Published in final edited form as:

Biochem Soc Trans. 2010 August ; 38(4): 1082–1087. doi:10.1042/BST0381082.

snRNA 3' end formation: the dawn of the Integrator complex

Jiandong Chen and Eric J. Wagner¹

Department of Biochemistry and Molecular Biology, University of Texas Health Science Center at Houston, Houston, TX 77030, U.S.A., and Graduate School of Biomedical Sciences, University of Texas at Houston, Houston TX 77030, U.S.A

Abstract

The ubiquitously expressed uridine-rich snRNAs (small nuclear RNAs) are essential for the removal of introns, proper expression of histone mRNA and biosynthesis of ribosomal RNA. Much is known about their assembly into snRNP (small nuclear ribonucleoprotein) particles and their ultimate function in the expression of other genes; however, in comparison, less is known about the biosynthesis of these critical non-coding RNAs. The sequence elements necessary for 3' end formation of snRNAs have been identified and, intriguingly, the processing of snRNAs is uniquely dependent on the snRNA promoter, indicating that co-transcriptional processing is important. However, the *trans*-acting RNA-processing factors that mediate snRNA processing remained elusive, hindering overall progress. Recently, the factors involved in this process were biochemically purified, and designated the Integrator complex. Since their initial discovery, Integrator proteins have been implicated not only in the production of snRNA, but also in other cellular processes that may be independent of snRNA biogenesis. In the present study, we discuss snRNA biosynthesis and the roles of Integrator proteins. We compare models of 3' end formation for different classes of RNA polymerase II transcripts and formulate/propose a model of Integrator function in snRNA biogenesis.

Keywords

cleavage and polyadenylation specificity factor (CPSF); Integrator complex; β -lactamase; RNA cleavage; RNA polymerase II; small nuclear RNA (snRNA)

Organization and transcription of snRNA (small nuclear RNA) genes

Human and *Drosophila* snRNA genes, transcribed by RNAPII (RNA polymerase II), are highly expressed throughout development and the cell cycle of all cells. Similarly to histone genes, the RNAPII-driven snRNA genes are present in multiple copies within genomes and are typically clustered [1–3]. They have a relatively uncomplicated gene structure, having no TATA box, no introns and no polyadenylation, which is consistent with the lack of an open reading frame [4]. The promoter contains two elements: a DSE (distal sequence element), which behaves like an enhancer; and a PSE (proximal sequence element) that is essential for snRNA transcription [5]. The 3' end of snRNA genes contains a 3'-box sequence element required for proper 3' end formation located 9–19 nt downstream of the snRNA-coding region [6,7].

The mechanism of transcription initiation at snRNA genes and the phosphorylation pattern of the RNAPII CTD (C-terminal domain) is different from mRNA-encoding genes (for more

comprehensive reviews, see [4,8]). Initiation is mediated by a complex of proteins specific to snRNA genes that is referred to by various names: SNAPc (snRNA activator protein complex), PTF (PSE-binding transcription factor) or PBP (PSE-binding protein) [4]. This complex recruits RNAPII to the promoter; however, the events that follow are snRNA-specific. Typical protein-encoding genes initiate transcription through the recruitment of RNAPII by transcription factors, followed by phosphorylation of Ser⁵ within the heptad repeat (YSPTS⁵PS) of the large subunit (Rpb1) of RNAPII by TFIIH (transcription factor IIH) [cyclin H–CDK (cyclin-dependent kinase) 7], which ensures promoter clearance (reviewed in [9,10]). After the transcription of the first 100 nucleotides, there is a reduction in the levels of pSer⁵ concomitant with an increase in the phosphorylation of Ser² by P-TEFb (positive transcription elongation factor-b), which contains the kinase cyclin T–CDK9 [11]. The implication of pSer² at mRNA genes is that it increases elongation efficiency and acts as a scaffold for RNA-processing factors such as Pcf11, which are required for cleavage and polyadenylation [12,13]. Given their relative short length and lack of polyadenylation, it is not surprising that P-TEFb is dispensable for efficient elongation of snRNA genes [14]. However, it has been shown that phosphorylation of Ser² of the CTD is required for proper 3' end formation, thereby establishing a distinct requirement for P-TEFb that is not related to elongation [14]. Data from the Eick and Murphy laboratories have also shown that RNAPII indeed has an additional phosphorylation at Ser⁷ and that this modification is essential for snRNA processing [15,16]. Surprisingly, it was shown that TFIIH phosphorylates Ser⁷ of the CTD, in addition to its established role of phosphorylation of Ser⁵ [17].

