

# The Integrator complex controls the termination of transcription at diverse classes of gene targets

Jeffrey R Skaar<sup>1</sup>, Andrea L Ferris<sup>3</sup>, Xiaolin Wu<sup>4</sup>, Anita Saraf<sup>5</sup>, Kum Kum Khanna<sup>6</sup>, Laurence Florens<sup>5</sup>, Michael P Washburn<sup>5,7</sup>, Stephen H Hughes<sup>3</sup>, Michele Pagano<sup>1,2</sup>

<sup>1</sup>Department of Pathology, Laura and Isaac Perlmutter Cancer Center, New York University School of Medicine, 522 First Avenue, New York, NY 10016, USA; <sup>2</sup>Howard Hughes Medical Institute, 522 First Avenue New York, NY 10016, USA; <sup>3</sup>HIV Drug Resistance Program, National Cancer Institute, Frederick, MD 21702, USA; <sup>4</sup>Cancer Research Technology Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA; <sup>5</sup>The Stowers Institute for Medical Research, Kansas City, MO 6411, USA; <sup>6</sup>QIMR Berghofer Medical Research Institute, Herston, Brisbane, Queensland 4006, Australia; <sup>7</sup>Department of Pathology and Laboratory Medicine, The University of Kansas Medical Center, Kansas City, KS 66160, USA

**Complexes containing INTS3 and either NABP1 or NABP2 were initially characterized in DNA damage responses, but their biochemical function remained unknown. Using affinity purifications and HIV Integration targeting-sequencing (HIT-Seq), we find that these complexes are part of the Integrator complex, which binds RNA Polymerase II and regulates specific target genes. Integrator cleaves snRNAs as part of their processing to their mature form in a mechanism that is intimately coupled with transcription termination. However, HIT-Seq reveals that Integrator also binds to the 3' end of replication-dependent histones and promoter proximal regions of genes with polyadenylated transcripts. Depletion of Integrator subunits results in transcription termination failure, disruption of histone mRNA processing, and polyadenylation of snRNAs and histone mRNAs. Furthermore, promoter proximal binding of Integrator negatively regulates expression of genes whose transcripts are normally polyadenylated. Integrator recruitment to all three gene classes is DSIF-dependent, suggesting that Integrator functions as a termination complex at DSIF-dependent RNA Polymerase II pause sites.**

**Keywords:** Integrator; snRNA; histone; termination; RNA Polymerase II; processing

*Cell Research* (2015) **25**:288-305. doi:10.1038/cr.2015.19; published online 13 February 2015

## Introduction

RNA Polymerase II (Pol II)-mediated transcription can be controlled at many levels, including the recruitment and assembly of the transcription machinery, and initiation, elongation, and termination of transcription [1, 2], and many RNA processing events are linked with transcription [3], providing an additional level of gene regulation. This coupling of multiple regulatory mechanisms provides precise control of gene expression.

The Pol II-transcribed snRNA genes and replica-

tion-dependent histone genes undergo a specialized processing of the 3' ends of their RNAs, which are not polyadenylated [4, 5]. These RNAs are processed by the nuclease activities of two distinct complexes, and these processing events are tightly linked to transcription termination. snRNAs are initially cleaved to a pre-snRNA form by the Integrator, a complex of over a dozen proteins defined by a combination of biochemical associations and functional assays [6-10]. Within Integrator, the heterodimeric INTS9/INTS11 nuclease directly cleaves snRNAs [11], and the functions of the remaining Integrator Subunits (INTS) remain unclear, although INTS3 and INTS10 are not required for snRNA cleavage. Notably, a recent report suggests that Integrator may not exist as a single complex and may load onto snRNA genes in a sequential fashion, but the composition and function of any such subcomplexes are unclear [12]. snRNA processing

Correspondence: Jeffrey R Skaar

E-mail: skaarlab@gmail.com

Received 26 September 2014; revised 22 November 2014; accepted 25 December 2014; published online 13 February 2015

also depends on the distance between the termination site and the promoter, and disruption of snRNA transcription termination blocks proper processing, demonstrating the coupling of these functions [13]. Conversely, inhibition of processing also blocks normal termination, further demonstrating the coupling of snRNA processing and transcription termination [14]. However, the mechanisms controlling snRNA transcription termination are largely undefined. The DRB Sensitivity-Inducing Factor (DSIF) and Negative Elongation Factor (NELF) complexes appear to play a role [14-16], and disruption of these complexes leads to extended snRNAs with poly A tails. Additionally, both chromatin structure and proteins involved in polyadenylation have been implicated in the termination of snRNA transcription [14].

Replication-dependent histone mRNA processing is comparable to snRNA processing, but histone mRNA cleavage is performed by a complex containing Stem Loop Binding Protein (SLBP), the U7 snRNP, and several components that are shared with the Cleavage and Polyadenylation Specificity Factor (CPSF) [4, 5]. Processing of the 3' ends of replication-dependent histone mRNAs imparts precise regulation that coordinates histone protein synthesis with the cell cycle through the regulation of mRNA stability. In contrast, the mRNAs of histone variants, such as H2AX, CENPA, and H3.3, are polyadenylated and are not cell cycle regulated [5].

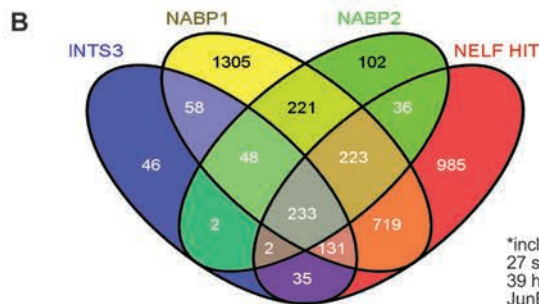
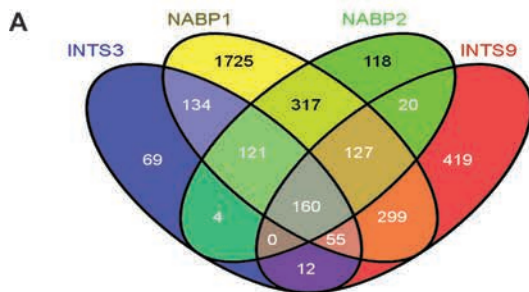
Similar to snRNA processing, replication-dependent histone mRNA processing is tightly coupled with transcription termination. Mutation of histone RNA elements or disruption of the processing machinery results in the polyadenylation of replication-dependent histone mRNAs, indicating that these mRNAs acquired polyadenylation signals, likely as a result of termination read-through [17-20]. Co-transcriptional engagement of the histone mRNA processing machinery appears to be mutually exclusive with engagement of the polyadenylation machinery, suggesting that termination failures are likely to interfere with normal replication-dependent histone mRNA processing [19]. Once polyadenylated, histone mRNAs are not substrates for processing, likely a consequence of coupled transcription and processing, and in contrast to properly processed replication-dependent histone mRNAs, the corresponding polyadenylated replication-dependent histone mRNAs are stable throughout the cell cycle [18]. The choice between proper histone mRNA processing and polyadenylation is regulated (in ways that are as yet unclear) by several additional factors, including the NELF complex [21].

The NELF complex (composed of WHSC2/NELF A, COBRA1/NELF B, THIL/NELF C/D, and NELF E) is best known for its role in promoter proximal pausing. In

higher eukaryotes, Pol II accumulates 10-60 bp downstream of the transcription start site (TSS) in 30% of all genes, and this accumulation at promoter proximal sites is thought to be the result of pausing, which is implemented by DSIF (composed of SPT4 and SPT5) and NELF [1, 2]. Productive elongation requires the removal of the pause by PTEF-b-mediated phosphorylation of DSIF and the C-terminal domain of Pol II, which results in NELF dissociation. Pausing has been proposed to be a key regulatory step controlling flux through highly inducible genes, such as stress response genes [2].

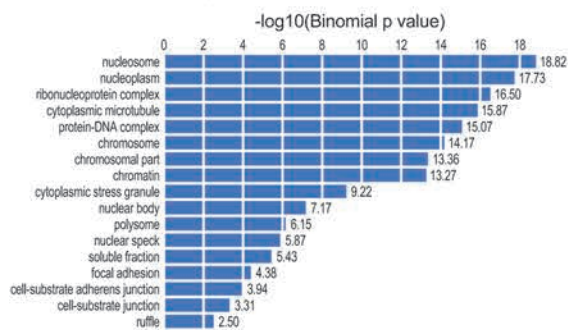
NELF plays a role in the transcription termination of snRNAs and the processing of replication-dependent histone mRNAs (a termination-associated function), but transcription termination at DSIF- and NELF-dependent promoter proximal pause sites has only recently been highlighted by several reports suggesting that pausing at promoter proximal sites could function as a decision point for transcription elongation or termination [22, 23]. Pausing may be a general feature of multiple different termination systems, including termination at the end of genes. For example, Senataxin controls termination at G-rich pause sites downstream of polyadenylation sites [24], and sequence-dependent pausing is coupled with termination and polyadenylation for many genes [25, 26]. Notably, Senataxin was recently reported to function at some early termination sites [27]. Exosome-sensitive short transcripts have also been reported in mammalian cells [23, 28, 29], and many of these transcripts are regulated by DSIF and NELF [23, 29]. Like NELF, DSIF was recently shown to function at the 3' end of snRNA genes, and although a role for DSIF at replication-dependent histone genes has not been reported, DSIF also binds these genes [16, 30, 31]. Therefore, it is likely that DSIF and NELF are functionally coupled at all loci, including the 3' end of snRNA genes, the 3' end of replication-dependent histone genes, and the promoter proximal regions of genes with polyadenylated messages.

