: (1) design dsRNA targeting the UTRs of the endogenous mRNA and rescue by the wild-type cDNA open reading frame (ORF) of the targeted gene; (2) dsRNA targeting the coding sequence and rescue with an ortholog cDNA (i.e. some other Drosophilids) of the targeted gene. The limitation for the first method is that many of the *Drosophila* genes have short UTR regions that when targeted elicit limited levels of reduction of endogenous proteins. Significant effort was wasted to determine that this was the case for the IntS12 3'UTR (~70 nt) (Data not shown). The limitation for the second method lies in that it rescues with a different protein, which may function

Table 6.1 Integrator proteins as positive hits in public screens

Gene	Off-Targets	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19
IntS1	1	M			W															M
IntS2	0	M			W			W	W							M	M			
IntS3	0	M	M	W	S					M				M						
IntS4	0							W	W						M					M
IntS5/Omd	1				W			W	W					M			M			
IntS6/DICE1	3	M	M		W			W				M		M						M
IntS7	0	M			S	M			W		M									M
IntS8	0		M		W			W	W											
IntS9	2								W									S		
IntS10	0																			
IntS11	1	M							W											
IntS12	0		M																	
IntS13/Asu	0						M													
IntS14	0						M													
CycC	0										W						M		W	
Cdk8	1	M	M							M		M	M				M		S	
Total		**7**	**5**	**1**	**7**	**1**	**2**	**5**	**7**	**2**	**2**	**2**	**1**	**3**	**1**	**1**	**4**	**1**	**2**	**4**

S1. Modifiers of Notch Signaling; S2. S2R+ Morphology Screen; S3. Functional genomic analysis of the Wnt-Wingless Signaling Pathway; S4. MAPK Signaling Pathway Screen; S5. RTK Specificity; S6. Identification and characterization of proteins required for MicroRNA and siRNA biogenesis and functioning; S7. Host factors important for uptake, survival and multiplication of chlamydia caviae; S8. Modifiers of Intracellular aggregates derived from mis-folded proteins; S9. General Secretion Screen; S10. CRACM1 Is a Plasma Membrane Protein Essential for Store-Operated Ca2+ Entry; S11. Hedghog secretion; S12. Interactions between Legionella pneumophila and its host; S13. Genes regulating lipid storage in Drosophila melanogaster; S14. Comparison of GPCR Induced vs. Small G-Protein Induced Signalling to the Nucleus; S15. Identifying factors that regulate the recruitment and function of the dosage compensation complex in Drosophila melanogaster.; S16. ERK Signaling Dynamics; S17. RNAI screen for genes required for processing of Cubitus interruptus (Ci); S18. Screen for novel regulators of G2/M checkpoint; S19. BCL9 Screen

Table 6.1 Integrator proteins as positive hits in public screens. The data for this statistical analysis is derived from the Drosophila RNAi Screening Center (DRSC). In total 78 genome-wide RNAi screen data were investigated and the screens that identified positive Integrator (including CycC/Cdk8) hits are shown here. In each screen, the hits are scored by Stong (S), Medium (M) or Weak (W). Off-targets here means predicted number of off-targets for each dsRNA used to target Integrator component.

different from the wild-type protein. Here we described another way to do the RNAi rescue in fly cells by using dsRNA targeting the coding sequence while rescue is conducted with an RNAi-resistant cDNA that encodes the same wild-type protein. This RNAi-resistant cDNA is generated by chemical synthesis of a DNA sequence with all of the possible silent mutations introduced to the cDNA region targeted by the dsRNA. This RNAi rescue method in fly cells can circumvent the obstacles generated by using the other two methods, and theoretically it is applicable to any gene in fly cells. With the development of synthetic biology, the cost using this technique will continue to decrease.

CONCLUDING REMARKS

This dissertation has contributed to the field of metazoan snRNA processing by providing detailed functional analysis of the snRNA 3' end processing machinery, the Integrator Complex. Specifically, this work has redefined the Integrator complex to comprise of fourteen core subunits and a regulatory kinase through a genome-wide functional RNAi screen. Moreover, this work also provides the first detailed functional domain analysis of one Integrator subunit IntS12, and reveals IntS12 mediates the function of Integrator complex in snRNA 3' end formation through stabilization of the putative scaffold protein in the complex.

Finally, the work presented in dissertation provides the basis for further elucidating the biochemical, cellular and biological function of the conserved IntS12 PHD finger, and for identification of those elusive RNA-binding proteins in the Integrator complex for snRNA sequence element recognition. The research into the Integrator Complex is only beginning and this thesis has contributed to laying foundation for future work.

Appendix. IntS12 is Preferentially Recruited to Promoter Region of snRNA Genes.