The 3'-box, although required for proper 3' end formation of snRNAs, is not likely to be a termination sequence, as it has been observed that RNAPII transcribes much further downstream of the snRNA cleavage site [18]. Thus the 3' box is likely to be a recognition site for a complex of proteins required to cleave the nascent snRNA from the gene (see below). Once the snRNA is cleaved off the template, it undergoes export to the cytoplasm through the activity of the snRNA-specific export factor PHAX (phosphorylated adapter for RNA export) [19] and then continues onward in the snRNP (small nuclear ribonucleoprotein) biogenesis pathway (reviewed in [20]).

The snRNA 3'-processing complex: the Integrator proteins

Until recently, the holes in our understanding of snRNA expression have been largely unfilled. What is the identity of the complex of proteins required for 3' end formation? How are these factors recruited specifically to snRNA genes and not to other RNAPII genes? An advance in the field came from research performed in the Shiekhataar laboratory where a complex of proteins associated with RNAPII was biochemically identified, which they termed the Integrator complex [21]. This evolutionarily conserved complex contains at least 12 proteins in humans, with each having readily identifiable orthologues in other metazoans. Conspicuously, there are no detectable Integrator subunits present in *Saccharomyces cerevisiae*, consistent with the observation of the Nrd1/Nab3/Sen1 complex carrying out snRNA 3' end formation in that species [22,23]. Initially, it was shown that RNAi (RNA interference)-mediated depletion of Integrator 11 or 1 resulted in the accumulation of misprocessed snRNA, thereby establishing a function in snRNA processing.

Figure 1 is a schematic diagram of all 12 proteins with respect to domains identified using Pfam analysis (in black) and we have also shown regions of high sequence conservation (in grey). These proteins are annotated in numerical order on the basis of gel migration, with Integrator 1 having the largest predicted molecular mass of 244 kDa and Integrator 12 being the smallest at 48 kDa (in humans). At the time of purification, virtually none of the Integrator proteins had been characterized by other laboratories and most lack identifiable domains that would suggest a role in RNA processing. Moreover, a recent purification of the

Integrator complex using specific antibodies against other subunits not tested in the original purification confirms the existence of all 12 proteins in the complex and demonstrates that there may indeed be many more members [24].

Since the discovery of the Integrator complex, several studies have emerged demonstrating a role for Integrator subunits throughout evolution. Some cases are consistent with an snRNA-processing function, whereas others suggest that Integrator proteins may be multifunctional. Targeted disruption of Integrator 1 in mouse embryos results in an early embryonic lethality illustrating a fundamental requirement for this gene [25]. Moreover, the identification of misprocessed snRNA in the mutant embryo is evidence that snRNA biosynthesis is disrupted. Inhibition of Integrator 5 expression in zebrafish using morpholinos results in severe developmental defects, including a lack of circulating blood cells [26]. This is due to a requirement for Integrator 5 in haemopoietic development through reduced differentiation of erythrocytes. Reduction in functional Integrator 5 or 11 in zebrafish leads to an accumulation of misprocessed snRNA and ultimately splicing defects in the *Smad1/5* pre-mRNAs, both of which are required for haemopoiesis. The observed defect in this organism is surprisingly specific and it may be that the splicing of these particular pre-mRNAs is sensitive to the levels of snRNA. A series of recent papers have identified Integrator 3 as a specific component of the single-stranded-DNA-binding heterotrimeric complex called the SOSS (sensor of single-stranded DNA) complex that is involved in detecting DNA double-stranded breaks [27–29]. The interaction of Integrator 3 with other members of this complex requires the N-terminal region and stretches into the highly conserved region shown in Figure 1. This complex appears to be distinct from the Integrator complex, as none of the other SOSS subunits were reported in the original Integrator purification and did not appear in the recent MS analysis of Integrator immunoprecipitations [24]. In an interesting set of studies, Wieland and colleagues identified that Integrator 6 is synonymous with a presumed tumour-suppressor gene termed DICE1 (deleted in cancer 1) and that overexpression of this protein in a prostate cancer cells reduces colony formation and causes a cell-cycle arrest [30,31]. It is not known why deletion of a specific Integrator protein, which would presumably inhibit snRNA biosynthesis, contributes to tumorigenesis, but this may be further evidence that different Integrator subunits may be members of distinct subcomplexes. Integrator 6 has a well-conserved vWF-A (von Willebrand factor A) domain at its N-terminus, which has been found to be present in proteins of diverse functions ranging from cell adhesion (i.e. integrin β -subunits) to transcription (i.e. TFIIP44) [32]. The lack of a signal motif in Integrator 6 argues for an intracellular function. Removal or down-regulation of the *Caenorhabditis elegans* orthologue of Integrator 6 also results in the cessation of development and defects in mitochondrial development [33]. Indeed, this protein localizes within the mitochondria in nematode worms and provides evidence for yet another divergent and specialized function for an individual Integrator subunit. Finally, *Drosophila* that are null for the *deflated* gene (Integrator 7) exhibit various defects at several stages of development and die as late second instar larvae [34]. Trans-heterozygotes display abdominal defects that present themselves as a ‘deflated balloon’, which is the derivation of the name of this gene in fruitflies. It will be interesting to determine whether other Integrator subunits phenocopy this observation when deleted. Although there have been no reports describing any phenotype associated with loss of Integrator 12 function, it is noteworthy that this protein contains a highly conserved PHD (plant homeodomain), which is a common motif among nuclear proteins that bind chromatin. This, along with the vWF-A domain of Integrator 6, the HEAT repeats present in Integrator 4, and the β -lactamase domains of Integrator 9 and 11 represent the most identifiable motifs present within this complex.