Nucleic Acid Binding Proteins 1 and 2 (NABP1 and 2; formerly known as OBFC2A/hSSB2/SOSS-S2 and OBFC2B/hSSB1/SOSS-S1, respectively) form complexes with Integrator Subunit 3 (INTS3; SOSS-A [32-35]) and the INTS3-NABP-Interacting Protein (INIP; formerly known as c9orf80/MISE/SOSS-C). Multiple mouse and cell culture model systems support the idea that the NABP proteins require binding to INTS3 for their functions, and disruption of INTS3 blocks the functions of both NABPs [36-38]. The INTS3/NABP complexes display low to undetectable binding to the other INTS proteins [32-35], and INTS3 does not appear to be required for snRNA cleavage [7]. The biochemical functions of complexes containing INTS3 and NABP proteins, and

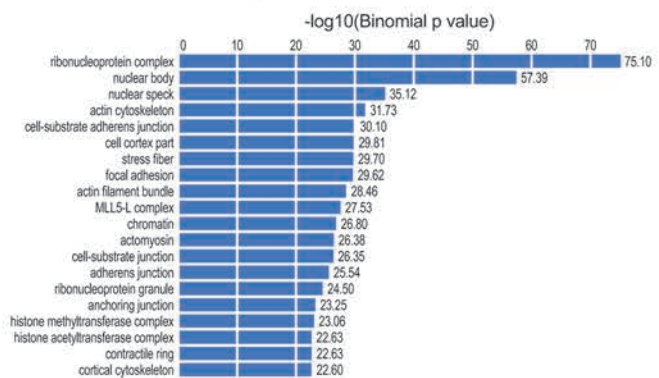


\*includes:  
27 snRNAs  
39 histones  
JunB, Sdc4, Fos1, Gadd45b

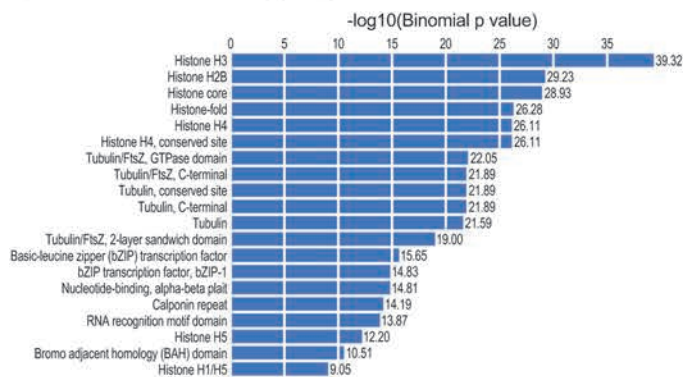
**C INTS3 Targets, Cell Compartment**



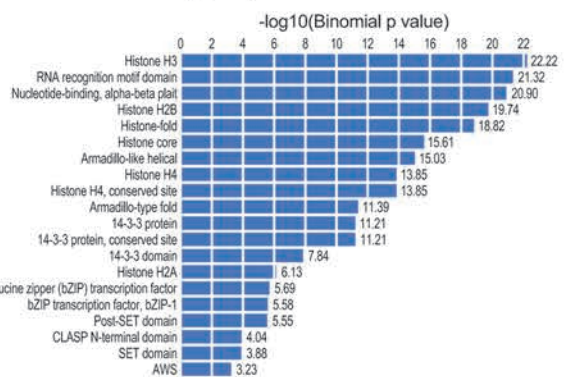
**C NABP1 Targets, Cell Compartment**



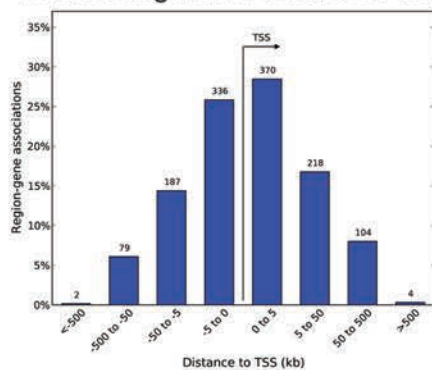
**D INTS3 Targets, Protein Families**



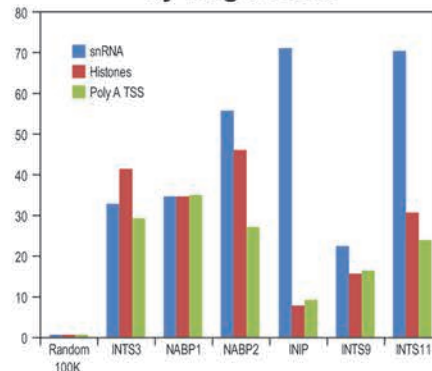
**D NABP2 Targets, Protein Families**



**E INTS3 integrations relative to TSSs**



**F Relative Integration Enrichment By Target Class**



**Figure 1** HIT-Seq analysis of Integrator, NELF and DSIF. **(A)** INTS3/NABP complexes are functionally associated with Integrator. The Venn diagram shows the intersections of target genes from the INTS3, NABP1, NABP2, and INTS9 HIT-Seq analyses. Target genes were determined by using the intersection tool of the UCSC mouse genome browser, using 2 kb windows containing six integrations as positive hits. Venn diagrams were constructed using Venny. **(B)** The Integrator complex is functionally associated with NELF. The Venn diagram shows the intersections of target genes from the INTS3, NABP1, NABP2, and a composite NELF (A, B, C, D, and E) HIT-Seq analyses. Target genes were determined by using the intersection tool of the UCSC mouse genome browser, using 2 kb windows containing six integrations as positive hits. Venn diagrams were constructed using Venny. **(C)** The Integrator complex regulates genes involved in ribonucleoprotein complexes. INTS3 and NABP1 target genes, as defined in **A**, were analyzed by GREAT for cell compartment. **(D)** The Integrator complex regulates the histone gene family. INTS3 and NABP1 target genes, as defined in **A**, were analyzed by GREAT for common protein families. **(E)** The Integrator complex binds near the transcriptional start sites (TSSs) of genes. Positive target windows, as defined in **A**, were analyzed by GREAT for their distance relative to TSS. **(F)** Enrichment scores for directed virus integrations events were calculated for snRNA genes, replication-dependent histone genes, and TSSs of genes with polyadenylated messages using the UCSC mouse genome browser. snRNA, replication-dependent histone, and polyadenylated message tracks were pre-selected using MEF GRO-Seq data to identify expressed genes, and a 700 bp window was extended from the end of the expressed genes for intersection with the HIT-Seq datasets. A track with 100 000 random integrations was used as a control.

whether they function with or without the rest of the Integrator complex, remains unclear.

In this study, we more fully characterize the proteins associated with the INTS3/NABP complexes, determine the relationship of these complexes to Integrator, define two new classes of Integrator target genes, and demonstrate that the Integrator complex participates in transcription termination at DSIF-dependent Pol II pause sites. Together, these data suggest a model in which Integrator provides a termination function that can be coupled to transcription-associated processes at multiple target genes, including snRNAs, replication-dependent histone mRNAs, and genes with polyadenylated mRNAs.

## Results

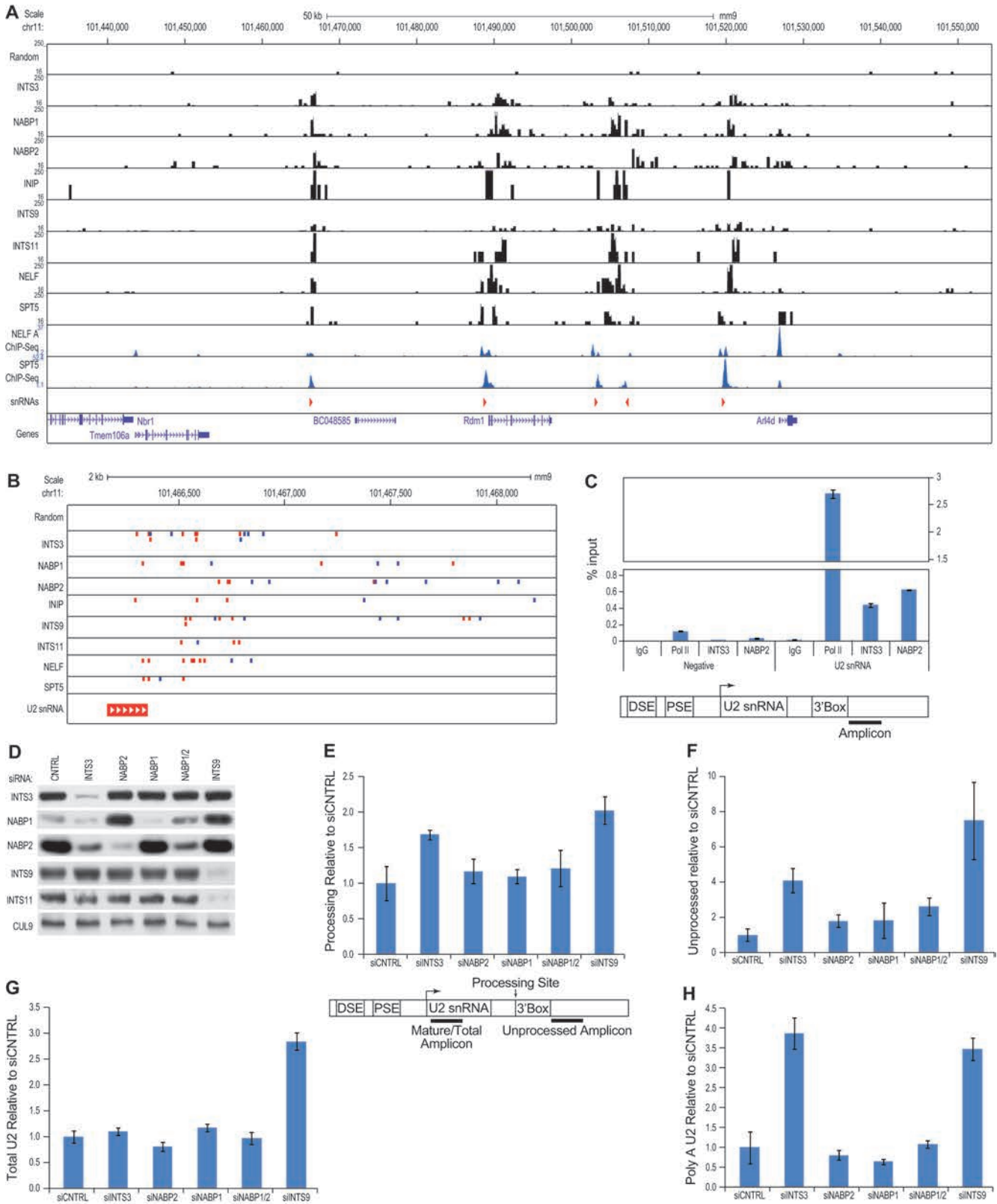
### *Composition of the INTS3/NABP complexes*

To address the composition of the INTS3/NABP complexes, tandem affinity purifications of each subunit were analyzed by Multidimensional Protein Identification Technology (MudPIT; Supplementary information, Figure S1A and S1B) [39, 40]. INTS3 and INIP co-purified with Pol II and additional INTS proteins that did not co-purify with NABP1 or 2. Using optimized conditions, the interaction of both INTS3/NABP complexes with Pol II was confirmed by immunoprecipitation and western blotting (Supplementary information, Figure S1C and S1D), consistent with a previously reported interaction between INTS3 and Pol II [41]. These results, combined with previous evidence that INTS3 regulates NABP2 at the mRNA level [32], suggested that the INTS3/NABP complexes might be protein modules that loosely associate with both the Integrator complex and Pol II to control transcription.