INTRODUCTION

Integrator Complex is the 3' end processing machinery for RNAPII-transcribed snRNAs in metazoan species that associates with the CTD of RNAPII, and IntS11 is the MBL/β-CASP superfamily endonuclease that cleaves 3' end of the nascent primary snRNA transcripts from the elongating polymerase (Baillat et al. 2005). Subsequent functional analysis of the Integrator complex has determined that all twelve subunits except IntS3 and IntS10, are required for snRNA 3' end formation in *Drosophila* cells (Ezzeddine et al. 2011). Several Integrator subunits (IntS2, 9,10,11) have also been determined to recruit to U1 and U2 snRNA genes but not the Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) encoding gene or the histone H3 gene (Baillat et al. 2005), supporting a specific functional requirement for Integrator Complex in snRNA biosynthesis. For the other two RNAPII transcripts, the poly(A) mRNA and histone mRNA, the shared cleavage factor including CPSF73, CPSF100 and Symplekin, is found to enrich at 3' end of both types of genes *in vivo* (Sullivan et al. 2009), supporting its cleavage role for both types of transcripts.

It has well documented that snRNA 3' end formation is a transcription-coupled process that requires transcription initiation from an snRNA promoter (de Vegvar et al. 1986; Hernandez and Weiner 1986; Ezzeddine et al. 2011). This is analogous to the poly(A) mRNA 3' end cleavage/polyadenylation process but is distinct from the histone mRNA 3' end formation. The pre-mRNA 3' end processing factors have been observed to be present at both 5' and 3' end of coding genes (He et al. 2003; Glover-Cutter et al. 2008), supporting the model for recruitment of processing factors to the promoter for coupling transcription and 3' end processing. Compared to mRNA encoding gene, the snRNA gene is very short with an average size of ~ 200 bp, which makes it difficult to discriminate the occupancy of Integrator proteins on the snRNA genomic loci by the Chromatin immunoprecipitations (CHIP) assay. The original study showed that four Integrator subunits tested occupied both ends of U1 and U2 snRNA genes in HEK293T cells (Baillat et al. 2005). However, the author described that due to the low resolution of CHIP, it was difficult to conclude the exact occupying loci of the Integrator proteins on snRNA genes.

Here I tried to develop a high-resolution CHIP assay for detection of Integrator proteins on snRNA genes in *Drosophila* S2 cells. This assay could be used to determine the accurate occupancy of Integrator proteins on snRNA genes (Promoter or 3' region), and combined with RNAi, it could be used to determine critical subunit for recruitment of Integrator complex to snRNA promoter or 3' end region, which will significantly advance our understanding for Integrator Complex function in snRNA 3' end formation. I started optimizing the CHIP conditions by study of the IntS12 occupancy on snRNA genes (U7, U1, U2, U4). I determined

the optimal chromatin fragmentation condition and screened good qPCR amplicons of various regions of the snRNA genes in S2 cells. By using the developed CHIP assay, I have found that IntS12 is preferentially recruited to the promoter region of U7 snRNA gene whereas the RNAPII is enriched across the whole gene region. Then I also tested the requirement of the conserved PHD and the T76 phosphorylation for IntS12 occupancy on promoter region of the U7 snRNA gene.

RESULTS

Optimization of Chromatin Immunoprecipitation (CHIP) Conditions in S2 Cells

A high resolution CHIP profile is important for determining the in vivo genomic occupancy of a protein under investigation, and several key factors are prerequisites to achieve that, including the qualities of CHIP antibodies, chromatin fragments, qPCR amplicons. Figure A.1A shows the standard procedure for CHIP assay, and the detailed procedure is described in the Chapter 2 Materials and Methods section. Here I evaluated all the factors required for development of a high resolution CHIP on snRNA genes. I purified the IntS12 polyclonal antibodies by affinity purification using the recombinant GST-IntS12 (1-163 a.a.) antigen-conjugated column, and showed that the antigen-purified IntS12 polyclonal antibodies can specifically recognize endogenous IntS12 proteins in S2 cells (Figure A.1B). The size of the fragmented chromatin is a critical determinant of the resolution of the CHIP assay. I tested different sonication conditions and determined the optimal program to be 4 times of 30 sec on/2.5 min off using the Ultrasonic Converter C5749. Under this condition, the major population of the chromatin fragments was distributed ~ 300-400 bp (Figure A.1C). I also tested qPCR primers for amplification of different regions upstream and downstream 1.6kb of the U7 snRNA gene (Figure A.1D), and determined seven amplicons (~200 bp) that could be specifically amplified across the U7 genomic loci (Figure A.1D). Collectively, I determined the optimal CHIP conditions for test IntS12 occupancy on U7 snRNA genes in S2 cells by controlling the quality of the antibodies, chromatin fragment size and amplicons.

IntS12 is preferentially occupying the U7 snRNA promoter region

Once optimized the CHIP condition in S2 cells, I tested the occupancy of IntS12 on U7 snRNA genomic loci. $1X10^8$ S2 cells were used for the CHIP assay. DNA-proteins were crosslinked, sonicated, immunoprecipitated, and precipitated DNA-protein were reverse crosslinked and DNA was purified for qPCR analysis. Antibodies for RNAPII, IntS11 and IntS12 were used for immunoprecipitation, and Guinea pig (GP) IgG were used as control for pulldown. The RNAPII was observed to enrich in both the promoter and 3' region of the U7 snRNA gene (Figure A.2d,e), and IntS12 was highly enriched in the U7 snRNA promoter (Figure A.2d). In contrast, the endonuclease IntS11 was found to be present across the whole region measured, with the highest occupancy in the 3' end region of the U7 snRNA gene (Figure A.2). This result indicates that there might be a sequential recruitment of different Integrator subunit onto specific region of the U7 snRNA gene.

