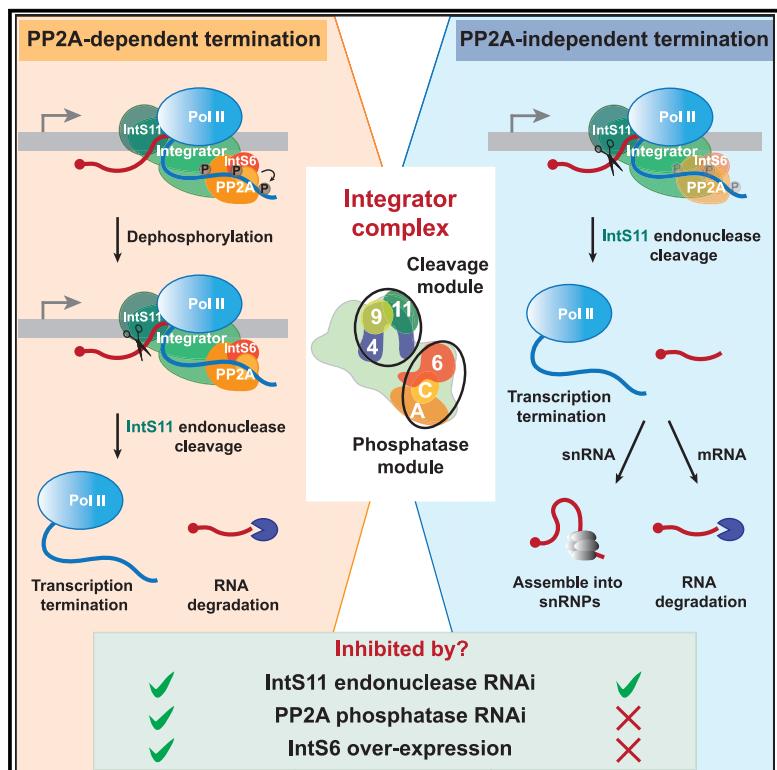


## IntS6 and the Integrator phosphatase module tune the efficiency of select premature transcription termination events

### Graphical abstract



### Authors

Rina Fujiwara, Si-Nan Zhai,  
Dongming Liang, ...,  
Deirdre C. Tatomer, Li Yang,  
Jeremy E. Wilusz

### Correspondence

[jeremy.wilusz@bcm.edu](mailto:jeremy.wilusz@bcm.edu)

### In brief

Fujiwara et al. show that the phosphatase module of the Integrator complex is differentially required for transcription termination across *Drosophila* gene loci. When protein phosphatase 2A subunits are limiting, Integrator complex function is blocked at a subset of protein-coding genes, yet snRNA processing and transcription attenuation of other loci continue.

### Highlights

- The Integrator phosphatase module is critical at only a subset of regulated genes
- IntS6 over-expression blocks Integrator function by titrating PP2A subunits
- IntS6 functions analogous to a PP2A regulatory B subunit
- PP2A recruitment can be tuned to modulate transcription termination efficiency

## Article

# IntS6 and the Integrator phosphatase module tune the efficiency of select premature transcription termination events

Rina Fujiwara,<sup>1,5</sup> Si-Nan Zhai,<sup>2,3,5</sup> Dongming Liang,<sup>4</sup> Aayushi P. Shah,<sup>1</sup> Matthew Tracey,<sup>4</sup> Xu-Kai Ma,<sup>3</sup> Christopher J. Fields,<sup>1</sup> María Sará Mendoza-Figueroa,<sup>4</sup> Michele C. Meline,<sup>4</sup> Deirdre C. Tatomer,<sup>4</sup> Li Yang,<sup>3</sup> and Jeremy E. Wilusz<sup>1,6,\*</sup>

<sup>1</sup>Verna and Marrs McLean Department of Biochemistry and Molecular Pharmacology, Therapeutic Innovation Center, Baylor College of Medicine, Houston, TX 77030, USA

<sup>2</sup>Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>3</sup>Center for Molecular Medicine, Children's Hospital, Fudan University and Shanghai Key Laboratory of Medical Epigenetics, International Laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

<sup>4</sup>Department of Biochemistry and Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead contact

\*Correspondence: [jeremy.wilusz@bcm.edu](mailto:jeremy.wilusz@bcm.edu)

<https://doi.org/10.1016/j.molcel.2023.10.035>

## SUMMARY

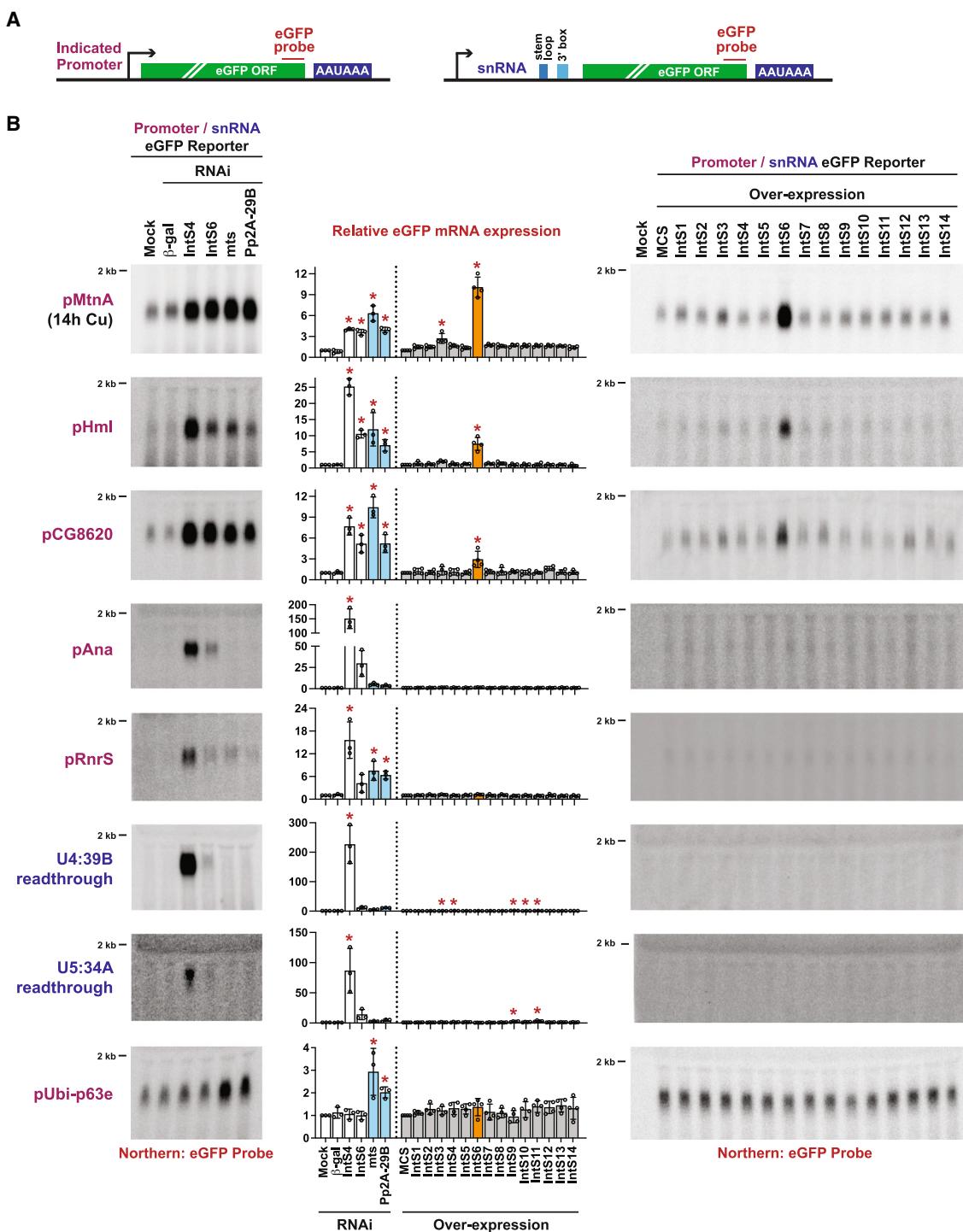
The metazoan-specific Integrator complex catalyzes 3' end processing of small nuclear RNAs (snRNAs) and premature termination that attenuates the transcription of many protein-coding genes. Integrator has RNA endonuclease and protein phosphatase activities, but it remains unclear if both are required for complex function. Here, we show IntS6 (Integrator subunit 6) over-expression blocks Integrator function at a subset of *Drosophila* protein-coding genes, although having no effect on snRNAs or attenuation of other loci. Over-expressed IntS6 titrates protein phosphatase 2A (PP2A) subunits, thereby only affecting gene loci where phosphatase activity is necessary for Integrator function. IntS6 functions analogous to a PP2A regulatory B subunit as over-expression of canonical B subunits, which do not bind Integrator, is also sufficient to inhibit Integrator activity. These results show that the phosphatase module is critical at only a subset of Integrator-regulated genes and point to PP2A recruitment as a tunable step that modulates transcription termination efficiency.