Although Integrator is known to regulate snRNA pro-

cessing, no additional targets for either Integrator and/or the INTS3/NABP complexes have been reported. To identify novel target genes and determine the functional overlap between Integrator and the INTS3/NABP complexes, HIT-Seq [42, 43] was used to determine the genome-wide binding sites of complex members (Supplementary information, Figure S1E). Briefly, the HIV Integrase protein binds the host chromatin protein LEDGF, which directs the viral preintegration complex to targets on the host chromosome, and viral DNA integration into the host chromosome is severely impaired (10-100-fold) in *Ledgf*-null cells [44]. This integration defect can be complemented by fusion proteins in which the Integrase Binding Domain (IBD) of LEDGF is fused to a chromatin-binding protein. These fusion proteins direct virus integrations to specific loci and/or chromatin marks bound by the chromatin-binding protein [42]. Selective amplification and sequencing of viral integration sites identifies target genes for the chromatin-binding protein [43].

LEDGF IBD fusions to INTS9, INTS11, INTS3, NABP1, NABP2, and INIP were generated and demonstrated to bind endogenous interaction partners (Supplementary information, Figure S1F and S1G). LEDGF IBD fusions to each member of the NELF complex and SPT5 were also constructed (Supplementary information, Figure S1H) for use as control markers for binding to the 3' end of snRNA genes, 3' end of replication-dependent histone genes, and promoter proximal regions of genes with polyadenylated messages [2, 15, 21, 30, 31]. *Ledgf*-null MEFs were reconstituted with individual LEDGF fusion proteins and were infected with HIV. HIV integration sites were mapped to the mouse genome (Supplementary information, Figure S1I). Integration sites were analyzed both pairwise and in combinations (Figure 1A, Supplementary information, Figure S2A and S2B). On each lev-



el, extensive overlap was observed between Integrator (as defined by the members of the dimeric, catalytic core of

INTS9 and INTS11) and INTS3/NABP complex members (Figure 1A), which, in conjunction with the MudPIT

**Figure 2** The INTS3/NABP complexes regulate snRNA genes. **(A)** Integrator, NELF, and DSIF bind snRNA genes. HIT-Seq data for a 200 kb section of mouse chromosome 11, containing a cluster of snRNA genes. The HIT-Seq data for each construct is presented as counts per kb per million (CPKM). The NELF dataset is a composite of all NELF subunits analyzed by HIT-Seq, and the NELF and SPT5 ChIP-Seq datasets are previously published [31]. The orientation of snRNA transcription is shown at the bottom. **(B)** Integrator, NELF and DSIF bind the 3' region of snRNA genes. A higher magnification of individual virus integrations at a U2 snRNA gene from the chromosome 11 locus is shown to illustrate binding in the 3' region. Blue and red marks indicate the orientation of each virus integration. **(C)** ChIP was performed from HeLa cells using the indicated antibodies and was analyzed by qPCR for the indicated amplicon at the U2 snRNA locus (+292 bp relative to the U2 snRNA start; [30]). A negative control primer in an intergenic region was also analyzed. Error bars show the standard deviation (SD) of three PCR reactions. **(D)** Western blotting was performed as indicated to show the effectiveness of the siRNA knockdowns. **(E)** Random primed cDNAs from HeLa cells transfected with the indicated siRNAs were analyzed by qPCR for unprocessed and total U2 snRNA transcripts as adapted from [59]. The ratio from each sample is presented relative to the control sample. The Data is presented as mean  $\pm$  SD. The location of the primers is shown below the graph. **(F)** Random primed cDNAs from HeLa cells transfected with the indicated siRNAs were analyzed by qPCR for unprocessed U2 snRNA transcripts versus an 18S rRNA control. Each sample is presented relative to the control sample. The data is presented as mean  $\pm$  SD. **(G)** Total U2 snRNA from HeLa cells transfected with the indicated siRNAs was analyzed by qPCR for the processed form of the U2 snRNA relative to 18S rRNA. Each sample is presented relative to the control sample. The data is presented as mean  $\pm$  SD. **(H)** Oligo-dT primed cDNAs from HeLa cells transfected with the indicated siRNAs were analyzed by qPCR for total U2 snRNA transcripts relative to actin. The results are presented relative to the control sample. The data is presented as mean  $\pm$  SD.

data demonstrating INTS3/NABP association with the full Integrator complex, suggests that the previously observed INTS3/NABP complexes are likely loosely associated components of Integrator. The Integrator HIT-Seq datasets also displayed many intersections with NELF and DSIF using either HIT-Seq or ChIP-Seq data (Figure 1B, Supplementary information, Figure S2A and S2B) [31].

As expected for regulators of snRNA genes, gene ontology analysis of the NELF, DSIF, and Integrator target genes showed enrichment for components of ribonucleoprotein complexes (Figure 1C and Supplementary information, Figure S2C). However, gene ontology analysis for enriched protein families unexpectedly revealed extensive binding of histone genes not only by NELF and DSIF (Supplementary information, Figure S3A), but also by Integrator (Figure 1D). Furthermore, even beyond the snRNA and histone genes, viral integrations directed by LEDGF-Integrator protein fusions displayed a bias towards TSSs that mirrors the localization of NELF and DSIF to TSSs (Figure 1E and Supplementary information, Figure S3B) [31].

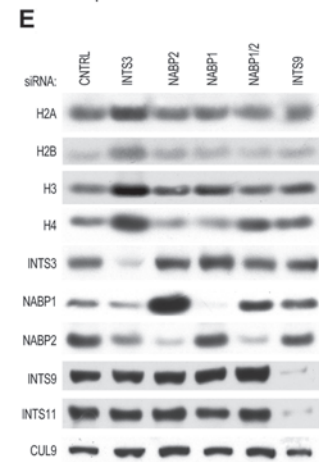
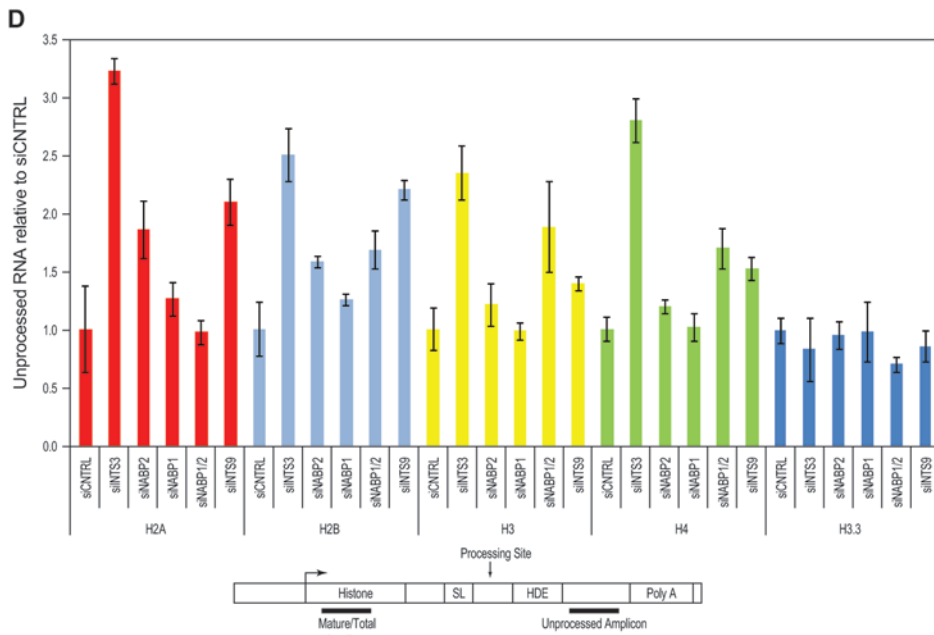
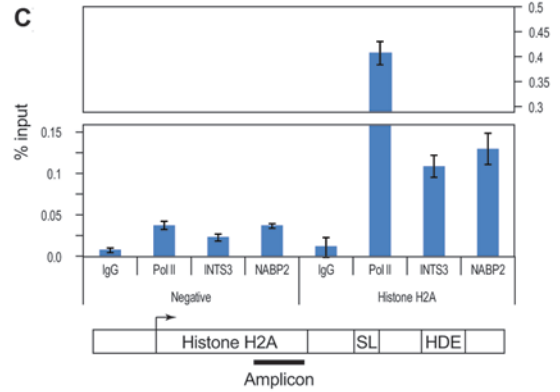
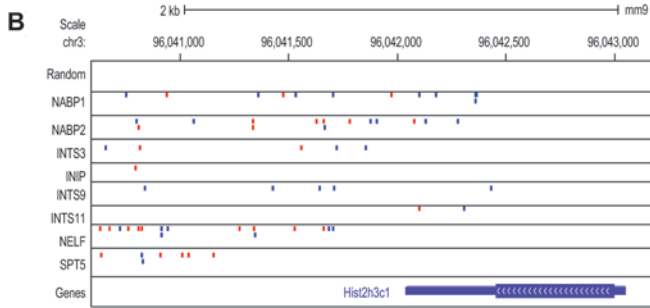
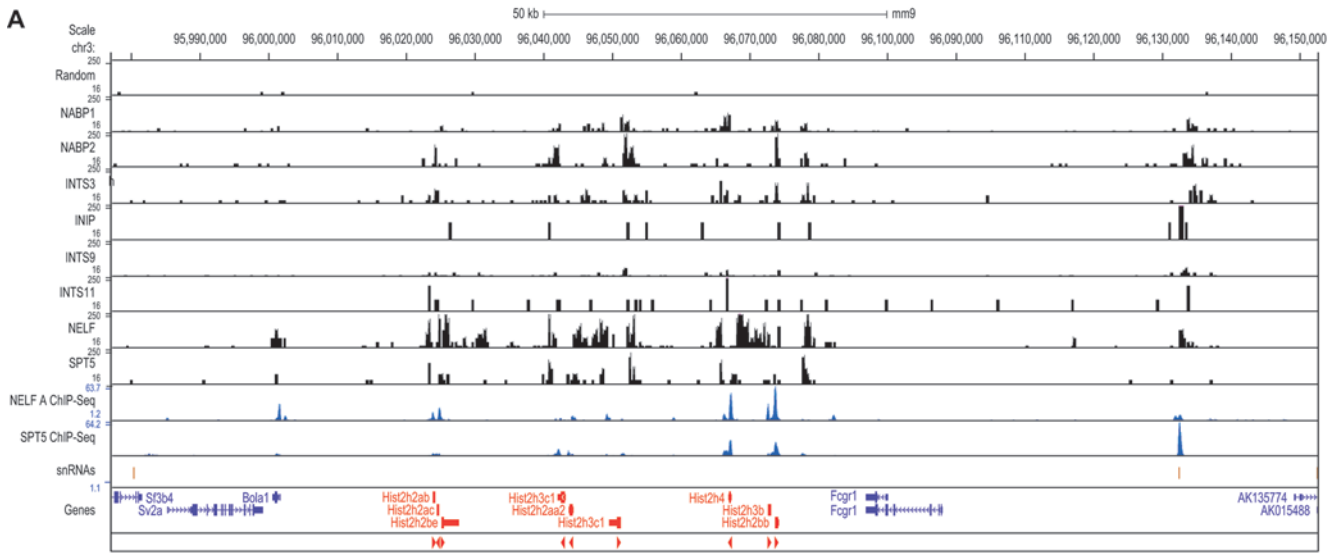
Based on this initial analysis, enrichment scores for each LEDGF fusion protein were calculated versus random integrations for three target gene classes: Pol II-transcribed snRNAs, replication-dependent histones, and the TSSs of genes with polyadenylated mRNAs. INTS3, NABP1, NABP2, INTS9, and INTS11 all displayed substantial enrichment at these loci, with enrichment values of 20-40 times of the random sample (Figure 1F). INIP did not show a similar degree of enrichment at the histones or poly A TSSs, possibly due to a smaller number of sequence reads. Overall, HIT-Seq analysis demonstrat-

ed that INTS3/NABP complexes associate with Integrator at snRNA genes, replication-dependent histones, and a subset of genes that produce polyadenylated messages.