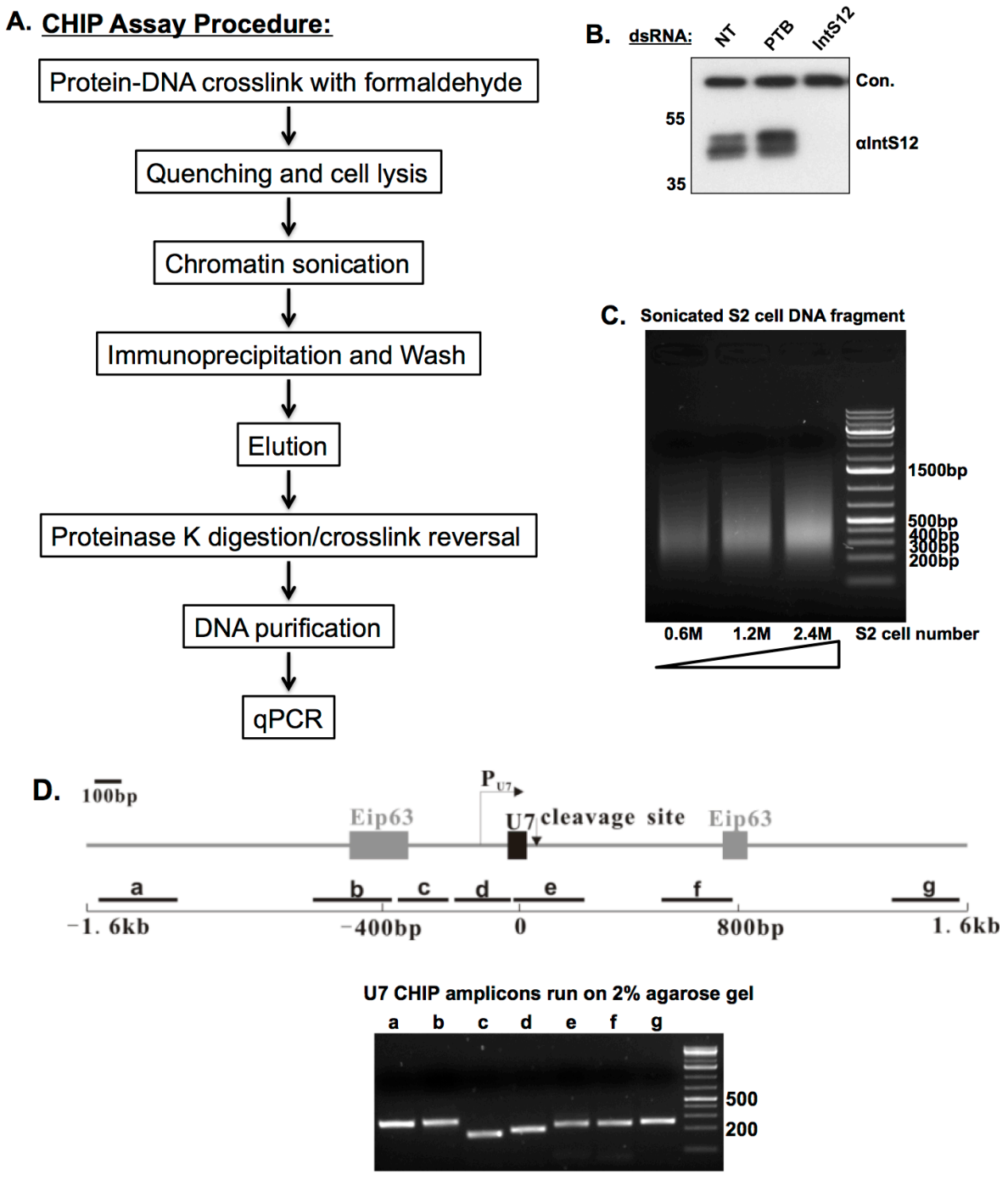


Figure A.1 Optimization of CHIP conditions on *Drosophila* U7 snRNA gene.

(**A**). Schematic representation of the standard procedure for CHIP assay. (**B**). Western blot analysis of S2 cell lysates using the antigen-purified IntS12 polyclonal antibodies. (**C**). Reverse crosslinked total DNA was run on agarose gel and stained with Ethidium bromide. (**D**). top panel, schematic representation of amplicon location relative to the U7 snRNA gene; bottom panel, amplicons from qPCR were run on agarose gel and stained with Ethidium bromide.