## INTRODUCTION

Transcription of all eukaryotic protein-coding genes and most non-coding RNAs is catalyzed by RNA polymerase II (Pol II), which itself is regulated by many transcription factors and co-factors that act during the initiation, elongation, and termination steps (for review, see Schier and Taatjes,<sup>1</sup> Proudfoot,<sup>2</sup> and Cramer<sup>3</sup>). These multiple macromolecular assemblies collectively help dictate the levels, processing, and functions of the transcripts produced. In recent years, the multi-subunit Integrator complex (sometimes referred to as INTAC [Integrator-containing PP2A-AC complex]) has emerged as a key transcriptional regulator across metazoans that controls the fates of many nascent RNAs (for review, see Mendoza-Figueroa et al.,<sup>4</sup> Welsh and Gardini,<sup>5</sup> Sabath and Jonas,<sup>6</sup> Kirstein et al.,<sup>7</sup> and Pfleiderer and Gale<sup>8</sup>). Integrator is >1.5 MDa and interacts with the C-terminal domain (CTD) of Pol II, with most of the more than 14 subunits in the Integrator complex

lacking identifiable paralogs.<sup>9</sup> The notable exceptions are Integrator subunits 11 (IntS11) and 9 (IntS9), which are homologous to the RNA endonuclease CPSF73 (cleavage and polyadenylation specificity factor 73, also known as CPSF3) and CPSF100 (also known as CPSF2) that cleave mRNA 3' ends prior to poly(A) tail addition.<sup>10</sup> The catalytic and scaffolding subunits of protein phosphatase 2A (PP2A) are also stably associated with Integrator, and thus, the complex can have dual RNA cleavage and phosphatase catalytic activities.<sup>11–18</sup>

It has long been recognized that Integrator catalyzes RNA cleavage at the 3' ends of nascent small nuclear RNA (snRNA) gene loci, enabling the termination of transcription and release of the snRNA transcript that goes on to function in pre-mRNA splicing.<sup>9,19</sup> Integrator has further been implicated in the processing of a variety of other Pol II transcripts, including enhancer RNAs,<sup>20,21</sup> telomerase RNA,<sup>22</sup> viral microRNAs,<sup>23</sup> piwi-interacting RNAs,<sup>24,25</sup> replication-dependent histone



**Figure 1. Integrator activity at a subset of reporter genes is lost upon over-expression of IntS6**

(A) eGFP-based reporters to examine Integrator activity at protein-coding genes (left) and snRNAs (right). The promoter and 5' UTR of each indicated protein-coding gene were cloned upstream of eGFP (left). The snRNA promoter, coding sequence, and downstream region were cloned upstream of eGFP, thereby enabling eGFP production when Integrator fails to process the snRNA 3' end between the stem loop and 3' box sequences (right).

(B) Each individual reporter plasmid was transfected into DL1 cells that had been treated with the indicated dsRNAs (left) or was co-transfected with 100 ng of a plasmid that over-expresses a FLAG-tagged Integrator subunit from the Ubi-p63e promoter (right). A plasmid containing a multi-cloning site (MCS) driven by the Ubi-p63e promoter was used as a control for the over-expression experiments. CuSO<sub>4</sub> was added for the last 14 h only when measuring eGFP production from

(legend continued on next page)

# Molecular Cell

## Article



mRNAs,<sup>26</sup> and many canonical protein-coding genes.<sup>27–32</sup> It is thus perhaps not surprising that misregulation of Integrator is associated with developmental and disease phenotypes, including in humans.<sup>4,33–37</sup>

Analogous to how Integrator functions at snRNAs, the complex can endonucleolytically cleave nascent mRNAs to enable transcription termination at protein-coding genes coupled to nascent RNA release.<sup>27–30</sup> These cleavage events predominantly do not occur at the 3' ends of genes where CPSF73 acts but instead close to the transcription start site (TSS), often within 100 nt at sites of Pol II pausing. This is at least in part because the Integrator phosphatase module can antagonize kinases that promote the release of paused Pol II.<sup>11–14,38</sup> Instead of allowing Pol II to productively elongate, Integrator causes transcription to prematurely terminate, and the released short mRNAs are rapidly degraded by the RNA exosome.<sup>27–29</sup> In some cases, Integrator cleavage has been proposed to activate transcription by enabling removal of stalled, inactive Pol II,<sup>32,38</sup> but termination driven by Integrator can also attenuate gene expression by blocking full-length mRNA production. Attenuation can be potent and readily observed in *Drosophila* cells, as exemplified by our prior genome-scale RNAi screen, which revealed that the depletion of Integrator subunits resulted in the largest increase (among all genes tested) in the output of the Metallothionein A (MtnA) promoter.<sup>28</sup> RNA sequencing (RNA-seq) further revealed more than 400 mRNAs that were upregulated in *Drosophila* cells upon the depletion of IntS9, some by more than 100-fold, compared with only 49 mRNAs that were downregulated. Integrator is thus a major attenuator of protein-coding gene outputs in *Drosophila*, yet how Integrator activity can be toggled on/off at a given gene locus depending on cellular transcriptional needs remains poorly understood.

There is emerging evidence that Integrator recruitment and/or activity is differentially regulated across gene loci (for review, see Mendoza-Figueroa et al.<sup>4</sup>). For example, depletion of many non-catalytic Integrator subunits has only a minimal effect on snRNA 3' end processing, yet these subunits are critical (especially IntS1, 2, 5, 6, 7, 8, and 12) for the ability of Integrator to attenuate the outputs of protein-coding genes.<sup>28,39,40</sup> This suggests these non-catalytic Integrator subunits may act to ensure the proper balance between full-length mRNA production and premature termination at protein-coding genes, but the underlying molecular mechanisms remain unclear. There is also evidence that Integrator subunits can function independently of the rest of the complex, e.g., as part of the DNA damage response<sup>41</sup> or to activate enhancers.<sup>21</sup>

Here, to identify regulatory subunits that enable the Integrator complex to be distinctly regulated in a locus-specific manner, we tested the effect of individually over-expressing each Integrator subunit in *Drosophila* cells. This revealed that IntS6 over-expression uniquely blocks Integrator function at a subset of protein-coding and long non-coding RNA genes. IntS6 can thus function in a dominant-negative manner, but IntS6 over-expression is not

sufficient to fully disable Integrator across the genome: nascent snRNAs continue to be processed, and the expression of many protein-coding genes remains attenuated by Integrator. The only loci affected by IntS6 over-expression are those where the PP2A phosphatase module is required for Integrator function, which is only a subset of gene loci bound by Integrator. We found that over-expressed IntS6 acts like a molecular sponge that titrates PP2A subunits and that IntS6 functions in a manner analogous to canonical PP2A regulatory B subunits. Altogether, these results show that the PP2A phosphatase module is critical at some, but not all genes for Integrator function and point to recruitment of PP2A via IntS6 as a tunable step that can control Integrator activity in a locus-specific manner.