#### *Integrator localization at snRNA genes*

Integrator has been shown to bind snRNA loci [8], and as expected, HIT-Seq detected Integrator — as defined by the binding of the catalytic core of both INTS9 and INTS11 — at snRNA genes (Figure 2A). Consistent with the observed interaction between INTS3/NABP complexes and Integrator Subunits (Supplementary information, Figure S1B, [32, 41]), HIT-Seq analysis showed that both NABP1- and NABP2- containing complexes also bound these snRNA loci (Figure 2A), and the binding of INTS3 and NABP2 to the 3' regions of snRNA genes (Figure 2B) was confirmed by Chromatin Immunoprecipitation (ChIP) (Figure 2C and Supplementary information, Figure S4A). In confirmation of multiple previous results, HIT-Seq also showed that NELF and SPT5 are localized to the 3' regions of snRNA genes (Figure 2B) [16, 30, 31].

Previously, INTS3 was reported to be dispensable for snRNA processing [7], therefore the effect of INTS3 and NABP protein depletion on snRNA processing was examined by quantitative PCR (qPCR). Western blotting demonstrated efficient knockdown of INTS3, NABP1, NABP2, and INTS9, and the established compensatory behavior of the NABP proteins was confirmed (Figure 2D) [36, 45, 46]. INTS3 depletion resulted in a small increase in unprocessed U2 snRNAs, as judged by measuring the ratio of unprocessed U2 to total U2 snRNA transcripts versus a control knockdown (Figure 2E). Notably, because NABP protein functions are dependent on



INTS3, the INTS3 knockdown more accurately reflects the combined functions of NABP1 and NABP2, [36–38], and the observed lack of effect of NABP knockdown (alone or in combination) was anticipated due to the overlapping functions of the NABPs and the inability of siRNA knockdowns to completely overcome compensatory upregulation [36, 45, 46].

The modest impact of the INTS3 knockdown on U2 snRNA processing is consistent with previous results suggesting that INTS3 is not required for Integrator-mediated processing, but in contrast to previous analyses of processing by RNase protection assays, knockdown of the INTS9 catalytic subunit of Integrator resulted in only a 2-fold increase in unprocessed U2 snRNA (Figure 2E) [8]. However, when processing was assayed by examining unprocessed U2 transcripts versus 18S rRNA (Figure 2F), INTS3 knockdown produced a 4-fold increase in misprocessing, and INTS9 knockdown produced an 8-fold increase in misprocessing. The relatively small increases in misprocessing observed by qPCR versus RNase protection assays are likely due to differences in methodology and are consistent with similar, recently reported qPCR-based examinations of snRNA processing by Integrator [16]. Additionally, a large increase in the total amount of U2 snRNA transcripts (Figure 2G) was observed with INTS9 depletion, which may indicate feedback regulation when processing is impaired, but this idea requires further investigation.

Although INTS3 appears to have a relatively minor effect on snRNA processing, because it binds the 3' end of snRNAs and associates with Pol II, it could play a role in processing-coupled termination of snRNA transcription. snRNAs are not normally polyadenylated, but based on the fact that termination failures at other genes with non-polyadenylated mRNAs (i.e., replication-dependent histones) result in readthrough, acquisition of cryptic poly A signals, and polyadenylation of the RNA

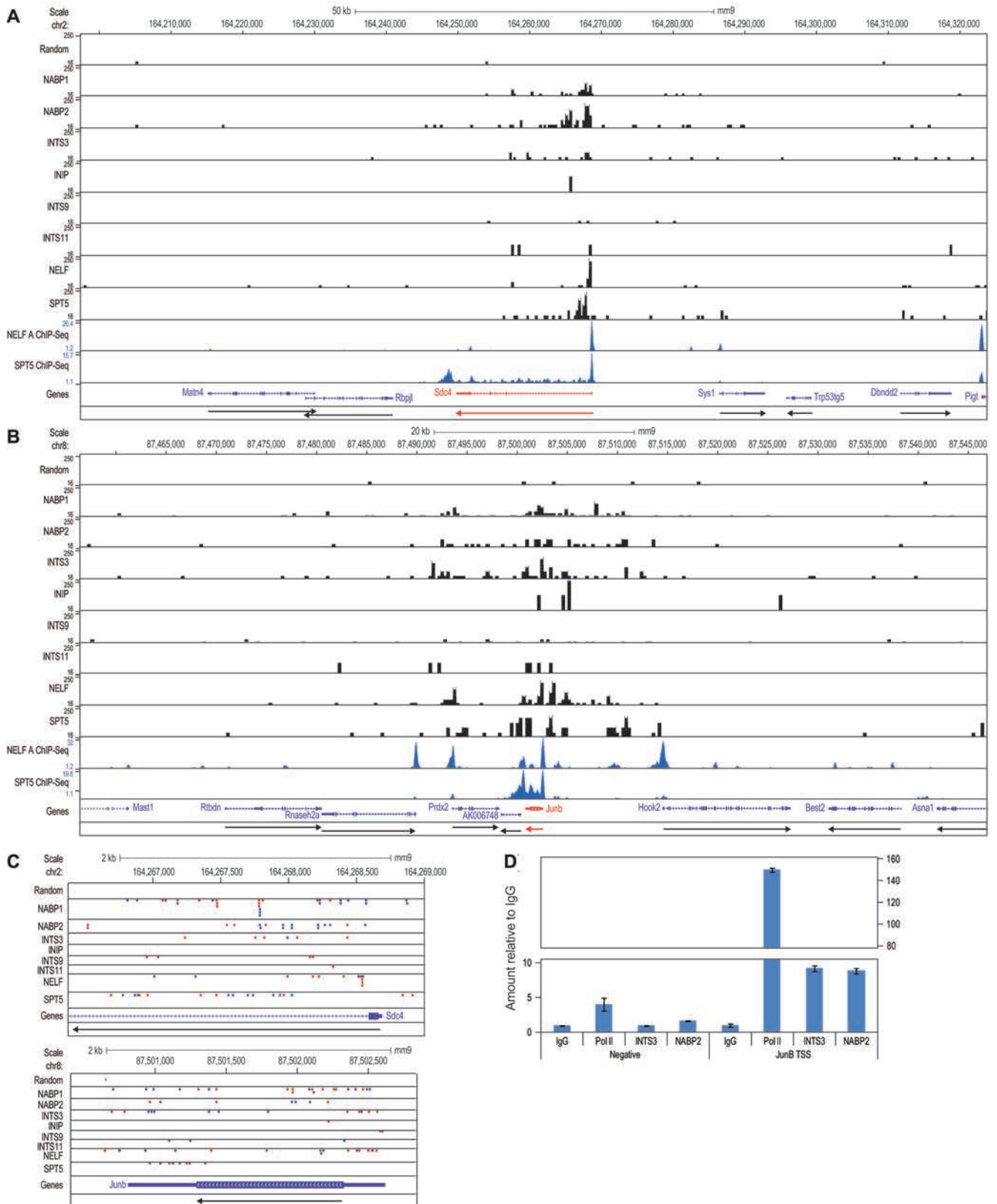
[21], the amount of U2 snRNA was measured by qPCR of oligo-dT primed cDNAs (which enrich for polyadenylated RNAs) following siRNA knockdowns of INTS3, NABP2, NABP1, NABP1 and NABP2, and INTS9 (Figure 2H). Consistent with the linkage of processing and termination, the level of U2 snRNA in the oligo-dT primed samples increased with INTS9 knockdown, and knockdown of INTS3 resulted in the accumulation of even more U2 snRNA, suggesting that in the absence of INTS9 or INTS3 the U2 snRNA is polyadenylated. Because polyadenylation requires the acquisition of a poly A signal, this result suggested that INTS3 regulates the termination of U2 snRNA transcription. Indeed, snRNA polyadenylation was recently reported following knockdown of INTS9, NELF E, or SPT5, and the sequencing of the resulting poly A snRNAs indicated transcription past the normal termination site [16].

#### *Integrator inhibition blocks replication-dependent histone mRNA processing*

The observation that disrupting INTS3 function affects snRNA polyadenylation (but not necessarily snRNA processing) suggested that the INTS3/NABP complexes provide a termination function to Integrator that is coupled with, but not intrinsic to, snRNA processing. Indeed, HIT-Seq showed that the Integrator, including INTS3 and the NABP proteins, binds to additional classes of genes beyond the snRNA genes, including replication-dependent histone gene clusters (Figure 3A and data not shown). Integrator binds to the 3' end of replication-dependent histone genes (Figure 3B), and this binding was confirmed via ChIP at a Histone H2A gene (Figure 3C and Supplementary information, Figure S4B). Similar to snRNA processing, replication-dependent histone mRNAs are not polyadenylated and undergo a processing event involving the CPSF73/CPSF100 nuclease that is coupled with transcription termination [5]. In an addi-

**Figure 3** The Integrator complex regulates replication-dependent histone genes. **(A)** Integrator, NELF, and DSIF bind replication-dependent histone genes. HIT-Seq data for a 200 kb section of mouse chromosome 3, containing histone cluster 2. The HIT-Seq data for each construct is presented as CPKM. The NELF dataset is a composite of all NELF subunits analyzed by HIT-Seq, and the NELF and SPT5 ChIP-Seq datasets are previously published [31]. The orientation of histone gene transcription is shown at the bottom. **(B)** Integrator, NELF, and DSIF bind the 3' region of replication-dependent histone genes. A higher magnification of individual virus integrations at the *Hist2h3c1* gene from the chromosome 11 locus is shown to illustrate binding in the 3' region. Blue and red marks indicate the orientation of each virus integration. **(C)** ChIP was performed from T98G cells using the indicated antibodies and was analyzed by qPCR for the indicated amplicon at the histone H2A locus. A negative control primer in an intergenic region was also analyzed. Error bars show the SD of three PCR reactions. **(D)** Random primed cDNAs from HeLa cells transfected with the indicated siRNAs were analyzed by qPCR for unprocessed and total (mature) histone transcripts as indicated. For histone H3.3, which is not processed, qPCR was performed for the 3'UTR and coding region. The results for each histone are presented relative to the control sample. The data is presented as mean  $\pm$  SD. The general location of each primer is indicated below the graph. **(E)** Western blotting was performed with lysates from HeLa cells transfected with the indicated siRNAs.