The most structural and functional information for the Integrator proteins has been obtained with Integrator 9 and 11. These two proteins are highly homologous with CPSF (cleavage and polyadenylation specificity factor) 100 and CPSF73, which are known to be involved in

the cleavage reaction of other RNAPII transcripts (see below). Integrator 9 and 11 form a heterodimeric complex in cells and have been found to interact directly using a yeast two-hybrid analysis [35]. Interestingly, RNAi-mediated down-regulation of Integrator 11 leads to a very specific G₁ arrest in human cells [35]. This specific cell-cycle defect may be a direct result of decreased U7 snRNA biosynthesis, which would be predicted to cause a G₁ arrest owing to a decreased potential to synthesize histone mRNA. Alternatively, a G₁ arrest could be an indirect effect related to a reduction in spliceosomal snRNA production, resulting in a global aberrant splicing, which could affect the expression of proteins essential for cell-cycle progression.

The Integrator ‘cleavage complex’ and its recruitment to pre-snRNA

It is not known how the Integrator complex brings about cleavage of pre-snRNA and this represents an open avenue of research. In Figure 2 we illustrate the three different RNAPII 3'-end-processing machinery and highlight what is known about the well-studied mRNA-cleavage reactions. We compare this with a speculative model of how the Integrator complex carries out the analogous reaction on snRNAs.

Polyadenylated mRNAs have two well-defined elements at their 3' end; the AAUAAA hexamer or PAS (polyadenylation signal) and the GU/U-rich downstream element (DSE) (for simplicity, we have omitted the less-well-conserved auxiliary elements) (reviewed in [36]). The PAS is recognized by the 160 kDa protein component of the CPSF through its RNA-recognition motifs [37]. The DSE is bound by the 64 kDa component of the CstF (cleavage-stimulation factor) complex [38], which is a trimeric set of three polypeptides. Once these two sequences are recognized, they recruit a ‘cleavage factor’ that comprises two CPSF subunits (CPSF73 and CPSF100) as well as a large scaffold protein called Symplekin. The recent crystal structure of CPSF73 demonstrates that it is indeed a member of the β -lactamase family of zinc-dependent hydrolases as suggested previously using bioinformatics [39,40]. Moreover, this protein falls within the β -CASP (metallo- β -lactamase-associated CPSF Artemis SNM1/PSO2) subfamily, which contains proteins that have evolved additional features allowing for the endonucleolytic cleavage of nucleic acids. Although the crystal structure of CPSF100 clearly demonstrates that it is also a β -CASP member [40], it has several of the critical residues changed in the active site probably rendering it catalytically inactive. The implication of this change is not known, neither is how this complex is assembled with Symplekin and recruited to the site of cleavage.