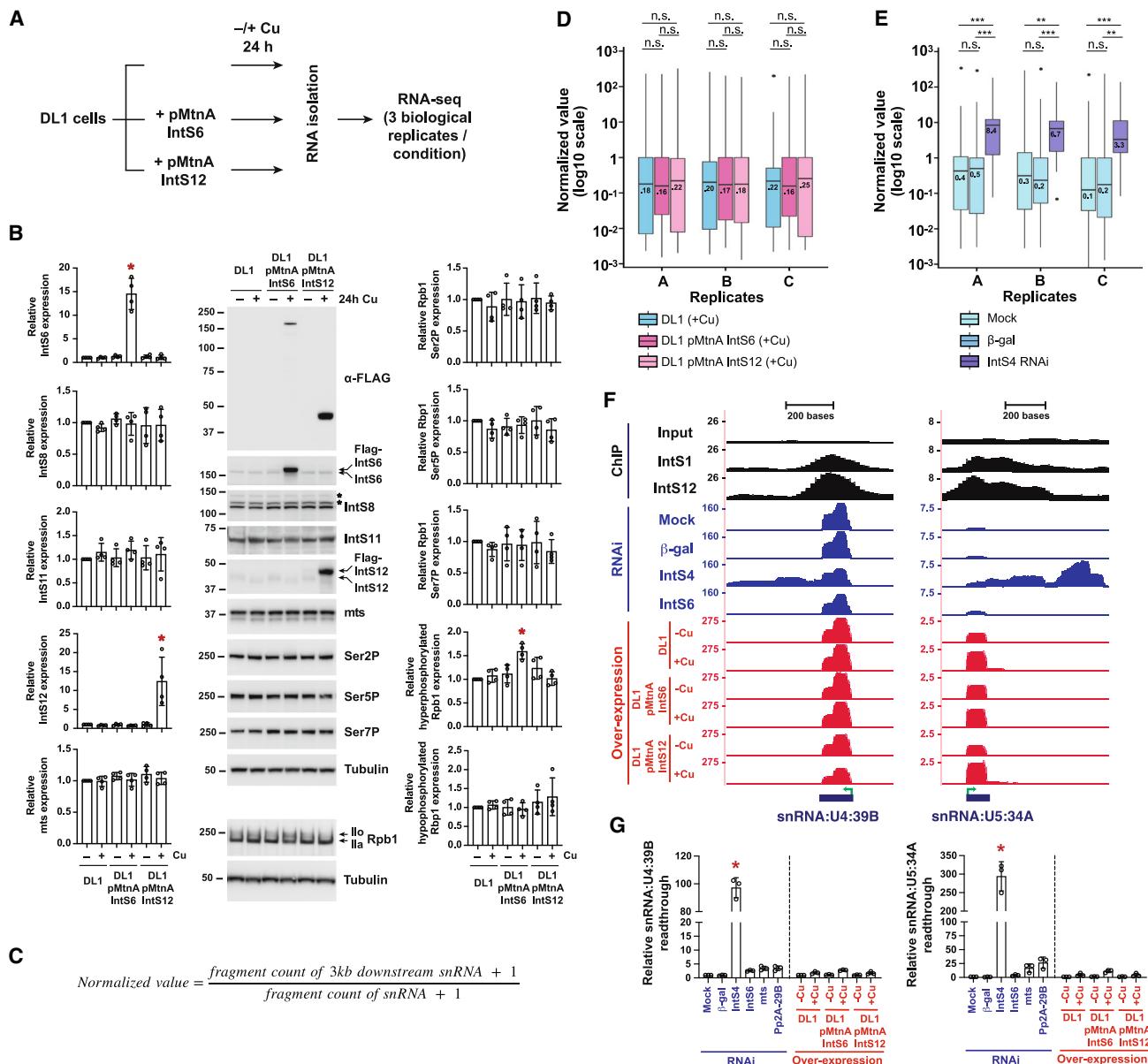
## RESULTS

### Over-expression of IntS6 uniquely inhibits Integrator function at a subset of protein-coding promoters

Cleavage of nascent snRNA transcripts by Integrator is critical for the production of functional snRNAs,<sup>9</sup> whereas cleavage of nascent mRNAs by Integrator triggers premature transcription termination coupled to nascent transcript degradation and attenuation of protein-coding gene expression.<sup>27,28</sup> Given these opposite effects on the ultimate gene outputs, we reasoned that the Integrator complex must be regulated in a distinct manner at protein-coding genes compared with snRNAs. To reveal the critical regulatory differences, we took advantage of a set of eGFP reporters that are driven by *Drosophila* protein-coding promoters (Figure 1A, left) and compared their regulation patterns with that of snRNA readthrough reporters that produce eGFP when Integrator fails to process the 3' end of the encoded snRNA (Figure 1A, right). Each of the examined protein-coding gene promoters, which included MtnA, Hemolectin (Hml), CG8620, anachronism (ana), and the small subunit of ribonucleoside reductase (RnrS), are normally attenuated by Integrator in *Drosophila* DL1 cells as eGFP mRNA expression increased upon treatment with double-stranded RNA (dsRNA) targeting IntS4, a scaffolding component of the Integrator endonuclease cleavage module,<sup>42,43</sup> relative to treatment with control ( $\beta$ -galactosidase [ $\beta$ -gal]) dsRNA (Figures 1B, left, and S1A–S1C). As expected, the U4:39B and U5:34A snRNA readthrough reporters were likewise sensitive to the depletion of IntS4 as demonstrated by a potent increase in eGFP levels due to Integrator not cleaving near the 3' box sequence (Figure 1B, left). By contrast, a reporter driven by the ubiquitin-63E (Ubi-p63E) promoter was not affected by IntS4 depletion (Figure 1B, left), consistent with prior RNA-seq results that showed endogenous Ubi-p63E mRNA levels are not regulated by Integrator.<sup>28</sup>

We next took advantage of the eGFP reporters and examined the functional effects of individually over-expressing a FLAG-tagged version of each Integrator subunit (IntS1–IntS14) (Figure S1D). Compared with over-expressing a control transgene that encodes a multicloning site (MCS), individual over-expression

the MtnA promoter. Total RNA was isolated, and northern blots (20  $\mu$ g/lane) were used to measure expression of each eGFP reporter mRNA. Representative blots are shown, and loading controls are provided in Figure S1C. RT-qPCR was used to quantify eGFP mRNA expression levels (middle). RNAi data were normalized to the mock samples, and over-expression data were normalized to the MCS samples. Data are shown as mean  $\pm$  SD, n  $\geq$  3. \*p < 0.05. See also Figures S1 and S2 and Table S5.



**Figure 2. Over-expression of IntS6 does not affect Integrator activity at endogenous *Drosophila* snRNA loci**

(A) Parental DL1 cells or DL1 cells stably maintaining IntS6 or IntS12 transgenes driven by the copper-inducible MtnA promoter were grown for 3 days. 500  $\mu$ M CuSO<sub>4</sub> was added for the last 24 h prior to total RNA isolation from three independent biological replicates. rRNA-depleted RNA-seq libraries were then generated, sequenced, and analyzed.

(B) Cell lines were seeded in 12-well plates ( $5 \times 10^5$  cells per well) and grown for 3 days. As indicated, a final concentration of 500  $\mu$ M CuSO<sub>4</sub> was added to cells for the last 24 h prior to harvesting total protein. Western blot analysis was then performed using antibodies that recognize FLAG, IntS6, IntS8, IntS11, IntS12, mts, Rbp1 phosphorylated at Ser2, Ser5, or Ser7, and Rbp1 (C-terminal domain). \* denotes non-specific band. Ilo denotes hyperphosphorylated Rbp1, whereas Ila denotes hypophosphorylated Rbp1.  $\alpha$ -Tubulin was used as a loading control. Subunit expression data were normalized to the parental DL1 cells without CuSO<sub>4</sub> treatment and are shown as mean  $\pm$  SD, n = 3. \*p < 0.05.

(C) To quantify readthrough transcription downstream of endogenous snRNAs, the levels of RNA-seq fragments that map to the 3 kb downstream of mature snRNA 3' ends were normalized to the levels of fragments that map to mature snRNA sequences.

(D and E) Normalized values of endogenous snRNA readthrough among (D) CuSO<sub>4</sub>-treated parental DL1 cells and DL1 cells stably maintaining IntS6 or IntS12 transgenes, and (E) DL1 cells subjected to mock, control ( $\beta$ -gal) dsRNA, or IntS4 dsRNA treatments. Center lines represent medians, boxes represent interquartile ranges (IQRs), and whiskers represent extreme data points within 1.5  $\times$  IQRs. Black points were outliers exceeding 1.5  $\times$  IQRs. p values were calculated by Wilcoxon signed-rank test. \*\*p < 0.01; \*\*\*p < 0.001; n.s., not significant.

(F) UCSC genome browser tracks depicting exemplar snRNA loci. IntS1 and IntS12 ChIP-seq profiles in DL1 cells (GEO: GSE114467) are shown in black. RNA-seq data generated from DL1 cells treated for 3 days with control ( $\beta$ -gal), IntS4, or IntS6 dsRNAs are shown in blue. RNA-seq data generated from parental

(legend continued on next page)

# Molecular Cell

## Article



of almost all Integrator subunits had no or minimal effect on the amounts of eGFP mRNA produced from the reporters (Figures 1B, right, and S1C). The notable exception was IntS6 over-expression, which caused significantly increased amounts of eGFP mRNA to be produced from the MtnA, Hml, and CG8620 promoters (Figure 1B, right). There was, however, no change in the outputs of the Ana, RnrS, or snRNA readthrough reporters upon IntS6 over-expression (Figure 1B, right). This suggested IntS6 over-expression may have a dominant-negative effect and be sufficient to disable Integrator function at a subset of protein-coding promoters while having no effect on the ability of Integrator to attenuate other protein-coding promoters or process nascent snRNAs. To confirm the uniqueness of the IntS6 over-expression effects, increasing amounts of plasmids that over-express IntS6, IntS8 (which forms part of the Integrator shoulder module<sup>12</sup>), or IntS12 (which contains a PHD finger and can interact with the Integrator backbone module<sup>44</sup>) were transfected into DL1 cells. Effects on the eGFP reporters were then examined (Figures S2A and S2B). A progressive increase in eGFP levels from the MtnA promoter was observed with titration of the IntS6 plasmid, whereas IntS8 and IntS12 over-expression had no or minimal effect even at the highest tested levels.