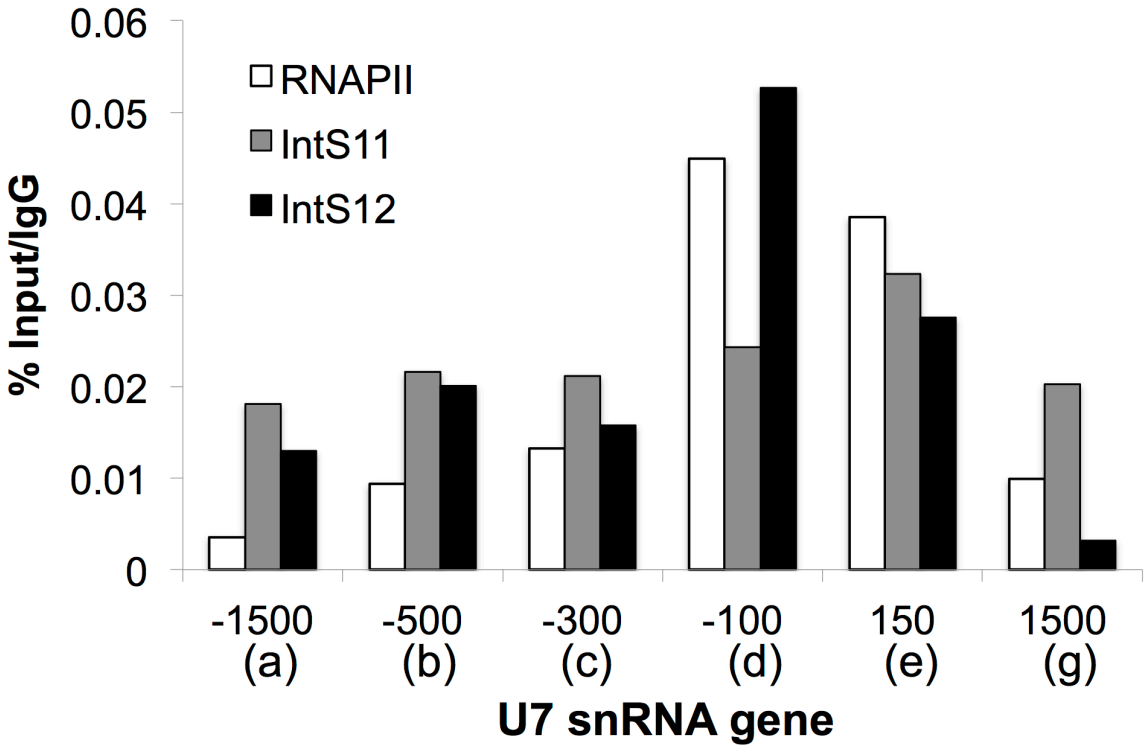


Figure A.2 CHIP profiles of IntS11, IntS12 and RNAPII on the U7 snRNA gene.
S2 cell CHIP DNA was prepared by following the optimized protocol described above, and RNAPII, IntS11 and IntS12 antibodies were used for immunoprecipitations with GP IgG as negative control. Purified DNA was used for qPCR analysis using primers described above. Data were presented as percentage of input with the IgG background subtracted.

DISSCUSSION

Chromatin Immunoprecipitation is a technique used to determine the occupancy of a protein factor on specific genomic region in vivo. Information garnered from CHIP assay is an important complementation for understanding the cellular function of nuclear acid associated proteins. Integrator proteins were found to recruit to U1 and U2 snRNA genes and catalyze the 3' end formation of RNAPII-transcribed snRNAs. However, it is still not clear whether it is recruited to the promoter or the 3' end region of snRNA. Both cases are possible as Integrator complex are determined to couple the snRNA transcription and 3' end formation as well as cleave the nascent primary transcripts (de Vegvar et al. 1986; Hernandez and Weiner 1986; Baillat et al. 2005). A high resolution CHIP assay will help answering these biologically important questions. In this study, I developed a CHIP assay for detection of the occupancy of IntS12 protein on U7 snRNA gene in vivo. The IntS12 proteins were found to preferentially occupy the promoter region of the U7 snRNA gene.

In Chapter 4 and 5, we have comprehensively investigated the domain functions of IntS12 protein, and characterized an N-terminal microdomain that is essential for snRNA 3' end formation. We also determined the roles of the T76 phosphorylation and the conserved PHD finger domain of IntS12 in Integrator interactions. It would be interesting to know whether these characterized features are important for IntS12 occupancy on the U7 snRNA promoter or not. To test that, we have generated S2 cell stable lines expressing FLAG-tagged PHD deletion or point mutants, T76 point mutant or the N-terminal microdomain point mutants. CHIP analysis using anti-FLAG antibodies in these stable cell lines gave ambiguous results that are hard to interpret (data not shown). One potential caveat to do the CHIP analysis using these stable lines is the uneven expression of tagged proteins in these cells, which makes the CHIP signal hard to compare between different cell lines. To circumvent this problem, it will be worth selecting the monoclonal S2 stable cell lines that give similar levels of expression for the tagged proteins or to generate isogenic cells that harbor only a single copy of gene insertion at a specific genomic site.

Nevertheless, using the CHIP assay we developed on snRNA genes combined with RNAi-mediated depletion of Integrator subunits, it will provide more details of Integrator function in snRNA 3' end formation, such as the promoter recruitment of Integrator complex and the 3' end cleavage of snRNA, through study of the occupancy of all other Integrator subunits on snRNA promoter and 3' end region.