The overall architecture of the 3' end elements is conserved in the replication-dependent histone pre-mRNAs, which are the only cellular mRNAs that are not polyadenylated. The key differences lie in the actual *cis* elements present at the 3' end (reviewed in [41]). The PAS is replaced by a small evolutionarily conserved SL (stem-loop) sequence, which is bound by the SLBP (stem-loop-binding protein) [42]. In lieu of the DSE is an HDE (histone downstream element) that base pairs to the U7 snRNA, which is a component of the U7 snRNP [43]. Similarly to their poly(A) counterparts, once the elements in the histone pre-mRNA are recognized, there is a recruitment of a cleavage factor that probably consists of CPSF73/CPSF100 and Symplekin, although, clearly, the method of recruitment and the type of interactions are distinct [44–47]. This suggests that there are two ‘pools’ of mRNA-cleavage factors recruited to nascent pre-mRNAs and that each pool mediates its interaction with the unique subunits of each processing complex.

The organization of the 3' elements of the pre-snRNA appear to be similar in some respects to its mRNA counterparts. The 3' box is located downstream of the cleavage site and a terminal SL structure present in all snRNAs; however, the size and content of this SL is not well conserved [6]. The 3' box is unique in that it is more tolerant to mutation than *cis*

elements present in the pre-mRNA sequences, suggesting that the recognition of this sequence to guide the cleavage event is more general rather than specific [7]. The obvious difference between the Integrator complex and the other 3'-end-processing machinery is the apparent lack of any component possessing an RNA-binding domain. This lack of an identifiable binding domain in the *trans*-acting factors leads to the question of how the 3' box is recognized. It is possible that binding to the 3' box (and potentially the 3' SL) is mediated by a yet-to-be identified RNA-binding motif present within an uncharacterized Integrator subunit. This is reminiscent of when SLBP was discovered to possess a unique RNA-binding domain not seen in other RNA-processing factors [42]. It is entirely possible that none of the Integrators binds to RNA, but, rather, this event is mediated by an associated factor.

A second interesting mechanistic question is how the Integrator complex interacts with RNAPII CTD. It has been determined by several groups that snRNA promoters play an essential role in the 3'-end processing of the precursors [48,49]. The use of non-snRNA promoters causes the misprocessing of the snRNA, suggesting that the snRNA promoter is responsible for the recruitment of the machinery that cleaves the 3' end of the snRNA. This suggests that the CTD is uniquely modified at the snRNA promoter through the activities of both P-TEFb and TFIIF, but the nature of this interaction has yet to be established. It is not clear how a general transcription factor, such as TFIIF, which is presumed to be present at all RNAPII genes, creates a phosphorylation pattern uniquely outfitted for Integrator binding, neither is it known whether Integrator proteins are present at non-snRNA genes.

Finally, how the cleavage complex is assembled is not known. The accepted model is that Integrator 11 and 9 behave analogously to CPSF73 and CPSF100 [35]. Integrator 11 is nearly 40% identical with the catalytic domain of CPSF73 in the first 450 amino acids. Both Integrator 9 and 11 are members of the β -CASP family [39]. Even more compelling is the fact that Integrator 9, like CPSF100, contains the same amino acid changes that are predicted to inactivate its catalytic activity, suggesting that this change is somehow important for the function of both cleavage factors. Interestingly, no protein related to Symplekin has been identified in the Integrator complex. It is unlikely, although possible, that Symplekin itself is a member of the Integrator complex, as it has not been identified in any pull-downs, neither is there any clear homology between an Integrator subunit and Symplekin. Given the importance of Symplekin in both histone pre-mRNA processing [50] and the cleavage of poly(A) mRNAs, it is reasonable to assume that one of the Integrator subunits behaves as a scaffold for Integrator 9 and 11. Once this cleavage complex is assembled, it must interact with the other subunits at the site of pre-snRNA cleavage to facilitate snRNA release and ultimately lead to termination and recycling of the other members of the complex.

Conclusions

The identification of the Integrator complex will probably regenerate interest in the mechanism of snRNA 3'-end formation that will undoubtedly be the focus of study in years to come. The reports that have emerged since its initial purification suggest that this complex may be multifunctional and play roles in various types of gene expression beyond snRNA processing. Many of the observations that we have highlighted in the present article, will probably become clearer as the molecular mechanism of the Integrator proteins is determined. We envisage that future research will establish connections between components of the Integrator complex and members of the general transcription apparatus.

Acknowledgments

We thank Phillip Carpenter and Todd Albrecht for their helpful comments.

Funding

J.C. acknowledges a Graduate School of Biomedical Sciences training grant and E.J.W. is funded by a National Institutes of Health Pathway to Independence Award [grant number 5R00GM080447-03].