The 1,284 amino acid (aa) *Drosophila* IntS6 protein contains a conserved Von Willebrand factor type A (VWA) domain at its N terminus, a >250 aa region in the middle of the protein (aa 950–1,220) that is absent from IntS6 homologs in other species, and a conserved CTD that is observed in several additional proteins (Figure S2C). To define the key regions in *Drosophila* IntS6 responsible for the dominant-negative effect, FLAG-tagged IntS6 over-expression plasmids with deletions from the N or C terminus were generated (Figure S2C) and transfected into DL1 cells (Figure S2D). This revealed that the N terminus (aa 1–600) and the C terminus (aa 1,197–1,284) are each sufficient on their own to inhibit Integrator activity at the MtnA promoter (Figure S2E). Both functional regions of IntS6 contain conserved protein domains, and we further addressed whether the *Drosophila*-specific IntS6 region may play a role by testing the effect of over-expressing IntS6 homologs from human or zebrafish (Figure S2F). Human IntS6 and IntS6L (IntS6-like) were sufficient to inhibit Integrator activity at select promoters in *Drosophila* cells, arguing against a key role for the *Drosophila*-specific region of IntS6 (Figure S2G).

Depletion of IntS6 using RNAi interestingly revealed that this subunit was differentially required for Integrator function across the eGFP reporters (Figures 1B, left, and S1B). When compared with IntS4 depletion, IntS6 depletion had only a minimal effect on the output of the snRNA readthrough reporters (e.g., 226- vs. 11-fold increase in eGFP expression for the U4:39B reporter), whereas IntS4 and IntS6 depletion resulted in near equal increases in eGFP mRNA expression from the MtnA promoter (4.0- and 3.6-fold, respectively) (Figure 1B, left). Hence, we concluded that IntS6 varies from dispensable (or near dispensable) to essential for Integrator function in *Drosophila* depending on which reporter is examined.

### A subset of endogenous protein-coding genes normally attenuated by Integrator are upregulated upon IntS6 over-expression

To extend the results from the eGFP reporters to endogenous genes, we generated DL1 cell lines that stably maintain IntS6 or IntS12 transgenes driven by the copper-inducible MtnA promoter (Figure 2A). Parental DL1 cells and the inducible cell lines were grown in the presence of Cu<sup>2+</sup> for 24 h, which enabled robust induction of IntS6 and IntS12 proteins (15- and 12-fold, respectively) in the stable cell lines (Figures 2B, left and S3A). As a control, we verified that expressions of additional Integrator subunits (IntS8, IntS11, and mts) were unchanged upon Cu<sup>2+</sup> addition in all cell lines (Figure 2B, left). Total RNA from 3 biological replicates was isolated from each cell line to generate ribosomal RNA-depleted RNA-seq libraries (Figure 2A; Table S1), which were then analyzed to identify differentially expressed transcripts (Figure S3B).

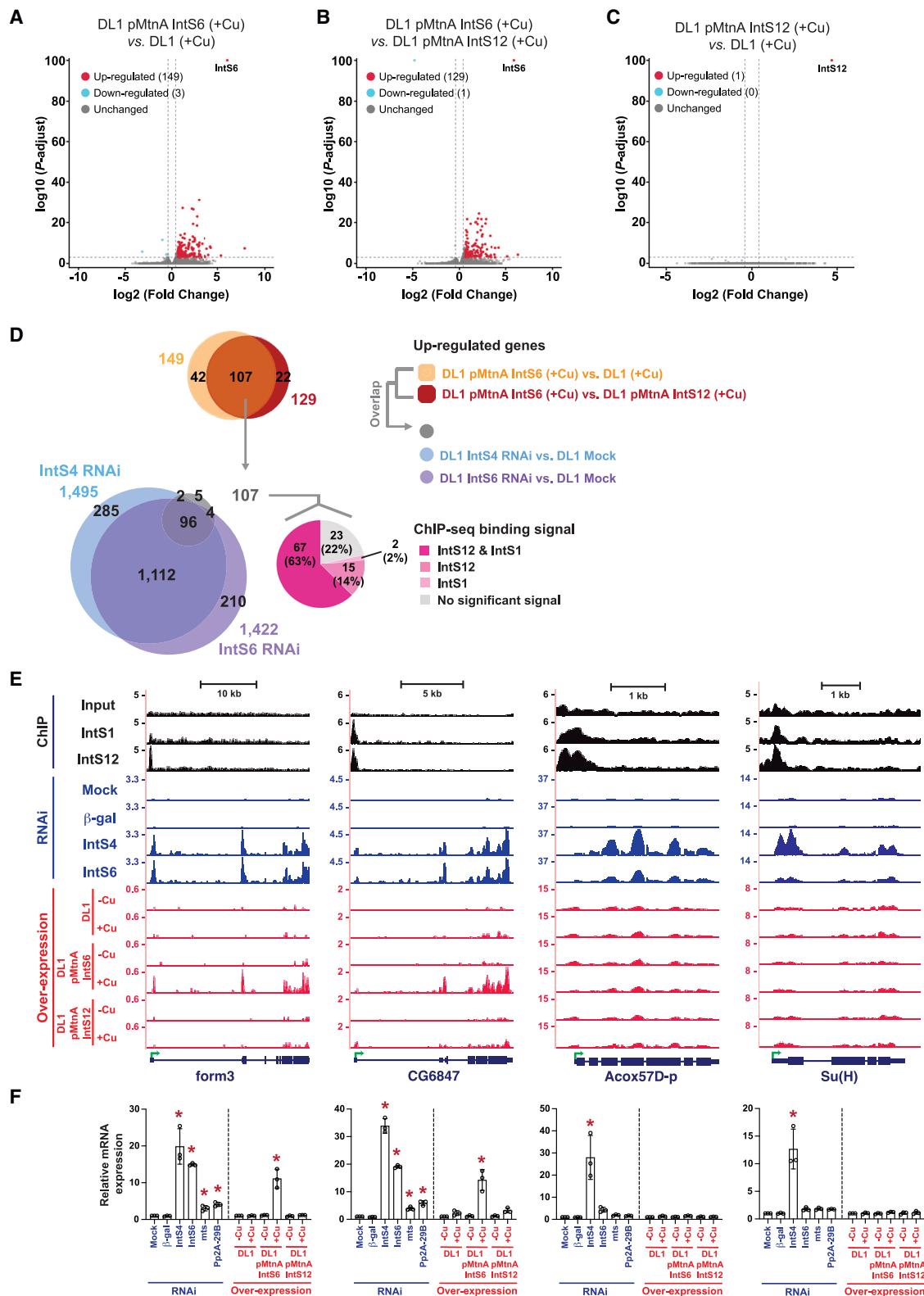
To first examine the effects on endogenous snRNA transcription termination, sequencing reads that mapped within 3 kb downstream of mature snRNA sequences were quantified and normalized to mature snRNA levels (Figure 2C). Over-expression of IntS6 or IntS12 had no significant effect on endogenous snRNA readthrough transcript levels (Figure 2D; Table S2), mirroring the results obtained with the eGFP reporters (Figure 1B, right). By contrast, depletion of IntS4 using RNAi resulted in the expected increase in endogenous snRNA readthrough (median increase > 30-fold) due to disabled Integrator endonuclease activity (Figure 2E; Table S2). These expression effects were readily observed at the endogenous U4:39B and U5:34A loci (Figure 2F) and confirmed using RT-qPCR (Figure 2G).

We next examined changes in protein-coding gene expression and found that over-expression of IntS6 (Figures 3A and 3B) but not IntS12 (Figure 3C) resulted in the upregulation of more than 100 genes with almost no genes being downregulated (fold-change > 1.5 and adjusted p < 0.001). Depending on whether the IntS6 over-expression data were compared with that of parental DL1 cells (Figure 3A) or the IntS12 over-expression cell line (Figure 3B), slightly different numbers of upregulated genes were identified, but the vast majority (107 genes) overlapped (Figure 3D; Table S3). These 107 genes are enriched in axon generation, protein folding, and heat response pathways (Figure S3C), are significantly longer in length than the average *Drosophila* gene (Figure S3D), and have a high Pol II pausing index in parental *Drosophila* DL1 cells as determined by precision run-on sequencing (PRO-seq) (Figures S3E and S3F).<sup>27</sup> Compared with genes that become upregulated in DL1 cells upon RNAi depletion of IntS4 or IntS6, genes affected by IntS6 over-expression normally have an even higher PRO-seq signal at the promoter regions coupled to low PRO-seq within the gene body (Figure S3F). Pol II thus typically initiates but fails to transition to productive elongation at these loci.

DL1 cells or DL1 cells stably maintaining copper-inducible IntS6 or IntS12 transgenes are shown in red. 500 μM CuSO<sub>4</sub> was added for 24 h as indicated. Green arrow, transcription start site (TSS).

(G) Readthrough downstream of snRNA transcripts was quantified using RT-qPCR. Data are shown as mean ± SD, n = 3. \*p < 0.05.

See also Figure S3 and Tables S1, S2, S4, S5, and S6.