**Figure 4** Integrator binds a subset of genes with polyadenylated mRNAs at promoter proximal sites. **(A)** Integrator, NELF, and DSIF bind genes with polyadenylated transcripts. HIT-Seq data for a 200 kb section of mouse chromosome 2, containing *Sdc4*. The HIT-Seq data for each construct is presented as CPKM. The NELF dataset is a composite of all NELF subunits

tional parallel to snRNA processing, NELF and SPT5 also bind to these 3' sites, and NELF is known to regulate histone mRNA processing (Figure 3A). Although the nuclease function of Integrator does not mediate histone RNA processing [9], Integrator binds to the 3' end of replication-dependent histone genes in a manner similar to snRNA genes, suggesting a role for the complex in processing and/or termination.

To explore the role of Integrator in the regulation of replication-dependent histones, siRNA-mediated knockdowns of INTS3, NABP2, NABP1, NABP1 and 2, and INTS9 were performed, and replication-dependent histone mRNAs were analyzed for processing, polyadenylation, and total RNA levels by qPCR (Figure 3D, Supplementary information, Figure S5A and S5B). The depletion of INTS3, and to a lesser extent INTS9, resulted in an increase in unprocessed RNAs for all four replication-dependent histones (Figure 3D), and this increase in unprocessed RNA corresponded to an increase in the amount of histone RNAs detected in cDNA samples enriched for polyadenylated RNAs by oligo dT priming (Supplementary information, Figure S5A). In contrast, the ratio of the Histone H3.3 UTR to total Histone H3.3 mRNA was unaffected, as expected for the mRNA of a histone variant that is polyadenylated and not processed. In general, the total levels of the replication-dependent histone mRNAs or histone H3.3 mRNA did not increase (Supplementary information, Figure S5B). These results are consistent with results previously obtained following NELF depletion, suggesting a functional link between Integrator and NELF at the replication-dependent histone loci [21].

Strikingly, western blotting of protein extracts from knockdown samples, showed that the decrease in processing and increase in polyadenylation of replication-dependent histone mRNAs correlated with an increase in histone protein levels (Figure 3E). The incorporation of histones into chromatin is dictated by the amount of DNA, and any excess histones cannot be incorporated into chromatin. The lysis buffer in Figure 3E does not fully solubilize chromatin, suggesting that the observed

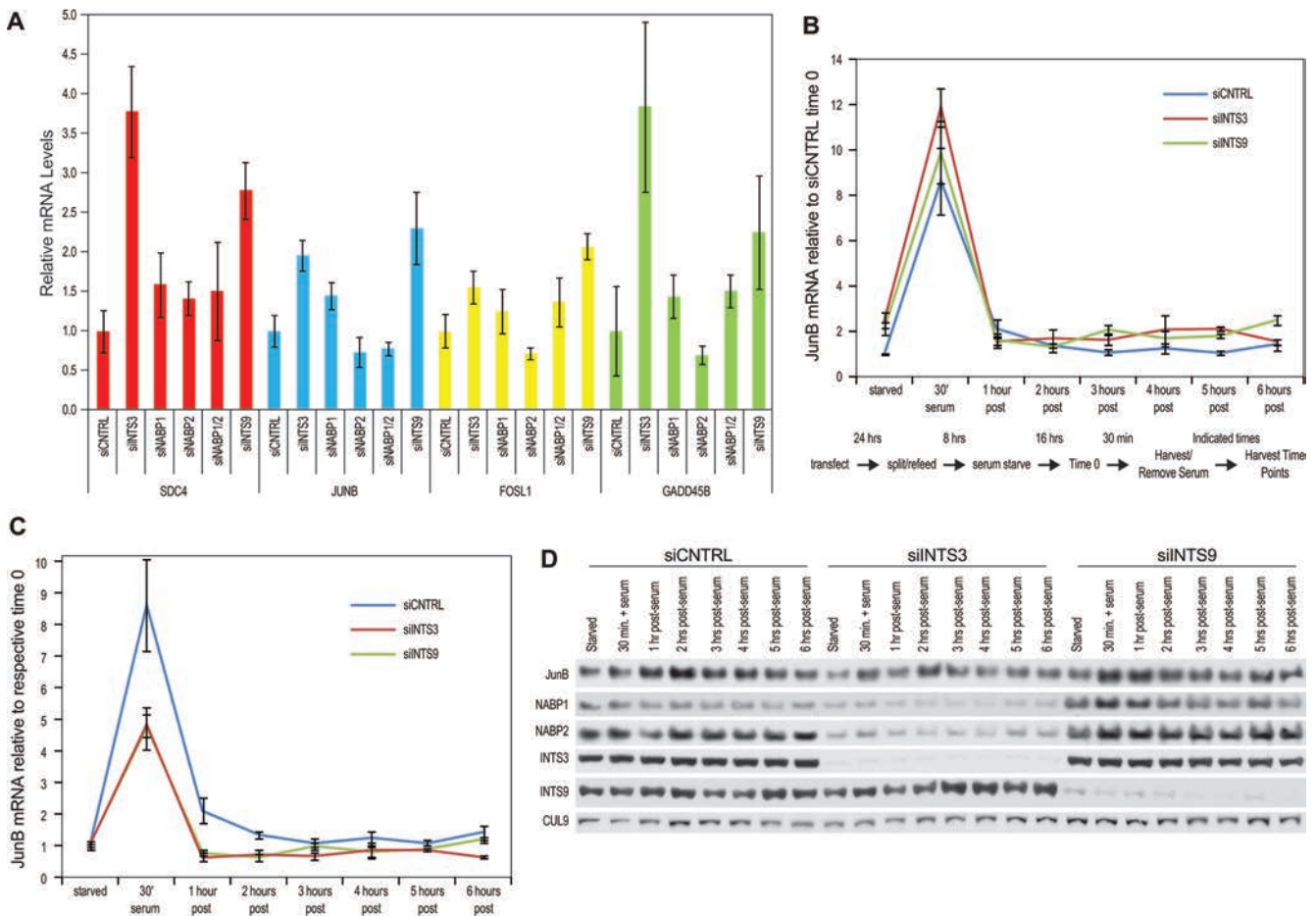
increase in histones is due to excess histones that are not incorporated into chromatin. To confirm this hypothesis, following knockdown of INTS3 or NABP2, cells were extracted with a low salt, high detergent CSK buffer to remove non-chromatin proteins before the generation of a chromatin extract via nuclease treatment and sonication. Western blotting of these extracts (Supplementary information, Figure S5C) demonstrates that INTS3 depletion resulted in an increase in replication-dependent histones only in the non-chromatin fraction, and the levels of replication-dependent histones in the chromatin fraction remained unchanged, suggesting that the observed increase in histone levels is due to the synthesis of excess histones that cannot be incorporated into chromatin.

Histone transcription is tightly regulated and is upregulated just prior to entry into S phase. Because replication-dependent histone mRNA processing is coupled to transcription, processing must also occur during this time period. The cell cycle expression profiles of NABP1 and NABP2 were analyzed by western blotting of extracts from T98G cells that were synchronized in G0 by serum starvation and released into the cell cycle by serum re-addition (Supplementary information, Figure S5D). While NABP2 levels are constant throughout the cell cycle, NABP1 levels fluctuate, with levels increasing at the G1-S transition, as indicated by the increase in cyclin A levels a few hours later (i.e., S phase). INTS3 immunoprecipitation shows that this increased level of NABP1 leads to increased binding with INTS3, at the expense of NABP2. Notably, the identical binding patterns observed for NABP1 and NABP2 in HIT-Seq may reflect the constitutive expression of the LEDGF fusion proteins. These data suggest that, under endogenous conditions, NABP1 may direct the Integrator complex to the replication-dependent histone genes.

#### *The Integrator complex regulates a subset of genes with polyadenylated mRNAs*

In contrast to the localization of Integrator to the 3' end of snRNA genes and replication-dependent histone