REFERENCE

Ach RA, Weiner AM. 1987. The highly conserved U small nuclear RNA 3'-end formation signal is quite tolerant to mutation. *Molecular and cellular biology* **7**: 2070-2079.

Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ. 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* **34**: 387-393.

Akoulitchev S, Chuikov S, Reinberg D. 2000. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature* **407**: 102-106.

Albrecht TR, Wagner EJ. 2012. snRNA 3' end formation requires heterodimeric association of integrator subunits. *Molecular and cellular biology* **32**: 1112-1123.

Anderson MA, Jodoin JN, Lee E, Hales KG, Hays TS, Lee LA. 2009. Asunder is a critical regulator of dynein-dynactin localization during Drosophila spermatogenesis. *Mol Biol Cell* **20**: 2709-2721.

Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P. 2001. Comparison of ARM and HEAT protein repeats. *Journal of molecular biology* **309**: 1-18.

Azzouz TN, Gruber A, Schumperli D. 2005. U7 snRNP-specific Lsm11 protein: dual binding contacts with the 100 kDa zinc finger processing factor (ZFP100) and a ZFP100-independent function in histone RNA 3' end processing. *Nucleic acids research* **33**: 2106-2117.

Baillat D, Gardini A, Cesaroni M, Shiekhattar R. 2012. Requirement for SNAPC1 in Transcriptional Responsiveness to Diverse Extracellular Signals. *Molecular and cellular biology* **32**: 4642-4650.

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

Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. 2002. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99**: 9697-9702.

Bernstein JA, Lin PH, Cohen SN, Lin-Chao S. 2004. Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proc Natl Acad Sci U S A* **101**: 2758-2763.

Bienz M. 2006. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem Sci* **31**: 35-40.

Boyd DC, Greger IH, Murphy S. 2000. In vivo footprinting studies suggest a role for chromatin in transcription of the human 7SK gene. *Gene* **247**: 33-44.

Brown KM, Gilmartin GM. 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* **12**: 1467-1476.

Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT. 2010. Protein annotation and modelling servers at University College London. *Nucleic acids research* **38**: W563-568.

Buratowski S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**: 541-546.

Burch BD, Godfrey AC, Gasdaska PY, Salzler HR, Duronio RJ, Marzluff WF, Dominski Z. 2011. Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila. *Rna* **17**: 1132-1147.

Burdach J, O'Connell MR, Mackay JP, Crossley M. 2012. Two-timing zinc finger transcription factors liaising with RNA. *Trends Biochem Sci* **37**: 199-205.

Burghes AH, Beattie CE. 2009. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nature reviews Neuroscience* **10**: 597-609.

Carpousis AJ. 2007. The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol* **61**: 71-87.

Cazalla D, Xie M, Steitz JA. 2011. A primate herpesvirus uses the integrator complex to generate viral microRNAs. *Mol Cell* **43**: 982-992.

Chandran V, Luisi BF. 2006. Recognition of enolase in the Escherichia coli RNA degradosome. *Journal of molecular biology* **358**: 8-15.

Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, Eick D. 2007. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**: 1780-1782.

Chen J, Ezzeddine N, Waltenspiel B, Albrecht TR, Warren WD, Marzluff WF, Wagner EJ. 2012. An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *Rna*.

Chen J, Wagner EJ. 2010. snRNA 3' end formation: the dawn of the Integrator complex. *Biochemical Society transactions* **38**: 1082-1087.

Coady TH, Lorson CL. 2011. SMN in spinal muscular atrophy and snRNP biogenesis. *Wiley interdisciplinary reviews RNA* **2**: 546-564.

Colledge VL, Fogg MJ, Levdikov VM, Leech A, Dodson EJ, Wilkinson AJ. 2011. Structure and organisation of SinR, the master regulator of biofilm formation in Bacillus subtilis. *Journal of molecular biology* **411**: 597-613.

Cotta-Ramusino C, McDonald ER, 3rd, Hurov K, Sowa ME, Harper JW, Elledge SJ. 2011. A DNA damage response screen identifies RHINO, a 9-1-1 and TopBP1 interacting protein required for ATR signaling. *Science* **332**: 1313-1317.

Cuello P, Boyd DC, Dye MJ, Proudfoot NJ, Murphy S. 1999. Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *Embo J* **18**: 2867-2877.

de Vegvar HE, Lund E, Dahlberg JE. 1986. 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell* **47**: 259-266.

Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic acids research* **11**: 1475-1489.

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

-. 2010. The hunt for the 3' endonuclease. *Wiley interdisciplinary reviews RNA* **1**: 325-340.

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

Dominski Z, Marzluff WF. 2007. Formation of the 3' end of histone mRNA: getting closer to the end. *Gene* **396**: 373-390.

Dominski Z, Yang XC, Marzluff WF. 2005a. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123**: 37-48.

Dominski Z, Yang XC, Purdy M, Marzluff WF. 2003. Cloning and characterization of the Drosophila U7 small nuclear RNA. *Proc Natl Acad Sci U S A* **100**: 9422-9427.