**Figure 3. Over-expression of IntS6 blocks Integrator activity at a subset of endogenous *Drosophila* protein-coding genes**

(A–C) The magnitude of change in mRNA expression compared with statistical significance (adjusted p value) is shown as volcano plots. Endogenous mRNA expression levels upon IntS6 over-expression were compared with that in parental DL1 cells (A) or upon IntS12 over-expression (B). mRNA expression levels (legend continued on next page)

# Molecular Cell

## Article



Consistent with the PRO-seq observations, there is evidence that almost all of the 107 genes are directly attenuated by Integrator in DL1 cells under normal conditions (Figure 3D). IntS1 and/or IntS12 chromatin immunoprecipitation sequencing (ChIP-seq) peaks were detected at 84 of the 107 genes (78%) (Figures 3D and S3G–S3J), and 102 of these genes (95%) were upregulated (fold-change > 1.5 and adjusted p < 0.001) upon RNAi depletion of IntS4 and/or IntS6 in the parental DL1 cells (Figures 3D, S4A, and S4B; Table S4). There was, in fact, a strong positive correlation (~0.8) between the degree of upregulation observed for these genes upon IntS6 over-expression and that observed upon depletion of IntS4 (Figure S4C) or IntS6 (Figure S4D). Upon examining additional RNA-seq datasets, we found that the vast majority of the genes upregulated upon IntS6 over-expression (89/107 genes, 83%) were also upregulated upon IntS8 depletion (Figure S4E, left; Table S1).<sup>11</sup> In addition, attenuation of many of the 107 genes is clearly dependent on the IntS11 endonuclease, as their mRNA levels increased upon depletion of IntS11, and this could be rescued by expression of a wild-type IntS11 transgene but not by a catalytically dead (E203Q) IntS11 transgene (Figure S4E, right; Table S1).<sup>9,27</sup>

Using the form3 and CG6847 genes as examples, Integrator subunits are normally bound at their 5' ends in DL1 cells (Figure 3E, top), and the encoded mRNAs were upregulated when IntS4 or IntS6 were depleted by RNAi (Figure 3E, middle). These mRNAs were likewise upregulated when IntS6, but not IntS12, was over-expressed (Figure 3E, bottom), and RT-qPCR confirmed the expression changes at both the mRNA (Figure 3F) and pre-mRNA levels (Figure S5A). The latter was done using qPCR primers that amplify across exon-intron boundaries. Similar effects on transcriptional outputs were confirmed at additional exemplar loci (CG8547, Kal1, wun2, and wdp) as shown in Figures S5B–S5E. To further confirm the unique ability of IntS6 over-expression to inhibit Integrator activity, we generated DL1 cell lines that stably maintain IntS7, IntS13, or IntS14 transgenes driven by the copper-inducible MtnA promoter (Figure S5F). Each cell line was grown in the presence of Cu<sup>2+</sup> for 24 h to induce robust transgene expression (Figure S5F), but RT-qPCR did not detect any effect of these other over-expressed Integrator subunits on endogenous gene outputs (Figure S5G).

IntS6 over-expression thus uniquely resulted in the upregulation of a set of protein-coding genes normally attenuated by Integrator, but the vast majority of Integrator-regulated genes were unaffected by IntS6 over-expression. More than 1,100 genes became upregulated (fold-change > 1.5 and adjusted p < 0.001) when IntS4 or IntS6 was depleted using RNAi

(Figures 3D, S4A, and S4B; Table S4), and many of these genes are normally bound by IntS1 and/or IntS12, suggesting that they are direct Integrator targets (discussed further in Figure 6). For example, Integrator subunits are bound at the 5' ends of the Acox57D-p and Su(H) genes (Figure 3E, top), and these mRNAs were upregulated upon depletion of IntS4 (Figure 3E, middle), yet they were minimally affected by IntS6 over-expression at both the mRNA (Figures 3E, bottom and 3F) and pre-mRNA levels (Figure S6A). Similar expression patterns were observed at the endogenous Ana (Figure S6B) and RnrS loci (Figure S6C), mirroring the lack of effect that was observed when these promoters were tested in the eGFP reporter assay (Figure 1B). We thus conclude that IntS6 over-expression has a dominant-negative effect on Integrator function at only a subset of endogenous target loci.

### IntS6 over-expression blocks Integrator function by titrating PP2A

We next aimed to understand why only a subset of Integrator-regulated loci are affected by IntS6 over-expression. Recent cryo-electron microscopy (cryoEM) efforts characterizing the Integrator complex revealed that IntS6 is located far (>75 Å) from the IntS11 endonuclease active site and instead contacts the PP2A subunits (Figure 4A).<sup>12,14,18</sup> This led us to hypothesize that IntS6 over-expression may somehow affect the phosphatase module of the Integrator complex (in line with recent work<sup>13</sup>). PP2A typically functions as a trimer in which a regulatory B subunit (twins [tws], widerborst [wdb], Connector of kinase to AP-1 [Cka], or well-rounded [wrd] in *Drosophila*) enables the scaffolding (A) and catalytic (C) subunits (Pp2A-29B and microtubule star [mts], respectively, in *Drosophila*) to be targeted to a variety of substrates in cells.<sup>45–47</sup> Both the A (Pp2A-29B) and C (mts) subunits are present in the Integrator complex, but no canonical B subunits have been detected in association with Integrator (Figure 4A).<sup>11–14</sup> Eliminating the association of Integrator with PP2A subunits by depleting IntS8,<sup>11,12,38</sup> knocking out IntS6,<sup>13</sup> or mutating/truncating the N terminus of IntS8<sup>11,38</sup> has been shown to increase phosphorylation levels of the CTD of Rpb1, the largest subunit of Pol II. Consistent with these studies, we observed increased (~2-fold) hyperphosphorylated Rpb1 and increased Ser7 CTD phosphorylation when IntS6 or IntS8 was depleted (Figure S7A). We also noted that the levels of IntS6 and IntS8 proteins appear to be co-regulated, as IntS6 protein was depleted when a dsRNA targeting IntS8 was used and vice versa (Figure S7A).

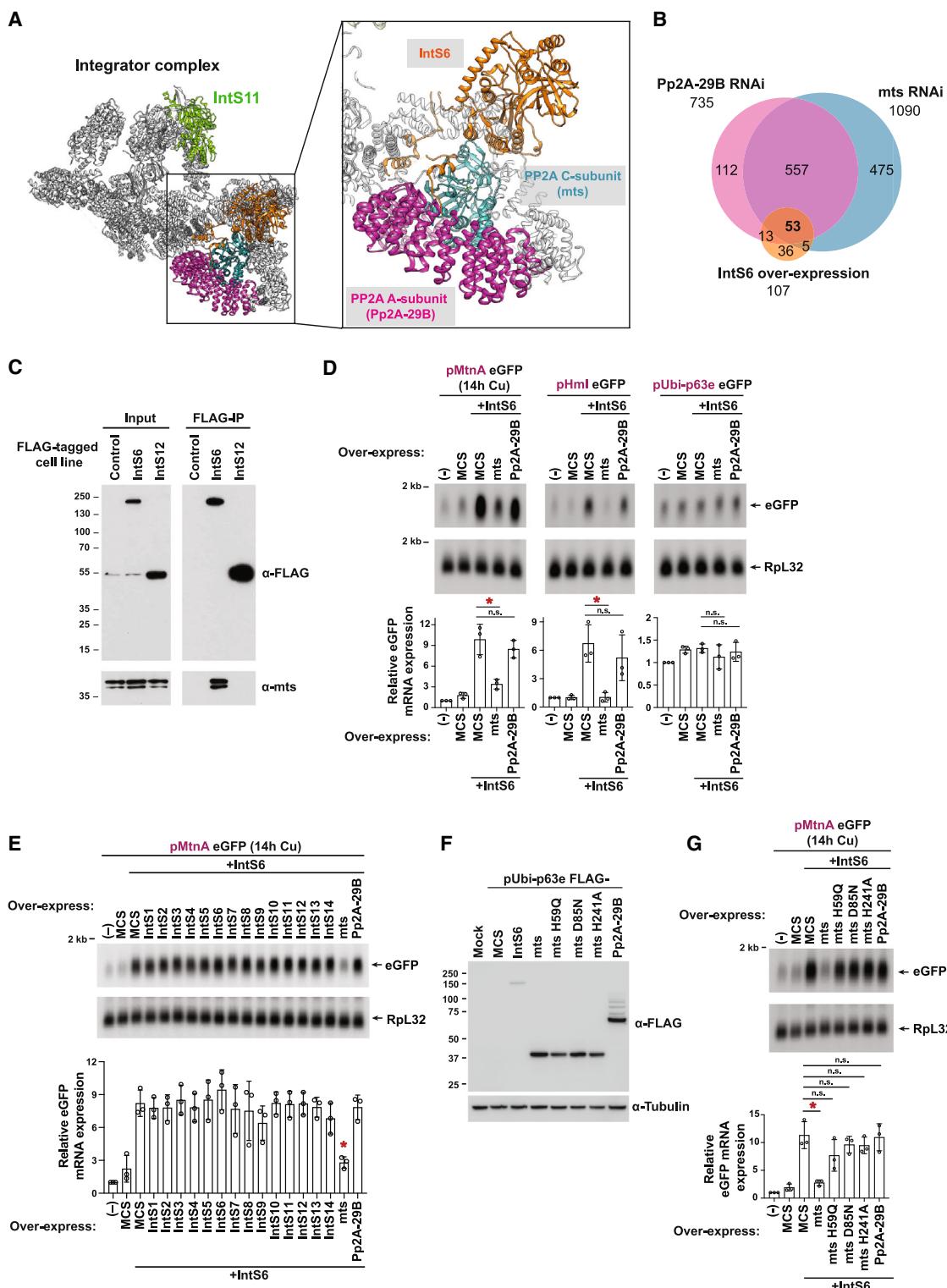
Upon over-expression of IntS6 in DL1 cells, there was also a statistically significant increase (~1.5-fold) in hyperphosphorylated Rpb1 levels, although no change was observed for any of

upon IntS12 over-expression were compared with parental DL1 cells (C). Threshold used to define differentially expressed mRNAs was |log<sub>2</sub>(fold-change)| > 0.585 and adjusted p < 0.001.