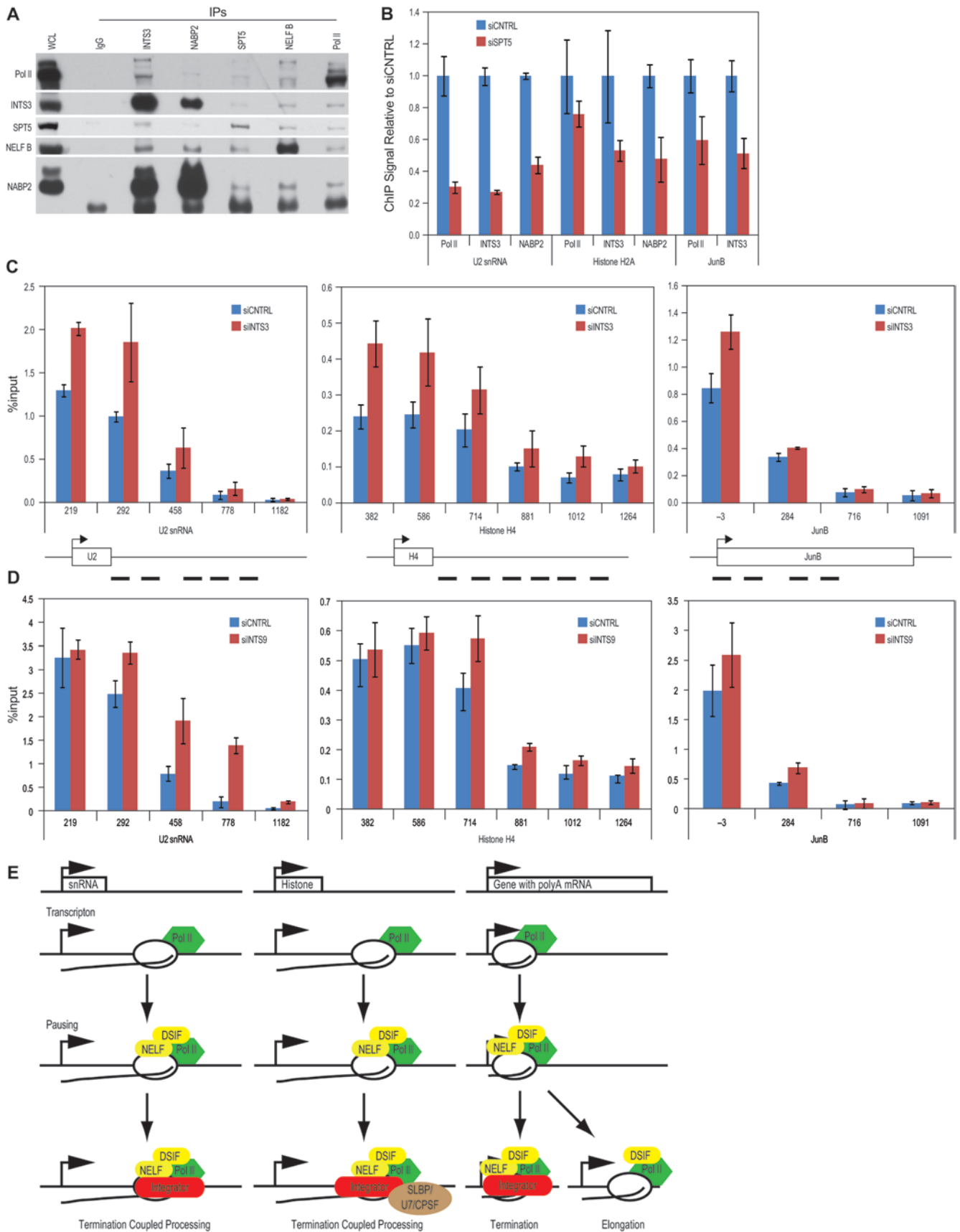
analyzed by HIT-Seq, and the NELF and SPT5 ChIP-Seq datasets are previously published [31]. The orientation of gene transcription is shown at the bottom. **(B)** Integrator, NELF and DSIF bind genes with polyadenylated transcripts. HIT-Seq data for a 200 kb section of mouse chromosome 2, containing *Junb*. The HIT-Seq data for each construct is presented as CPKM. The NELF dataset is a composite of all NELF subunits analyzed by HIT-Seq, and the NELF and SPT5 ChIP-Seq datasets are previously published. The orientation of gene transcription is shown at the bottom. **(C)** Integrator, NELF, and DSIF bind the promoter proximal sites of genes with polyadenylated transcripts, including *Sdc4* and *Junb*. Higher magnifications of the *Sdc4* and *Junb* loci are shown to illustrate binding in the promoter proximal sites. Blue and red marks indicate the orientation of each virus integration. **(D)** ChIP was performed from HeLa cells using the indicated antibodies and was analyzed by qPCR for an amplicon at the *JUNB* TSS (−3 bp; [47]). A negative control primer in an intergenic region was also analyzed. Error bars show the SD of three PCR reactions.



**Figure 5** Integrator regulates genes with polyadenylated mRNAs. **(A)** HeLa cells were transfected with the indicated siRNAs. Oligo-dT primed cDNAs were generated and analyzed by qPCR for the indicated transcripts relative to actin. The results for each gene are presented relative to the control sample. The data is presented as mean  $\pm$  SD. **(B)** HeLa cells were transfected with either a control siRNA (siCTRL), an siRNA against INTS3 (siINTS3), or an siRNA against INTS9. Twenty-four hours posttransfection, cells were split/re-fed, before a 16 h serum starvation (0.1% serum). Serum-starved cells were stimulated for 30 min with 20% serum before removal of serum (0.1% serum). Oligo-dT primed cDNAs were generated and analyzed by qPCR for *JUNB* relative to actin. All data points are shown relative to the time zero for the control siRNA samples. The data is presented as mean  $\pm$  SD. A schematic of the serum starvation/stimulation experiment is shown below the graph. **(C)** The data points from **B** are shown relative to the time zero for their respective siRNAs. The data is presented as mean  $\pm$  SD. **(D)** Western blotting was performed on protein extracts from **B** as indicated.

genes, HIT-Seq analysis also showed that Integrator binds to the 5' end of a subset of genes whose mRNAs are polyadenylated (Figure 4A-4C), and this localization was confirmed by ChIP at the *JUNB* locus for INTS3 and NABP2 (Figure 4D and Supplementary information, Figure S4C). Significantly, these sites directly overlap with the integration sites produced by NELF and SPT5 (Figure 4A and 4B), suggesting that Integrator plays a role in promoter proximal regulation of these genes by NELF and DSIF. Indeed, while many of the Integrator target genes may display largely constitutive expression (e.g., *Sdc4*, *Dot1l*, etc.), many others are inducible genes

(e.g., *JunB*, *Fos1l*, *Gadd45b*, *Mdm2*, *Cdkn1a*, *Vegfa*, *Arc*, etc.), including many immediate early genes and other stress response genes. (NELF and DSIF regulate a large percentage of genes (~30%) through promoter proximal sites, including both constitutive and inducible genes, but promoter proximal regulation has been studied predominantly at inducible genes [31].) To confirm that Integrator regulates genes with polyadenylated mRNAs through promoter proximal sites, the levels of target genes were assayed by qPCR following siRNA-mediated knockdowns of INTS3, NABP2, NABP1, NABP2 and NABP1, and INTS9 (Figure 5A). Following INTS3 or



**Figure 6** DSIF-dependent pausing is required for proper Integrator complex localization to target genes. **(A)** Integrator binds NELF and DSIF. Lysates from HeLa cells were immunoprecipitated and western blotting was performed. **(B)** HeLa cells were transfected with a control or DSIF siRNA as indicated. ChIP was performed for Pol II, INTS3, and NABP2, as indicated, with qPCR for the 3' region of the U2 snRNA, the 3' region of histone H2A, or the promoter proximal site of *JUNB*. Error bars show the SD of three PCR reactions. **(C)** Integrator controls transcription termination. HeLa cells were transfected with a control or INTS3 siRNA as indicated. ChIP was performed for Pol II, with qPCR for a series of primers in 3' region of the U2 snRNA [30], the 3' region of histone H4 [30], or the promoter proximal site of *JUNB* [47], as shown above the graphs. Data are presented relative to the control knockdown, and error bars show the SD of three independent experiments. **(D)** Integrator controls transcription termination. HeLa cells were transfected with a control or INTS9 siRNA as indicated. ChIP was performed for Pol II, with qPCR for a series of primers in 3' region of the U2 snRNA [30], the 3' region of histone H4 [30], or the promoter proximal site of *JUNB* [47], as shown below the graphs. Data are presented relative to the control knockdown, and error bars show the standard deviation of three independent experiments. **(E)** A model for the Integrator-mediated regulation of snRNAs, replication-dependent histone genes, and genes with polyadenylated mRNAs. Integrator may function as a termination module, coupling termination to a variety of co-transcriptional processes.

INTS9 knockdown, the levels of *SDC4*, *JUNB*, *FOSL1*, and *GADD45B* increased, suggesting that the Integrator complex negatively regulates the transcription of these genes.

*JUNB* is a prototypical gene model for the promoter proximal regulation of immediate early genes [47], and it is rapidly induced in response to serum. To examine the functional role of Integrator at the *JUNB* promoter proximal site, cells were depleted of INTS, serum-starved for 16 h, serum-stimulated for 30 min to induce *JUNB* transcription, and subsequently serum-starved to remove further mitogenic stimulation (Figure 5B). The induction of *JUNB* mRNA and protein was monitored by qPCR and western blotting (Figure 5B-5D). When normalized to the initial control knockdown sample (Figure 5B), INTS3 and INTS9 depletion increased the amount of *JUNB* mRNA at the zero time point over 2-fold, consistent with the steady state analysis, and the maximal *JUNB* induction level following 30 min of serum stimulation in INTS3- or INTS9-depleted cells reached ~12- and 10-fold level of the serum-starved control knockdown, respectively. Notably, beyond its negative impact on elongation through promoter proximal pausing, DSIF is also required for productive elongation [1], therefore Integrator could conceivably play both a negative and a positive role in regulating transcription elongation. However, the successful induction of *JUNB* expression following the knockdown of Integrator components suggests that Integrator does not have a DSIF-like role in elongation. Furthermore, following the withdrawal of serum, all the cells responded similarly, with *JUNB* levels dropping substantially by 1 hour post serum withdrawal, indicating that Integrator is unlikely to play a key role in restoring *JUNB* to the steady state of regulation (i.e., reducing transcription in response to feedback through early termination) and suggesting that Integrator-based regulation is secondary to regulation via transcription

initiation.

Despite the higher levels of *JUNB* at the zero time points, the maximal amount of *JUNB* induced in the INTS3- or INTS9-depleted cells, though greater than the amount in the control cells, was increased only 1.2-1.4-fold relative to the control. This result may reflect the multi-factorial nature of *JUNB* regulation. However, when the time courses for each knockdown were normalized to their own zero time points, the INTS3- and INTS9-knockdown samples showed a marked decrease in the dynamic range of *JUNB* induction (Figure 5C). From the serum-deprived state, control cells showed over an 8-fold activation of *JUNB*, but the INTS3- and INTS9-knockdown samples showed only a 4-fold activation range. Overall, Integrator does not appear to act as a binary switch in *JUNB* regulation, and instead, it appears to exert fine regulation over *JUNB*, consistent with a role in regulating transcriptional flux/the dynamic range of activation.

#### *The Integrator complex regulates transcription termination in a DSIF-dependent manner*

The consistent overlap of Integrator, NELF, and DSIF binding sites and the similar phenotypes that resulted from inhibiting these complexes suggested a functional relationship between these complexes. To define the nature of these relationships, the physical associations of INTS3 and NABP2 with NELF and DSIF were examined by immunoprecipitation and western blotting (Figure 6A). Both INTS3 and NABP2 reciprocally co-immunoprecipitated with NELF B and SPT5, as well as Pol II, supporting the idea that Integrator, NELF, and DSIF work together to regulate transcription. To define epistatic relationships between these complexes, the recruitment of these complexes to target genes was examined by ChIP following depletion of INTS3 or SPT5. While INTS3 depletion had no effect on SPT5 or NELF

B binding to the U2 snRNA or *JUNB* promoter proximal site (Supplementary information, Figure S6A and S6B), depletion of SPT5 (Supplementary information, Figure S6C) greatly reduced the recruitment of INTS3 and NABP2 to the U2 snRNA, histone H2A, and *JUNB* (Figure 6B) genes. Notably, the levels of Pol II at these sites were largely unaffected by INTS3 knockdown (Supplementary information, Figure S6B) and were greatly reduced by SPT5 knockdown (Figure 6B), particularly at the *JUNB* promoter proximal site. Since this site is a DSIF-sensitive pause site, these results suggest that Integrator is not required for Pol II pausing.