Dominski Z, Yang XC, Purdy M, Wagner EJ, Marzluff WF. 2005b. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Molecular and cellular biology* **25**: 1489-1500.

Dominski Z, Zheng LX, Sanchez R, Marzluff WF. 1999. Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. *Molecular and cellular biology* **19**: 3561-3570.

Egloff S, Al-Rawaf H, O'Reilly D, Murphy S. 2009. Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. *Molecular and cellular biology* **29**: 4002-4013.

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

Egloff S, O'Reilly D, Murphy S. 2008. Expression of human snRNA genes from beginning to end. *Biochemical Society transactions* **36**: 590-594.

Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S. 2010. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem* **285**: 20564-20569.

Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. 2012. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell* **45**: 111-122.

Ezzeddine N, Chen J, Waltenspiel B, Burch B, Albrecht T, Zhuo M, Warren WD, Marzluff WF, Wagner EJ. 2011. A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation. *Molecular and cellular biology* **31**: 328-341.

Fair K, Anderson M, Bulanova E, Mi H, Tropschug M, Diaz MO. 2001. Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. *Molecular and cellular biology* **21**: 3589-3597.

Filleur S, Hirsch J, Wille A, Schon M, Sell C, Shearer MH, Nelius T, Wieland I. 2009. INTS6/DICE1 inhibits growth of human androgen-independent prostate cancer cells by altering the cell cycle profile and Wnt signaling. *Cancer Cell Int* **9**: 28.

Gabanella F, Butchbach ME, Saieva L, Carissimi C, Burghes AH, Pellizzoni L. 2007. Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. *PloS one* **2**: e921.

Galbraith MD, Donner AJ, Espinosa JM. 2010. CDK8: a positive regulator of transcription. *Transcription* **1**: 4-12.

Glover-Cutter K, Kim S, Espinosa J, Bentley DL. 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature structural & molecular biology* **15**: 71-78.

Han SM, Lee TH, Mun JY, Kim MJ, Kritikou EA, Lee SJ, Han SS, Hengartner MO, Koo HS. 2006. Deleted in cancer 1 (DICE1) is an essential protein controlling the topology of the inner mitochondrial membrane in C. elegans. *Development* **133**: 3597-3606.

Hata T, Nakayama M. 2007. Targeted disruption of the murine large nuclear KIAA1440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death. *Biochim Biophys Acta* **1773**: 1039-1051.

He X, Khan AU, Cheng H, Pappas DL, Jr., Hampsey M, Moore CL. 2003. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes Dev* **17**: 1030-1042.

Hernandez G, Jr., Valafar F, Stumph WE. 2007. Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. *Nucleic acids research* **35**: 21-34.

Hernandez N. 1985. Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *Embo J* **4**: 1827-1837.

-. 2001. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem* **276**: 26733-26736.

Hernandez N, Weiner AM. 1986. Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell* **47**: 249-258.

Hom RA, Chang PY, Roy S, Musselman CA, Glass KC, Selezneva AI, Gozani O, Ismagilov RF, Cleary ML, Kutateladze TG. 2010. Molecular mechanism of MLL PHD3 and RNA recognition by the Cyp33 RRM domain. *Journal of molecular biology* **400**: 145-154.

Huang J, Gong Z, Ghosal G, Chen J. 2009. SOSS complexes participate in the maintenance of genomic stability. *Mol Cell* **35**: 384-393.

Hung KH, Stumph WE. 2011. Regulation of snRNA gene expression by the Drosophila melanogaster small nuclear RNA activating protein complex (DmSNAPc). *Crit Rev Biochem Mol Biol* **46**: 11-26.

Jacobs EY, Ogiwara I, Weiner AM. 2004. Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. *Molecular and cellular biology* **24**: 846-855.

Johnen G, Kaufman S. 1997. Studies on the enzymatic and transcriptional activity of the dimerization cofactor for hepatocyte nuclear factor 1. *Proc Natl Acad Sci U S A* **94**: 13469-13474.

Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* **292**: 195-202.

Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**: 664-668.

Kiger AA, Baum B, Jones S, Jones MR, Coulson A, Echeverri C, Perrimon N. 2003. A functional genomic analysis of cell morphology using RNA interference. *Journal of biology* **2**: 27.

Kitao S, Segref A, Kast J, Wilm M, Mattaj IW, Ohno M. 2008. A compartmentalized phosphorylation/dephosphorylation system that regulates U snRNA export from the nucleus. *Molecular and cellular biology* **28**: 487-497.

Knuesel MT, Meyer KD, Donner AJ, Espinosa JM, Taatjes DJ. 2009. The human CDK8 subcomplex is a histone kinase that requires Med12 for activity and can function independently of mediator. *Molecular and cellular biology* **29**: 650-661.

Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, Kimura H, Proudfoot NJ, Akoulitchev A. 2002. U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nature structural biology* **9**: 800-805.

Lau CK, Bachorik JL, Dreyfuss G. 2009. Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. *Nature structural & molecular biology* **16**: 486-491.