(D) The overlapping set of 107 protein-coding genes that were upregulated upon IntS6 over-expression were compared with the sets of genes upregulated upon RNAi depletion of IntS4 or IntS6 (left) and the sets of genes bound by IntS12 and/or IntS1 in DL1 cells in previously published (GEO: GSE114467) ChIP-seq experiments (right).

(E) UCSC genome browser tracks depicting example protein-coding loci that are (form3, CG6847) or are not (Acox57D-p, Su(H)) affected by IntS6 over-expression. IntS1 and IntS12 ChIP-seq profiles in DL1 cells (GEO: GSE114467) are shown in black. RNA-seq data generated from DL1 cells treated for 3 days with control (β-gal), IntS4, or IntS6 dsRNAs are shown in blue. RNA-seq data generated from parental DL1 cells or DL1 cells stably maintaining copper-inducible IntS6 or IntS12 transgenes are shown in red. 500 μM CuSO<sub>4</sub> was added for 24 h as indicated. Green arrow, transcription start site (TSS).

(F) Expression of the indicated mRNAs (order as in E) was quantified using RT-qPCR. Data are shown as mean ± SD, n = 3. \*p < 0.05. See also Figures S3–S6 and Tables S1, S3, S4, and S5.



**Figure 4. IntS6 over-expression titrates the catalytic subunit of PP2A and causes it to be limiting for Integrator activity**

(A) Cryo-EM structure (PDB: 7PKS) of the human Integrator complex,<sup>14</sup> highlighting the positions of the RNA endonuclease IntS11 (green), IntS6 (orange), and PP2A subunits (teal and pink). There are direct contacts between IntS6 and the PP2A subunits.

(legend continued on next page)

the individual CTD phosphorylation sites (Ser2, Ser5, or Ser7) examined (Figure 2B, right). To address whether Integrator phosphatase activity may indeed be disrupted, we asked whether IntS6 over-expression altered the expression of genes normally controlled by PP2A subunits. Indeed, a significant overlap was observed when comparing the list of 107 upregulated genes with those that were upregulated when the A (Pp2A-29B) or C (mts) subunit was depleted using RNAi in DL1 cells (Figure 4B; Tables S3 and S4). We thus hypothesized that IntS6 over-expression may sponge or titrate PP2A subunits from the rest of the Integrator complex. Consistent with such a model, the PP2A C (mts) subunit was co-immunoprecipitated with over-expressed FLAG-tagged IntS6, but not with over-expressed IntS12 (Figure 4C). This result nicely mirrors our observation that only IntS6 over-expression (not IntS12) alters Integrator complex function (Figures 1, 2, and 3).

Prior work in *S. cerevisiae* has shown that the PP2A A subunit is expressed in excess over the C subunit,<sup>48</sup> suggesting a possible role for stoichiometry in the control of PP2A function. If PP2A subunits are indeed limiting in DL1 cells, we reasoned that over-expressing the C (mts) and/or A (Pp2A-29B) subunits should be sufficient to reverse the gene expression changes caused by IntS6 over-expression. Increasing the expression of the C (mts) subunit, but not the A (Pp2A-29B) subunit, was sufficient to reduce the amount of eGFP mRNA produced from the MtnA (Figure 4D, left) and Hml promoter reporters (Figure 4D, middle) to near wild-type levels, indicating a restoration of Integrator activity at these promoters. As expected, there was no change in the output of the control Ubi-p63e promoter, which is not regulated by Integrator (Figure 4D, right). Over-expression of other individual Integrator subunits (Figures 4E and S7B) or catalytically dead PP2A C (mts) subunits that bear mutations that disrupt metal binding (H59Q, D85N, or H241A)<sup>49,50</sup> all were unable to rescue the gene expression changes (Figures 4F, 4G, and S7C). Collectively, these data indicate that the phosphatase activity of the PP2A C (mts) subunit can be limiting for Integrator function and that over-expression of IntS6 blocks the ability of PP2A to function at specific gene loci.

(B) DL1 cells were treated for 3 days with control ( $\beta$ -gal), Pp2A-29B, or mts dsRNAs, and RNA-seq data were generated. The sets of endogenous genes upregulated upon Pp2A-29B or mts depletion (fold-change > 1.5 and adjusted p value < 0.001) were compared with the set of 107 genes that were upregulated upon over-expression of IntS6.

(C) Parental DL1 cells (control) and DL1 cells stably maintaining inducible FLAG-tagged IntS6 or IntS12 transgenes were treated with 500  $\mu$ M CuSO<sub>4</sub> for 24 h to induce transgene expression. Immunoprecipitation (IP) using anti-FLAG resin was then performed. Western blots of input nuclear extracts (left, 0.16% input for FLAG and 0.25% for mts) and IP (right, 0.2% eluate for FLAG and 25% for mts) are shown.

(D and E) DL1 cells were co-transfected with 300 ng of eGFP reporter plasmid, 100 ng of IntS6 over-expression plasmid (driven by the Ubi-p63e promoter), and 100 ng of the indicated PP2A subunit (D) or Integrator subunit (E) over-expression plasmid (driven by the Ubi-p63e promoter). Empty vector (pUb 3xFLAG MCS) was added as needed so that 500 ng DNA was transfected in all samples. CuSO<sub>4</sub> was added for the last 14 h only when measuring eGFP production from the MtnA promoter. Northern blots (20  $\mu$ g/lane) were used to quantify expression of each eGFP reporter mRNA. Representative blots are shown, and Rpl32 mRNA was used as a loading control. Data are shown as mean  $\pm$  SD, n = 3. \*p < 0.05; n.s., not significant.

(F) DL1 cells were transfected with 500 ng of the indicated FLAG-tagged expression plasmids, and total protein was harvested after 48 h. A plasmid containing a multi-cloning site (MCS) was used as a control. Western blot analysis using an antibody that recognizes FLAG was used to confirm expression of individual Integrator/PP2A subunits.  $\alpha$ -Tubulin was used as a loading control.

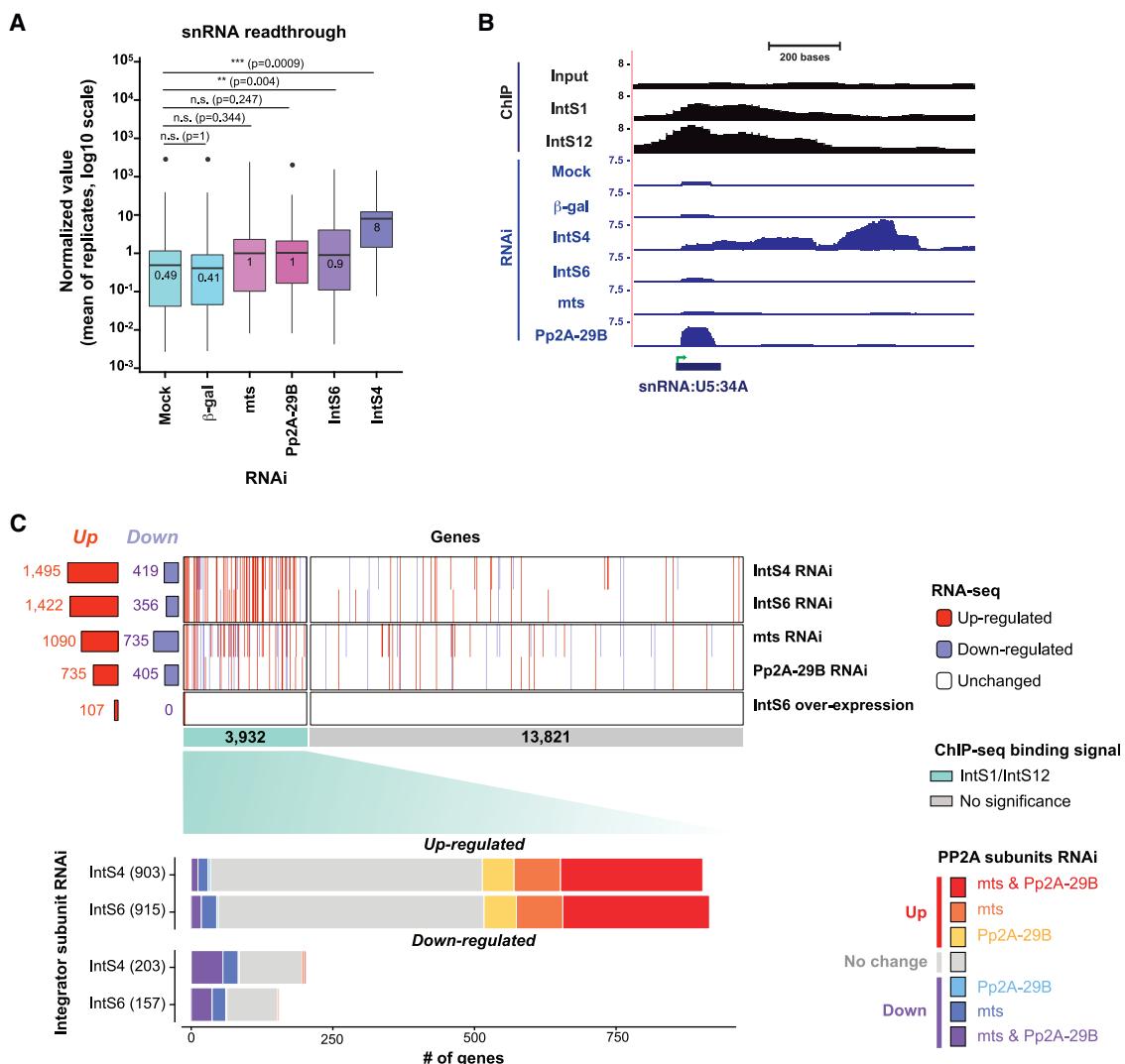
(G) DL1 cells were co-transfected with 300 ng of eGFP reporter plasmid, 100 ng of IntS6 over-expression plasmid (driven by the Ubi-p63e promoter), and 100 ng of the indicated PP2A subunit over-expression plasmid (driven by the Ubi-p63e promoter). CuSO<sub>4</sub> was added for the last 14 h, and northern blots (20  $\mu$ g/lane) were used to quantify expression of the eGFP reporter mRNA. Representative blots are shown, and Rpl32 mRNA was used as a loading control. Data are shown as mean  $\pm$  SD, n = 3. \*p < 0.05; n.s., not significant.