Because Pol II pausing is maintained following Integrator inhibition, the ability of Integrator to regulate transcription termination was investigated. A potential role in transcription termination was suggested by previous experiments demonstrating phenotypes associated with termination failure (e.g., the acquisition of poly A tails by a U2 snRNA and replication-dependent histone mRNAs following Integrator inhibition, suggesting transcription readthrough, and the negative regulation of genes with polyadenylated mRNAs by Integrator), and multiple previous studies that have suggested the existence of a DSIF- and NELF-dependent early termination mechanism. To assay termination of Pol II transcription, ChIP for Pol II was performed using primers spanning the U2, Histone H4, and *JUNB* loci after INTS3 or INTS9 knockdown (Figure 6C and 6D). At the 3' end of the U2 snRNA and Histone H4 genes, Pol II binding decreases as a function of termination [15, 30]. Following INTS3 knockdown, more and persistent downstream binding of Pol II was detected at the U2 and Histone H4 loci compared to the control knockdown, demonstrating that Pol II termination was impaired. The increase in Pol II at these sites is unlikely to reflect increased initiation at these genes because INTS3 knockdown has little effect on U2 snRNA levels and causes a slight decrease in Histone H4 levels (Figure 2G and Supplementary information, Figure S4B). Notably, a similar result was obtained in the region of the *JUNB* promoter proximal site. The peak of Pol II at the pause site increased, and there was also an increase in the gene body. Although this effect is seen at the 5' end of the gene, it is similar to the termination defect observed at the U2 and Histone H4 genes.

## Discussion

In summary, we demonstrate that the INTS3/NABP complexes are physically and functionally components of the Integrator complex and, as such, also bind Pol II. Furthermore, we have used HIT-Seq to identify Integrator target genes on a genome-wide level. Beyond the

snRNA loci, we show that Integrator binds to the 3' end of replication-dependent histone genes and promoter proximal sites in a subset of genes that produce polyadenylated transcripts. At all three types of target genes, Integrator colocalizes with NELF and DSIF, but Integrator does not affect the recruitment of these pausing factors. Conversely, DSIF depletion blocks Integrator binding, suggesting that Pol II pausing is required for Integrator recruitment.

NELF-dependent transcription termination has been observed for both snRNAs and replication-dependent histone mRNAs [15, 21], and similar phenotypes (i.e., increased polyadenylation of snRNAs and replication-dependent histone mRNAs) are observed following disruption of Integrator function (Figure 2H and Supplementary information, Figure S4A, [16]). Furthermore, depletion of INTS3 resulted in progression of Pol II beyond termination sites at the U2 snRNA and histone H4 loci, confirming a role for these complexes in transcription termination. Overall, these data support a model in which Integrator functions in pause-dependent termination of transcription (Figure 6E). These data are further supported by published data showing that the sequences of polyadenylated U1 snRNAs generated following depletion of INTS9 extend beyond their normal termination sites, which demonstrates defective termination [16].

snRNAs and replication-dependent histone mRNAs display clear coupling of processing and termination [13, 14, 20], but these RNAs are processed by distinct machineries. We propose that Integrator provides a modular termination function that can be coupled with multiple co-transcriptional functions, including 3' end processing of snRNAs and replication-dependent histone mRNAs. The concept of coupled pausing, processing, and termination is analogous to termination at the 3' end of genes with polyadenylated mRNAs, although the biochemical underpinnings are different [48].

Integrator binding to promoter proximal regions, the increase in Pol II at *JUNB* following INTS3 depletion, and the increase in Pol II in the body of *JUNB* following INTS3 depletion parallel termination at both the U2 snRNA and histone H4 genes and suggest that early termination occurs at promoter proximal sites, supporting previous studies implicating early termination in a DSIF- and NELF-dependent manner [23, 29]. However, while the increase in Pol II at the *JUNB* promoter proximal site and in the gene body following INTS3 depletion supports a termination function, this increase is also consistent with an increase in pausing. Existing assays for termination, including ChIP and nuclear run-ons, only demonstrate the presence of Pol II and do not measure an actual termination activity. Formal proof of termination at pro-

moter proximal sites may require the establishment of *in vitro* systems that can accurately discriminate between pausing and termination [22], especially when these functions appear to be linked. The potential contributions of early termination to the regulation of metazoan genes remain an active area of debate, as several recent studies have suggested that relatively few transcripts are regulated by early termination and that the half-life of paused RNA Polymerase is ~7 min [49, 50]. In agreement with this data, the impact of Integrator inhibition on genes with polyadenylated transcripts is only 2- to 4-fold, and Pol II pausing is not disrupted by Integrator depletion.

Notably, both DSIF and Integrator are conserved from *C. elegans* to humans, and NELF is only present from *Drosophila* to humans [51]. This evolutionary perspective and the murky status of Pol II pausing in worms may indicate an Integrator-dependent function at promoter proximal sites that is unrelated to NELF-regulated pausing. Given its termination function at both the U2 snRNA and Histone H4 genes, we favor a model in which Integrator also has a termination function at promoter proximal sites.

Several questions remain about the biochemical mechanisms of Integrator-mediated termination. It remains unclear whether the nuclease activity of Integrator is required for termination. Depletion of INTS9 affects termination at multiple loci, but this effect could be structural and related to the stability of the complex following INTS9 depletion. The association of Integrator with Pol II also suggests that two other activities in the Integrator complex may play important roles. NABP proteins, which bind single-stranded DNA (ssDNA) [37], could bind ssDNA in the transcription bubble at pause sites, and INTS6, a helicase, could be required to resolve structures such as the RNA-DNA hybrids in transcription bubbles. The impact of the NABP proteins on target gene selection is also unclear. Although the NABP proteins appear largely functionally equivalent and show compensatory behaviors (Figure 1G, [36, 45, 46]), the cell cycle-dependent regulation of NABP1 suggests that the regulation of the histone genes may be NABP1-dependent under normal circumstances, and the peri-natal lethality of the *Nabp2* mouse demonstrates that the functions of the NABP proteins are not entirely overlapping [36, 45, 46].

Finally, the INTS3/NABP complexes display several parallels to yeast Sen1, which functions in the termination of snRNAs and the mRNAs of other selective genes in yeast [52]. Human Sen1 controls termination, but it does not regulate snRNA processing. Therefore, the Integrator complex may act as a functional analog of yeast Sen1. However, at least in yeast, Sen1-dependent termi-

nation appears to involve kinetic competition with Pol II elongation [53], and our data suggest a model in which NELF- and DSIF-dependent pausing allows the Integrator complex, containing either NABP1 or NABP2, to bind ssDNA behind Pol II and actively contribute to the termination of transcription.

## Materials and Methods

### Cell lines, transfection and siRNAs

T98G, HeLa, and HEK-293T cells were maintained in DMEM supplemented with 10% fetal bovine serum (Invitrogen). T98G cells stably expressing FLAG-HA tagged NABP/INTS3 complex components were generated by retroviral infection. HEK-293T cells were transfected using PEI, and siRNA transfections were performed with Lipofectamine RNAiMAX (Invitrogen). siRNA sequences are from Dharmacon as follows:

INTS3, GAUGAGAGUUGCUAUGACA(#D-018360-01);  
SPT5, AAGAAGAACUGGGCGAGUA (#J-016234-05);  
INTS9, GAAAGCGGGUGAGCGAUGA (#D-020275-01);  
NABP2, GUUCGGACCUGCAAAGUGG (#D-014288-02);  
NABP1, GAUAUUAAGCCCGGACUGA (#D-014224-08).

### Western blotting and antibodies

Normal rabbit IgG and rabbit polyclonal antibodies to INTS3, Actin, NABP2, INTS9, INTS11, CUL9, COBRA1 (NELFB), JunB, and HA were obtained from Bethyl Laboratories, Inc. Other rabbit antibodies used include NABP2, NABP1 (Proteintech), Pol II (N-20, Santa Cruz), Spt5 (N-20, Santa Cruz), Histone H2A (Abcam), Histone H3 (Abcam), Histone H2B (Millipore), Histone H4 (gift from CD Allis), INIP (gift from W Wang), Lamin A/C (Cell Signaling), and Cyclin A. The mouse monoclonal Pol II antibody 8WG16 (Millipore) was also used (Supplementary information, Figure S1). Cell lysates were generated in lysis buffer (50 mM Tris, pH 8.0, 1 mM EDTA, 50 mM NaF, 0.5% Triton X-100, and either 150 or 250 mM NaCl, as indicated) supplemented with protease and phosphatase inhibitors. CSK (100 mM NaCl, 300 mM Sucrose, 3 mM MgCl<sub>2</sub>, and 10 mM PIPES, pH 6.8) extractions were generated as indicated. Select samples were generated by sonication (Bioruptor; Diagenode) in buffer containing 1 U/μl benzonase. Western blotting was performed as described previously [32].

### Cloning, RT-PCR and ChIP

All constructs were cloned by PCR into the designated vectors, and oligo sequences are available upon request. The cDNAs were verified by sequencing. Total RNA was generated using TRIzol (Invitrogen). cDNA was generated using either EcoDry kits (Random Hexamer or Oligo dT kits; Clontech) or SuperScript (Invitrogen). qRT-PCR was performed using Absolute SYBR green (Thermo-Fisher) on a Roche LightCycler 480. Primers for RT-PCR are available in Supplementary information, Table S1. ChIP was performed as described [43], and primers are available in Supplementary information, Table S1.

### HIT-Seq

For HIT-Seq analysis, E2 (-/-) MEF-LEDGF KO cell lines were used as previously described [42, 43, 54]. The LEDGF fu-

sion protein constructs were nucleofected into MEF-LEDGF KO cells according to the manufacturer's directions (Amaxa). After 18 h, the cells were harvested and sorted for GFP expression using a Becton Dickinson FACS Aria 1 cell sorter. Approximately  $1 \times 10^6$  GFP-positive cells were plated in a 100-mm dish and infected with 500 ng of VSV-G pseudotyped pNLN<sub>g</sub>oMIVR-Emod Luc HIV-1 in the presence of 8  $\mu$ g/ml polybrene. The binding sites for the LEDGF fusion proteins were identified by amplifying HIV integration sites using linker-mediated PCR. The methods used to prepare the DNA fragments, and to amplify and sequence the integration sites are previously described [55]. Briefly, 5  $\mu$ g of the DNA was sheared into ~300-500 bp fragments by adaptive focused acoustics (Covaris Inc., Woburn, MA) and purified using AMPure XP magnetic beads (Beckman Coulter Genomics, Danvers, MA). The sheared DNA was end-repaired, dA-tailed, and ligated to a double-stranded DNA linker as described in Illumina's sequencing protocol. The resulting libraries were sequenced on the Illumina Gallx. Integration junction site sequences were trimmed for LTR and linker sequences and mapped to the mouse genome (mouse genome build mm9, July 2007, University of California Santa Cruz (UCSC) genome website) using BLAT. Integration sites were considered to be authentic if the sequence (i) began with 3 bp of the end of the HIV-1 LTR, (ii) had a match to the mouse genome with at least 20 bp in length and 95% identity, (iii) had a unique best hit to the mouse genome, and (iv) the paired ends were within 1 kb on the same chromosome. Because PCR amplification can produce multiple copies of the same integration site, only tags with unique paired end alignments were used for further analysis. Venn diagrams of target genes were generated using Venny (<http://bioinfogp.cnb.csic.es/tools/venny/>). Gene ontology analysis was performed using the Genomic Regions Enrichment of Annotations Tool [56]. GRO-Seq and Chip-Seq data were previously published [31, 57]. HIT-Seq data has been deposited under GEO accession number GSE65090.