Leclerc V, Tassan JP, O'Farrell PH, Nigg EA, Leopold P. 1996. Drosophila Cdk8, a kinase partner of cyclin C that interacts with the large subunit of RNA polymerase II. *Mol Biol Cell* **7**: 505-513.

Li Y, Bolderson E, Kumar R, Muniandy PA, Xue Y, Richard DJ, Seidman M, Pandita TK, Khanna KK, Wang W. 2009. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem* **284**: 23525-23531.

Loncle N, Boube M, Joulia L, Boschiero C, Werner M, Cribbs DL, Bourbon HM. 2007. Distinct roles for Mediator Cdk8 module subunits in Drosophila development. *Embo J* **26**: 1045-1054.

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**: 996-1000.

MacDonald CC, Wilusz J, Shenk T. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Molecular and cellular biology* **14**: 6647-6654.

Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, Ding C, Shi Y, Yucer N, Krenciute G et al. 2011. Analysis of the human endogenous coregulator complexome. *Cell* **145**: 787-799.

Malovannaya A, Li Y, Bulynko Y, Jung SY, Wang Y, Lanz RB, O'Malley BW, Qin J. 2010. Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc Natl Acad Sci U S A* **107**: 2431-2436.

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**: 953-956.

Marzluff WF, Wagner EJ, Duronio RJ. 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9**: 843-854.

Matera AG, Terns RM, Terns MP. 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8**: 209-220.

McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404-405.

Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, Murphy S. 2005. P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *Embo J* **24**: 4154-4165.

Medlin JE, Uguen P, Taylor A, Bentley DL, Murphy S. 2003. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *Embo J* **22**: 925-934.

Miller TC, Rutherford TJ, Johnson CM, Fiedler M, Bienz M. 2010. Allosteric remodelling of the histone H3 binding pocket in the Pygo2 PHD finger triggered by its binding to the B9L/BCL9 co-factor. *Journal of molecular biology* **401**: 969-984.

Mouaikel J, Verheggen C, Bertrand E, Tazi J, Bordonne R. 2002. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol Cell* **9**: 891-901.

Mourikis P, Lake RJ, Firnhaber CB, DeDecker BS. 2010. Modifiers of notch transcriptional activity identified by genome-wide RNAi. *BMC developmental biology* **10**: 107.

Mowry KL, Oh R, Steitz JA. 1989. Each of the conserved sequence elements flanking the cleavage site of mammalian histone pre-mRNAs has a distinct role in the 3'-end processing reaction. *Molecular and cellular biology* **9**: 3105-3108.

Mowry KL, Steitz JA. 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.

Murthy KG, Manley JL. 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* **9**: 2672-2683.

Musselman CA, Kutateladze TG. 2011. Handpicking epigenetic marks with PHD fingers. *Nucleic acids research* **39**: 9061-9071.

Nurmohamed S, McKay AR, Robinson CV, Luisi BF. 2010. Molecular recognition between Escherichia coli enolase and ribonuclease E. *Acta Crystallogr D Biol Crystallogr* **66**: 1036-1040.

Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW. 2000. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**: 187-198.

Org T, Chignola F, Hetenyi C, Gaetani M, Rebane A, Liiv I, Maran U, Mollica L, Bottomley MJ, Musco G et al. 2008. The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO reports* **9**: 370-376.

Paige JS, Wu KY, Jaffrey SR. 2011. RNA mimics of green fluorescent protein. *Science* **333**: 642-646.

Patel SB, Bellini M. 2008. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic acids research* **36**: 6482-6493.

Pavelitz T, Bailey AD, Elco CP, Weiner AM. 2008. Human U2 snRNA genes exhibit a persistently open transcriptional state and promoter disassembly at metaphase. *Molecular and cellular biology* **28**: 3573-3588.

Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG. 2006. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**: 100-103.

Phatnani HP, Greenleaf AL. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20**: 2922-2936.

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

Pillai RS, Will CL, Luhrmann R, Schumperli D, 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.

Py B, Higgins CF, Krisch HM, Carpousis AJ. 1996. A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature* **381**: 169-172.

Rickert P, Corden JL, Lees E. 1999. Cyclin C/CDK8 and cyclin H/CDK7/p36 are biochemically distinct CTD kinases. *Oncogene* **18**: 1093-1102.

Rogers SL, Rogers GC. 2008. Culture of Drosophila S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. *Nature protocols* **3**: 606-611.

Ropke A, Buhtz P, Bohm M, Seger J, Wieland I, Allhoff EP, Wieacker PF. 2005. Promoter CpG hypermethylation and downregulation of DICE1 expression in prostate cancer. *Oncogene* **24**: 6667-6675.

Rutkowski RJ, Warren WD. 2009. Phenotypic analysis of deflated/Ints7 function in Drosophila development. *Dev Dyn* **238**: 1131-1139.

Sanchez R, Zhou MM. 2011. The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**: 364-372.

Scharl EC, Steitz JA. 1994. The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. *EMBO J* **13**: 2432-2440.

Schramm L, Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16**: 2593-2620.

Segref A, Mattaj IW, Ohno M. 2001. The evolutionarily conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export. *Rna* **7**: 351-360.

Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Pena P, Lan F, Kaadige MR et al. 2006. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**: 96-99.

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**: 74-79.

Skaar JR, Richard DJ, Saraf A, Toschi A, Bolderson E, Florens L, Washburn MP, Khanna KK, Pagano M. 2009. INTS3 controls the hSSB1-mediated DNA damage response. *J Cell Biol* **187**: 25-32.

Smith ER, Lin C, Garrett AS, Thornton J, Mohaghegh N, Hu D, Jackson J, Saraf A, Swanson SK, Seidel C et al. 2011. The little elongation complex regulates small nuclear RNA transcription. *Mol Cell* **44**: 954-965.

Stebbings L, Grimes BR, Bownes M. 1998. A testis-specifically expressed gene is embedded within a cluster of maternally expressed genes at 89B in Drosophila melanogaster. *Dev Genes Evol* **208**: 523-530.

Stunkel W, Kober I, Seifart KH. 1997. A nucleosome positioned in the distal promoter region activates transcription of the human U6 gene. *Molecular and cellular biology* **17**: 4397-4405.

Sullivan KD, Steiniger M, Marzluff WF. 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.

Takata H, Nishijima H, Maeshima K, Shibahara K. 2012. The integrator complex is required for integrity of Cajal bodies. *Journal of cell science* **125**: 166-175.

Tao S, Cai Y, Sampath K. 2009. The Integrator subunits function in hematopoiesis by modulating Smad/BMP signaling. *Development* **136**: 2757-2765.

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

Wagner EJ, Burch BD, Godfrey AC, Salzler HR, Duronio RJ, Marzluff WF. 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.

Wang ZF, Whitfield ML, Ingledue TC, 3rd, Dominski Z, Marzluff WF. 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.

Weiner AM. 2005. E Pluribus Unum: 3' end formation of polyadenylated mRNAs, histone mRNAs, and U snRNAs. *Mol Cell* **20**: 168-170.

West S. 2012. The increasing functional repertoire of U1 snRNA. *Biochemical Society transactions* **40**: 846-849.

Whittaker CA, Hynes RO. 2002. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell* **13**: 3369-3387.

Wieland I, Arden KC, Michels D, Klein-Hitpass L, Bohm M, Viars CS, Weidle UH. 1999. Isolation of DICE1: a gene frequently affected by LOH and downregulated in lung carcinomas. *Oncogene* **18**: 4530-4537.

Will CL, Luhrmann R. 2005. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biological chemistry* **386**: 713-724.

-. 2011. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* **3**.

Workman E, Kolb SJ, Battle DJ. 2012. Spliceosomal small nuclear ribonucleoprotein biogenesis defects and motor neuron selectivity in spinal muscular atrophy. *Brain research* **1462**: 93-99.

Xiang K, Manley JL, Tong L. 2012. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. *Nat Commun* **3**: 946.

Xue Y, Ren J, Gao X, Jin C, Wen L, Yao X. 2008. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Molecular & cellular proteomics : MCP* **7**: 1598-1608.

Yang XC, Burch BD, Yan Y, Marzluff WF, 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 XC, Sabath I, Debski J, Kaus-Drobek M, Dadlez M, Marzluff WF, Dominski Z. 2012. 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. *Molecular and cellular biology*.

Yong J, Kasim M, Bachorik JL, Wan L, Dreyfuss G. 2010. Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol Cell* **38**: 551-562.

Zhang S, Binari R, Zhou R, Perrimon N. 2010. A genomewide RNA interference screen for modifiers of aggregates formation by mutant Huntingtin in Drosophila. *Genetics* **184**: 1165-1179.

Zhang Z, Fu J, Gilmour DS. 2005. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev* **19**: 1572-1580.

Zhang Z, Gilmour DS. 2006. Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell* **21**: 65-74.

Zhang Z, Lotti F, Dittmar K, Younis I, Wan L, Kasim M, Dreyfuss G. 2008. SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133**: 585-600.

Zhao X, Pendergrast PS, Hernandez N. 2001. A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain. *Mol Cell* **7**: 539-549.

Zhou R, Hotta I, Denli AM, Hong P, Perrimon N, Hannon GJ. 2008. Comparative analysis of argonaute-dependent small RNA pathways in Drosophila. *Mol Cell* **32**: 592-599.

VITA

Jiandong Chen was born in Suzhou (Jiangsu, China) on Monday, December 28, 1981, the son of Yingen Chen and Fumei Xi. After graduating from Suzhou Experimental High School in 2001, he enrolled in Nanjing Agricultural University (Nanjing, China) where he received a Bachelor of Science in Biotechnology (National Base) and a Master of Science in Microbiology, graduating with honor in June 2006 and 2008 respectively. In August of 2008, he entered the Ph.D. program in the University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences (Houston, Texas). In May 2009, he joined the laboratory of Dr. Eric John Wagner for his Ph.D. training.

Permanent Address:

6 Tongwei Road, Unit 30 Room 404, Nanjing, Jiangsu, China, 210095.

童卫路6号，南京农业大学家属区30幢404室，南京，江苏，210095