See also Figure S7 and Tables S5 and S6.

### The phosphatase module of Integrator is critical at only a subset of Integrator-regulated loci

Given that IntS6 over-expression only affects the outputs of a subset of Integrator-regulated eGFP reporters (Figure 1) and endogenous genes (Figures 2 and 3), the titration model suggests that the PP2A phosphatase activity is differentially required across loci for regulation by the Integrator complex. To test this hypothesis, RNAi was used to deplete the PP2A A (Pp2A-29B) or C (mts) subunits (Figure S1A), and we then examined the effect on the outputs of the eGFP reporters (Figure 1B, left). eGFP expression from the MtnA, Hml, and CG8620 promoters all increased upon depletion of PP2A subunits, consistent with a critical role for phosphatase activity in enabling Integrator function at these promoters. In stark contrast, there was no or minimal change in the outputs of the Ana, RnrS, and snRNA readthrough reporters upon depletion of PP2A subunits (Figure 1B, left). These results perfectly mirror the effects observed with IntS6 over-expression: only at promoters where PP2A subunits are required (MtnA, Hml, and CG8620) does IntS6 over-expression result in increased eGFP reporter expression (Figure 1B, right). Promoters unaffected by the depletion of PP2A subunits were further only minimally affected by the depletion of IntS6 (Figure 1B, left), with modest but reproducible depressive effects observed that may be due to co-depletion of other Integrator subunits, including IntS8 (Figure S7A) and perhaps IntS5.<sup>38</sup>

We next examined individual endogenous snRNA (Figure 2G) and protein-coding genes (Figures 3F, S5B–S5E, S6B, and S6C) using RT-qPCR and again observed that only genes that require PP2A subunits for Integrator function were significantly affected by IntS6 over-expression. For example, the Integrator phosphatase module is not required at the endogenous Acox57D-p and Su(H) genes as there was only a minimal increase in expression of these mRNAs upon depletion of the PP2A A (Pp2A-29B) or C (mts) subunits, and these loci were unaffected when IntS6 was over-expressed (Figure 3F). By contrast, PP2A subunits were required for Integrator function at the form3 and CG6847 loci, and the expression of these genes was increased with IntS6 over-expression (Figure 3F). Mirroring the results with the eGFP reporter genes (Figure 1B, left), the



**Figure 5. The phosphatase module is differentially required for Integrator activity across the genome**

(A) Means of normalized values of endogenous snRNA readthrough. Readthrough values (see Figure 2C) were calculated from RNA-seq data of DL1 cells subjected to a mock treatment or treatment with β-gal, mts, Pp2A-29B, IntS6, or IntS4 dsRNAs (3 independent biological replicates). Center lines represent medians, boxes represent interquartile ranges (IQRs), and whiskers represent extreme data points within 1.5× IQRs. Black points were outliers exceeding 1.5× IQRs. p values were calculated by Wilcoxon signed-rank test. \*\*p < 0.01; \*\*\*p < 0.001; n.s., not significant.

(B) ChIP-seq and RNA-seq tracks at the U5:34A snRNA locus. IntS1 and IntS12 ChIP-seq profiles in DL1 cells (GEO: GSE114467) are shown in black. RNA-seq data generated from DL1 cells treated for 3 days with control (β-gal), IntS4, IntS6, mts, or Pp2A-29B dsRNAs are shown in blue. Green arrow, transcription start site (TSS).

(C) RNA-seq was used to define genes that were up- or down-regulated ( $|\log_2(\text{fold-change})| > 0.585$  and adjusted  $p < 0.001$ ) upon IntS4, IntS6, mts, or Pp2A-29B depletion using RNAi or upon IntS6 over-expression (top). These gene lists were then stratified by ChIP-seq data that identified 3,932 protein-coding genes with peaks of IntS1 and/or IntS12 binding located  $\pm 1$  kb of gene bodies in DL1 cells (green, middle). For genes bound by Integrator subunits and differentially regulated upon IntS4 or IntS6 depletion, the effect of depleting PP2A subunits on their expression is graphed (bottom).

See also Tables S1, S2, S3, and S4.

expression of PP2A-independent endogenous loci was only modestly increased upon IntS6 depletion (Figures 3F and S6B).

To determine how broadly the phosphatase activity is required for Integrator function across the genome, we first examined readthrough transcription downstream of all endogenous snRNA transcripts. Depletion of PP2A C (mts) or A (Pp2A-29B) resulted in modest effects (compared with IntS4 depletion) that were not statistically significant (Figures 5A and 5B). Note that the levels

of mature U5:34A snRNA increased upon the depletion of Pp2A-29B (Figure 5B), but similar changes were not observed at the vast majority of snRNA loci, suggesting an idiosyncratic effect particular to this snRNA (Table S2). We then used ChIP-seq data from DL1 cells<sup>27</sup> to identify 3,932 protein-coding genes with peaks of IntS1 and/or IntS12 binding located  $\pm 1$  kb of gene bodies (Figures 5C, middle, and S3G–S3J). Most (>70%) of these peaks were close to TSSs (Figures S3H and S3I), and we

# Molecular Cell

## Article



reasoned that these 3,932 genes represent loci where Integrator may have direct effects. We thus determined if their expression changed upon the depletion of IntS4, IntS6, PP2A C (mts), or PP2A A (Pp2A-29B) (Figure 5C, top; Table S4). Consistent with Integrator catalyzing premature transcription termination, most of the protein-coding genes that changed were upregulated (fold-change > 1.5 and adjusted p < 0.001) upon the depletion of IntS4 or IntS6 (Figure 5C, top). Of the 3,932 genes bound by Integrator, 903 and 915 were upregulated upon IntS4 or IntS6 depletion, respectively, and ~40% of these genes were also upregulated upon the depletion of PP2A C (mts) and/or A (Pp2A-29B) (Figure 5C, bottom). This suggests that the phosphatase activity of Integrator may be required at ~40% of the protein-coding genes attenuated by Integrator, whereas it is dispensable at many other genes that are nonetheless still controlled by the complex. This result helps explain why IntS6 over-expression had no effect on most Integrator-regulated genes (Figure 3D). The 107 protein-coding genes that were upregulated by IntS6 over-expression may represent those genes that are most sensitive to tuning by PP2A, and indeed, these loci had higher Pol II pausing indexes (Figures S3E and S3F).