#### Tandem affinity purification and mass spectrometry

Tandem affinity purification was performed as previously described [32]. MudPIT of TCA-precipitated proteins was performed as previously described [32, 39]. Tandem mass spectrometry were interpreted using SEQUEST against a database of 61 738 sequences, consisting of 30 709 human proteins (NCBI Protein database on July 9, 2009), 160 usual contaminants, and, to estimate false discovery rates, 30 869 randomized amino acid sequences derived from each nonredundant protein entry. Peptide/spectrum matches were sorted and selected using DTASelect with the following criteria: spectra/peptide matches were only retained if they had a DeltCn of at least 0.08 and minimum XCorr of 1.8 for singly, 2.5 for doubly, and 3.5 for triply charged spectra. Peptides had to be fully tryptic and at least 7 aa long, and positive identification required two unique peptides or one peptide with two independent spectra. The final false discovery rates at the protein and spectral levels were 1.9% and  $0.14 \pm 0.085\%$ , respectively. dNSAF values were calculated as described [58].

#### Acknowledgments

We thank E McIntush/Bethyl Laboratories, Inc., CD Allis, W Wang and BD Dynlacht for antibodies, and CM Hughes, T Milne and S Buratowski for critical reading of the manuscript. We are

grateful to J Meyer for his help with figure assembly. This work was funded by grants from the National Health and Medical Research Council of Australia to KKK and grants from the National Institutes of Health (R01-GM057587, R37-CA076584, and R21-CA161108) to MP. JRS is a Special Fellow of The Leukemia & Lymphoma Society, received support from the Mr. and Mrs. William G Campbell Postdoctoral Fellowship in Memory of Carolyn Cabott from the American Cancer Society, and is currently supported by NIH T32 CA009161 (Levy). A Saraf, L Florens, and MP Washburn are supported by the Stowers Institute for Medical Research. The Hughes laboratory was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. This project has been funded in part (XW) with Federal funds from the Frederick National Laboratory for Cancer Research, National Institutes of Health, under contract HHSN261200800001E.

#### Disclaimer

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

#### References

- 1 Zhou Q, Li T, Price DH. RNA polymerase II elongation control. *Annu Rev Biochem* 2012; **81**:119-143.
- 2 Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* 2012; **13**:720-731.
- 3 Hsin JP, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* 2012; **26**:2119-2137.
- 4 Chen J, Wagner EJ. snRNA 3' end formation: the dawn of the Integrator complex. *Biochem Soc Trans* 2010; **38**:1082-1087.
- 5 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.
- 6 Huang Q, Jacobson MR, Pederson T. 3' processing of human pre-U2 small nuclear RNA: a base-pairing interaction between the 3' extension of the precursor and an internal region. *Mol Cell Biol* 1997; **17**:7178-7185.
- 7 Ezzeddine N, Chen J, Waltenspiel B, *et al.* A subset of *Drosophila* integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation. *Mol Cell Biol* 2011; **31**:328-341.
- 8 Baillat D, Hakimi MA, Naar AM, *et al.* Integrator, a multi-protein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 2005; **123**:265-276.
- 9 Chen J, Ezzeddine N, Waltenspiel B, *et al.* An RNAi screen identifies additional members of the *Drosophila* Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA* 2012; **18**:2148-2156.
- 10 Uguen P, Murphy S. The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *EMBO J* 2003; **22**:4544-4554.
- 11 Albrecht TR, Wagner EJ. snRNA 3' end formation requires heterodimeric association of integrator subunits. *Mol Cell*



- Biol* 2012; **32**:1112-1123.
- 12 Eglhoff S, Zaborowska J, Laitem C, Kiss T, Murphy S. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell* 2012; **45**:111-122.
  - 13 Ramamurthy L, Ingledue TC, Pilch DR, Kay BK, Marzluff WF. Increasing the distance between the snRNA promoter and the 3' box decreases the efficiency of snRNA 3'-end formation. *Nucleic Acids Res* 1996; **24**:4525-4534.
  - 14 O'Reilly D, Kuznetsova OV, Laitem C, et al. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res* 2014; **42**:264-275.
  - 15 Eglhoff S, Al-Rawaf H, O'Reilly D, Murphy S. Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. *Mol Cell Biol* 2009; **29**:4002-4013.
  - 16 Yamamoto J, Hagiwara Y, Chiba K, et al. DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat Commun* 2014; **5**:4263.
  - 17 Sullivan E, Santiago C, Parker ED, et al. *Drosophila* stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. *Genes Dev* 2001; **15**:173-187.
  - 18 Levine BJ, Chodchoy N, Marzluff WF, Skoultchi AI. Coupling of replication type histone mRNA levels to DNA synthesis requires the stem-loop sequence at the 3' end of the mRNA. *Proc Natl Acad Sci USA* 1987; **84**:6189-6193.
  - 19 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.
  - 20 Chodchoy N, Pandey NB, Marzluff WF. An intact histone 3'-processing site is required for transcription termination in a mouse histone H2a gene. *Mol Cell Biol* 1991; **11**:497-509.
  - 21 Narita T, Yung TM, Yamamoto J, et al. NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Mol Cell* 2007; **26**:349-365.
  - 22 Buratowski S. Progression through the RNA polymerase II CTD cycle. *Mol Cell* 2009; **36**:541-546.
  - 23 Brannan K, Kim H, Erickson B, et al. mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. *Mol Cell* 2012; **46**:311-324.
  - 24 Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* 2011; **42**:794-805.
  - 25 Gromak N, West S, Proudfoot NJ. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* 2006; **26**:3986-3996.
  - 26 Enriquez-Harris P, Levitt N, Briggs D, Proudfoot NJ. A pause site for RNA polymerase II is associated with termination of transcription. *EMBO J* 1991; **10**:1833-1842.
  - 27 Wagschal A, Rousset E, Basavarajaiah P, et al. Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* 2012; **150**:1147-1157.
  - 28 Preker P, Nielsen J, Kammler S, et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 2008; **322**:1851-1854.
  - 29 Flynn RA, Almada AE, Zamudio JR, Sharp PA. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci USA* 2011; **108**:10460-10465.
  - 30 Glover-Cutter K, Kim S, Espinosa J, Bentley DL. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol* 2008; **15**:71-78.
  - 31 Rahl PB, Lin CY, Seila AC, et al. c-Myc regulates transcriptional pause release. *Cell* 2010; **141**:432-445.
  - 32 Skaar JR, Richard DJ, Saraf A, et al. INTS3 controls the hSSB1-mediated DNA damage response. *J Cell Biol* 2009; **187**:25-32.
  - 33 Huang J, Gong Z, Ghosal G, Chen J. SOSS complexes participate in the maintenance of genomic stability. *Mol Cell* 2009; **35**:384-393.
  - 34 Li Y, Bolderson E, Kumar R, et al. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem* 2009; **284**:23525-23531.
  - 35 Zhang F, Wu J, Yu X. Integrator3, a partner of single-stranded DNA-binding protein 1, participates in the DNA damage response. *J Biol Chem* 2009; **284**:30408-30415.
  - 36 Gu P, Deng W, Lei M, Chang S. Single strand DNA binding proteins 1 and 2 protect newly replicated telomeres. *Cell Res* 2013; **23**:705-719.
  - 37 Ren W, Chen H, Sun Q, et al. Structural basis of SOSS1 complex assembly and recognition of ssDNA. *Cell Rep* 2014; **6**:982-991.
  - 38 Zhang F, Chen Y, Li M, Yu X. The oligonucleotide/oligosaccharide-binding fold motif is a poly(ADP-ribose)-binding domain that mediates DNA damage response. *Proc Natl Acad Sci USA* 2014; **111**:7278-7283.
  - 39 Florens L, Washburn MP. Proteomic analysis by multidimensional protein identification technology. *Methods Mol Biol* 2006; **328**:159-175.
  - 40 Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001; **19**:242-247.
  - 41 Malovannaya A, Li Y, Bulynko Y, et al. Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc Natl Acad Sci USA* 2010; **107**:2431-2436.
  - 42 Ferris AL, Wu X, Hughes CM, et al. Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc Natl Acad Sci USA* 2010; **107**:3135-3140.
  - 43 Lauberth SM, Nakayama T, Wu X, et al. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 2013; **152**:1021-1036.
  - 44 Shun MC, Raghavendra NK, Vandegraaff N, et al. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev* 2007; **21**:1767-1778.
  - 45 Feldhahn N, Ferretti E, Robbiani DF, et al. The hSSB1 orthologue Obfc2b is essential for skeletogenesis but dispensable for the DNA damage response *in vivo*. *EMBO J* 2012; **31**:4045-4056.
  - 46 Shi W, Bain AL, Schwer B, et al. Essential developmental, genomic stability, and tumour suppressor functions of the mouse orthologue of hSSB1/NABP2. *PLoS Genet* 2013; **9**:e1003298.

- 47 Aida M, Chen Y, Nakajima K, *et al.* Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the *junB* gene. *Mol Cell Biol* 2006; **26**:6094-6104.
- 48 Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA processing with transcription. *Cell* 2002; **108**:501-512.
- 49 Henriques T, Gilchrist DA, Nechaev S, *et al.* Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol Cell* 2013; **52**:517-528.
- 50 Jonkers I, Kwak H, Lis JT. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* 2014; **3**:e02407.
- 51 Bowman EA, Kelly W. RNA Polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: a tail of two kinases. *Nucleus* 2014; **5**:224-236.
- 52 Kuehner JN, Pearson EL, Moore C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* 2011; **12**:283-294.
- 53 Hazelbaker DZ, Marquardt S, Wlotzka W, Buratowski S. Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Mol Cell* 2013; **49**:55-66.
- 54 Richard DJ, Bolderson E, Cubeddu L, *et al.* Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature* 2008; **453**:677-681.
- 55 Gillet NA, Malani N, Melamed A, *et al.* The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood* 2011; **117**:3113-3122.
- 56 McLean CY, Bristor D, Hiller M, *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010; **28**:495-501.
- 57 Min IM, Waterfall JJ, Core LJ, *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev* 2011; **25**:742-754.
- 58 Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal Chem* 2010; **82**:2272-2281.
- 59 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.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>