Abbreviations used

CDK	cyclin-dependent kinase
CPSF	cleavage and polyadenylation specificity factor
β-CASP	metallo- β -lactamase-associated CPSF Artemis SNM1/PSO2
CTD	C-terminal domain
DSE	distal sequence element
PAS	polyadenylation signal
PSE	proximal sequence element
P-TEFb	positive transcription elongation factor b
RNAi	RNA interference
RNAPII	RNA polymerase II
SL	stem-loop
SLBP	stem-loop-binding protein
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SOSS	sensor of single-stranded DNA
TFIIH	transcription factor IIH
vWF-A	von Willebrand factor A

References

1. Matera AG, Weiner AM, Schmid CW. Structure and evolution of the U2 small nuclear RNA multigene family in primates: gene amplification under natural selection? *Mol Cell Biol.* 1990; 10:5876–5882.
2. Card CO, Morris GF, Brown DT, Marzluff WF. Sea urchin small nuclear RNA genes are organized in distinct tandemly repeating units. *Nucleic Acids Res.* 1982; 10:7677–7688.
3. Westin G, Zabielski J, Hammarstrom K, Monstein HJ, Bark C, Pettersson U. Clustered genes for human U2 RNA. *Proc Natl Acad Sci USA.* 1984; 81:3811–3815.
4. Hernandez N. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem.* 2001; 276:26733–26736.
5. Carbon P, Murgo S, Ebel JP, Krol A, Tebb G, Mattaj LW. A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II. *Cell.* 1987; 51:71–79.
6. Hernandez N. Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO J.* 1985; 4:1827–1837.
7. Ach RA, Weiner AM. The highly conserved U small nuclear RNA 3'-end formation signal is quite tolerant to mutation. *Mol Cell Biol.* 1987; 7:2070–2079.

8. Egloff S, O'Reilly D, Murphy S. Expression of human snRNA genes from beginning to end. *Biochem Soc Trans.* 2008; 36:590–594.
9. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 2006; 20:2922–2936.
10. Buratowski S. Progression through the RNA polymerase II CTD cycle. *Mol Cell.* 2009; 36:541–546.
11. Peng J, Marshall NF, Price DH. Identification of a cyclin subunit required for the function of *Drosophila* P-TEFb. *J Biol Chem.* 1998; 273:13855–13860.
12. Zhang Z, Fu J, Gilmour DS. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev.* 2005; 19:1572–1580.
13. Zhang Z, Gilmour DS. 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.* 2006; 21:65–74.
14. Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, Murphy S. P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *EMBO J.* 2005; 24:4154–4165.
15. Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, Eick D. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science.* 2007; 318:1780–1782.
16. Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science.* 2007; 318:1777–1779.
17. Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell.* 2009; 34:387–393.
18. Cuello P, Boyd DC, Dye MJ, Proudfoot NJ, Murphy S. Transcription of the human U2 snRNA genes continues beyond the 3' box *in vivo*. *EMBO J.* 1999; 18:2867–2877.
19. Ohno M, Segref A, Bachi A, Wilm M, Mattaj JW. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell.* 2000; 101:187–198.
20. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol.* 2007; 8:209–220.
21. Baillat D, Hakimi MA, Naar AM, Shilatfard A, Cooch N, Shiekhattar R. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell.* 2005; 123:265–276.
22. Steinmetz EJ, Brow DA. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol.* 1996; 16:6993–7003.
23. Steinmetz EJ, Conrad NK, Brow DA, Corden JL. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature.* 2001; 413:327–331.
24. Malovannaya A, Li Y, Bulynko Y, Jung SY, Wang Y, Lanz RB, O'Malley BW, Qin J. Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc Natl Acad Sci USA.* 2010; 107:2431–2436.
25. Hata T, Nakayama M. 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.* 2007; 1773:1039–1051.
26. Tao S, Cai Y, Sampath K. The Integrator subunits function in hematopoiesis by modulating Smad/BMP signaling. *Development.* 2009; 136:2757–2765.
27. Huang J, Gong Z, Ghosal G, Chen J. SOSS complexes participate in the maintenance of genomic stability. *Mol Cell.* 2009; 35:384–393.
28. Li Y, Bolderson E, Kumar R, Muniandy PA, Xue Y, Richard DJ, Seidman M, Pandita TK, Khanna KK, Wang W. hSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem.* 2009; 284:23525–23531.

29. Skaar JR, Richard DJ, Saraf A, Toschi A, Bolderson E, Florens L, Washburn MP, Khanna KK, Pagano M. INTS3 controls the hSSB1-mediated DNA damage response. *J Cell Biol.* 2009; 187:25–32.
30. Filleur S, Hirsch J, Wille A, Schon M, Sell C, Shearer MH, Nelius T, Wieland I. INTS6/DICE1 inhibits growth of human androgen-independent prostate cancer cells by altering the cell cycle profile and Wnt signaling. *Cancer Cell Int.* 2009; 9:28.
31. Wieland I, Arden KC, Michels D, Klein-Hitpass L, Bohm M, Viars CS, Weidle UH. Isolation of *DICE1*: a gene frequently affected by LOH and downregulated in lung carcinomas. *Oncogene.* 1999; 18:4530–4537.
32. Whittaker CA, Hynes RO. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell.* 2002; 13:3369–3387.
33. Han SM, Lee TH, Mun JY, Kim MJ, Kritikou EA, Lee SJ, Han SS, Hengartner MO, Koo HS. Deleted in cancer 1 (DICE1) is an essential protein controlling the topology of the inner mitochondrial membrane in *C. elegans*. *Development.* 2006; 133:3597–3606.
34. Rutkowski RJ, Warren WD. Phenotypic analysis of deflated/Ints7 function in *Drosophila* development. *Dev Dyn.* 2009; 238:1131–1139.
35. Dominski Z, Yang XC, Purdy M, Wagner EJ, Marzluff WF. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol.* 2005; 25:1489–1500.
36. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci.* 2008; 65:1099–1122.
37. Murthy KG, Manley JL. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev.* 1995; 9:2672–2683.
38. MacDonald CC, Wilusz J, Shenk T. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol.* 1994; 14:6647–6654.
39. Callebaut I, Moshous D, Mornon JP, de Villartay JP. Metallo- β -lactamase fold within nucleic acids processing enzymes: the β -CASP family. *Nucleic Acids Res.* 2002; 30:3592–3601.
40. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature.* 2006; 444:953–956.
41. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet.* 2008; 9:843–854.
42. Wang ZF, Whitfield ML, Ingledue TC 3rd, Dominski Z, Marzluff WF. The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev.* 1996; 10:3028–3040.
43. Mowry KL, Steitz JA. Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNAs. *Science.* 1987; 238:1682–1687.
44. Dominski Z, Yang XC, Marzluff WF. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell.* 2005; 123:37–48.
45. Kolev NG, Yario TA, Benson E, Steitz JA. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. *EMBO Rep.* 2008; 9:1013–1018.
46. Sullivan KD, Steiniger M, Marzluff WF. A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. *Mol Cell.* 2009; 34:322–332.
47. Wagner EJ, Burch BD, Godfrey AC, Salzler HR, Duronio RJ, Marzluff WF. A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing. *Mol Cell.* 2007; 28:692–699.
48. de Vegvar HE, Lund E, Dahlberg JE. 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell.* 1986; 47:259–266.
49. Hernandez N, Weiner AM. Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell.* 1986; 47:249–258.

50. Kolev NG, Steitz JA. Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. *Genes Dev.* 2005; 19:2583–2592.

Figure 1. Schematic diagram of Integrator protein domains and conservation

Regions identified using Pfam analysis are labelled in black, whereas regions of conservation greater than 80% over a stretch of ten amino acids are labelled in grey (conservation was compared between human, cow, chicken, *Drosophila* and zebrafish). Abbreviations: ARM, Armadillo-fold repeats; β -CASP, metallo- β -lactamase-associated CPSF Artemis SNM1/PSO2 (*Integrator 9 has these domains inactivated through single amino acid changes); DEAD, RNA helicase DEAD box (*key residues are mutated, probably rendering this domain inactive); DUF, domain of unknown function; HEAT, HEAT repeat units; IntS, Integrator subunit; β -lactamase, metallo- β -lactamase domain (*Integrator 9 has these domains inactivated through single amino acid changes); PHD, plant homeodomain finger; TPR repeats, tetratricopeptide repeats; VWA, von Willebrand factor A fold.

Figure 2. Comparison of the three 3'-end-processing complexes for RNA polymerase II transcripts

The three 'cleavage factor' proteins are denoted in light grey for all three systems with the Integrator cleavage factor having one of its unknown subunits labelled with a question mark. The dark-grey proteins represent factors involved in either CTD recognition or *cis* element binding. CstF, cleavage-stimulation factor; HDE, histone downstream element; RNA Pol II, RNA polymerase II.