### IntS6 behaves like a canonical PP2A B subunit

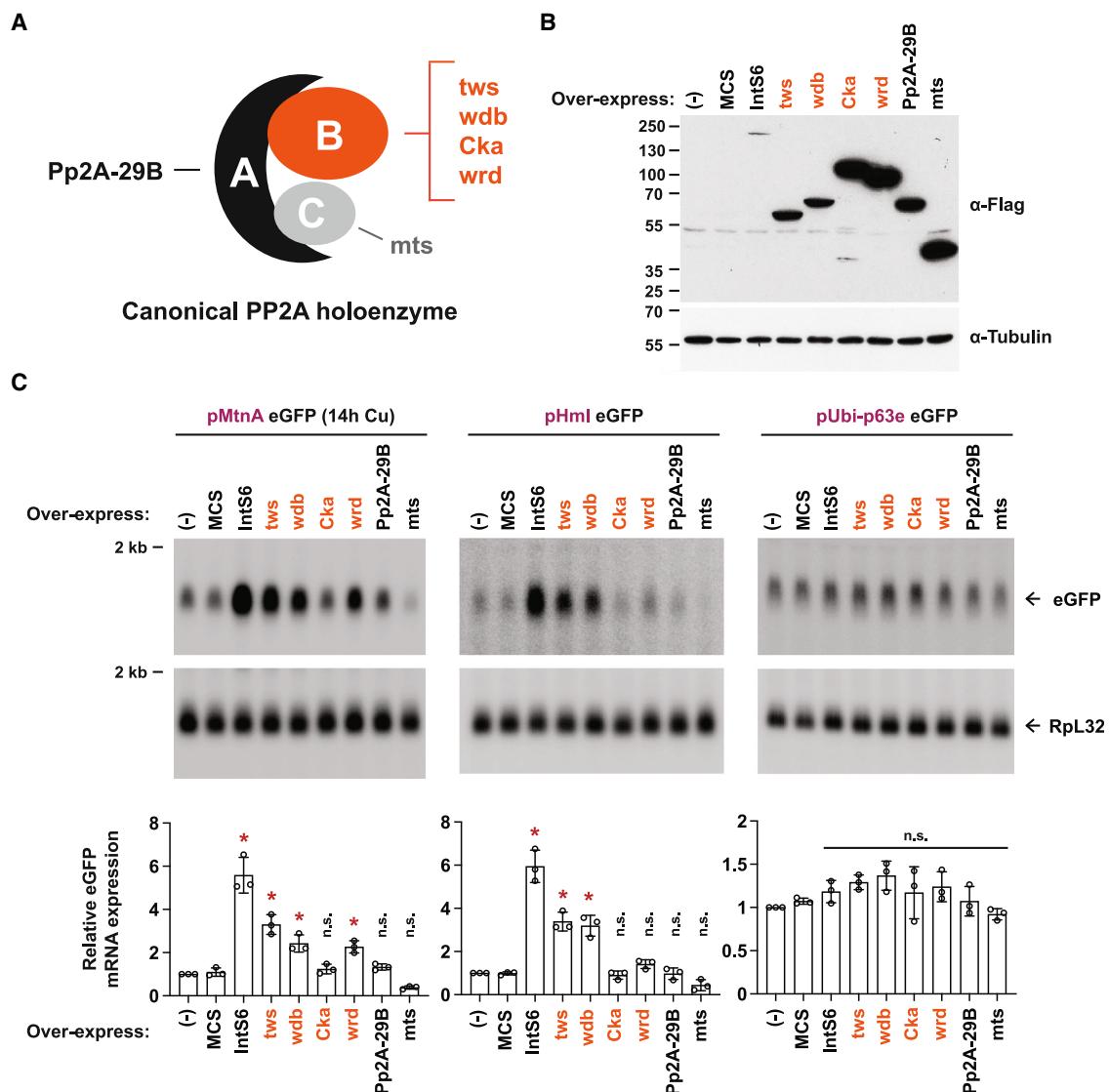
The PP2A holoenzyme typically functions as a heterotrimer consisting of an A, B, and C subunit (Figure 6A), but no canonical B subunits have been detected in the Integrator complex.<sup>11–14</sup> This has made it somewhat unclear how the Integrator phosphatase activity is controlled and regulated. Because IntS6 binds the PP2A A and C subunit in Integrator (Figure 4A) and IntS6 over-expression is sufficient to titrate PP2A activity (Figure 4D), we reasoned that IntS6 may be a previously unappreciated regulatory B subunit. If true, this model predicts that the over-expression of canonical B subunits (tws, wdb, Cka, or wrd in *Drosophila*) may likewise be sufficient to titrate the PP2A C subunit and inhibit the attenuation activity of Integrator. We thus individually over-expressed FLAG-tagged versions of each PP2A B subunit (Figure 6B) and examined the functional effects on the eGFP reporters (Figure 6C). Over-expression of the canonical B subunits tws and wdb did indeed lead to increased eGFP expression from the MtnA and Hml promoters, but not from the control Ubi-p63e promoter that is not regulated by Integrator (Figure 6C). It should be noted that the effects were not as large as those observed with IntS6 over-expression, despite the FLAG-tagged versions of tws and wdb being expressed at higher levels than FLAG-tagged IntS6 (Figure 6B). In total, these results indicate that IntS6 behaves, at least in some ways, similar to canonical B subunits and that alterations in the levels of B subunits not normally associated with Integrator can still affect the efficiency of select premature transcription termination events.

### DISCUSSION

The Integrator complex has RNA endonuclease and protein phosphatase activities, but it has remained unclear if the coordinated action of both activities is required for complex function or if they can be independently harnessed in a locus-specific manner. Here, using reporter genes and transcriptomics, we revealed that (1) the phosphatase module is functionally required only at a sub-

set of *Drosophila* genes that are regulated by Integrator and (2) that the ability of the PP2A catalytic subunit to act as part of Integrator can be tuned by the levels of IntS6 and canonical PP2A regulatory B subunits. Integrator regulates transcription elongation (Figure 7A), and on one hand, we identified a set of protein-coding genes that became potently de-attenuated when IntS4 (a component of the endonuclease module), IntS6, or PP2A subunit levels were modulated, indicating that the endonuclease and phosphatase modules are both critical for Integrator function at these genes (Figure 7B, left). Meanwhile, cleavage of nascent snRNAs and attenuation of many other protein-coding genes required the Integrator endonuclease module but were largely unaffected by modulation of IntS6 or PP2A subunit levels. This suggests that the phosphatase module plays little or perhaps even no functional role at these loci (Figure 7B, right). Prior work suggested the phosphatase module is broadly necessary for Integrator function across *Drosophila* protein-coding genes and snRNAs,<sup>11</sup> which contrasts with the locus specificity we observed. Most of the prior contrasting conclusions were based on either (1) depleting IntS8 using RNAi, which we found to also deplete IntS6 protein (Figure S7A) and possibly other closely associated Integrator subunits (e.g., IntS5, as was shown recently in human cells<sup>38</sup>), or (2) mutating a 4 aa region in IntS8 that binds the Pp2A-29B subunit. Direct roles for PP2A subunits themselves were only previously examined at 4 *Drosophila* protein-coding genes after RNAi depletion of PP2A subunits or treatment with chemical PP2A inhibitors.<sup>11</sup> When the catalytic subunit of PP2A, IntS6, or IntS8 were individually depleted in human cell lines, no or minimal effect on transcription termination of exemplar snRNA genes was observed,<sup>12,13</sup> which mirrors the results described here.

Structural studies using cryoEM have made it increasingly clear that Integrator contains discrete endonuclease, phosphatase, and auxiliary modules that each bind around a core set of Integrator subunits.<sup>12,14,18,42,51–53</sup> Overall complex integrity is disrupted upon the depletion of a core subunit (e.g., IntS2), whereas the depletion of module-specific subunits, including IntS6, does not appreciably affect the binding of other modules.<sup>11,13</sup> Our functional data support this concept of Integrator complex modularity as we found that altering IntS6 or PP2A sub-unit levels had no effect on endonuclease activity at many gene loci (e.g., snRNAs continued to be efficiently processed at their 3' ends). Why then is the Integrator phosphatase module required at a subset of loci? It has been recently shown that PP2A can dephosphorylate components of the Pol II complex, e.g., Spt5 or the CTD of Pol II itself, to antagonize kinases that stimulate Pol II pause release and productive elongation.<sup>11–13</sup> Based on our data, it appears that the order by which the Integrator RNA endonuclease and protein phosphatase activities act can be critical. In particular, at a subset of genes, the Integrator phosphatase module must act prior to or, at minimum, simultaneously with the IntS11 endonuclease to slow/pause Pol II and thereby enable cleavage of the nascent RNA (Figure 7B, left). In the absence of phosphatase activity, the temporal window of opportunity for IntS11 to cleave may be too short, and/or IntS11 may fail to achieve an active conformation that accommodates RNA.<sup>12,14</sup> The Pol II juggernaut<sup>2</sup> is hence unable to be stopped and instead elongates to produce full-length mRNAs. Given that RNA cleavage by IntS11 is presumably a non-reversible step, modulating



**Figure 6. Over-expression of canonical PP2A B subunits can inhibit Integrator activity**

(A) Schematic of the canonical *Drosophila* PP2A holoenzyme that consists of a scaffolding A subunit (Pp2A-29B), a catalytic subunit (mts), and a variable regulatory B subunit (tws, wdb, Cka, or wrd).

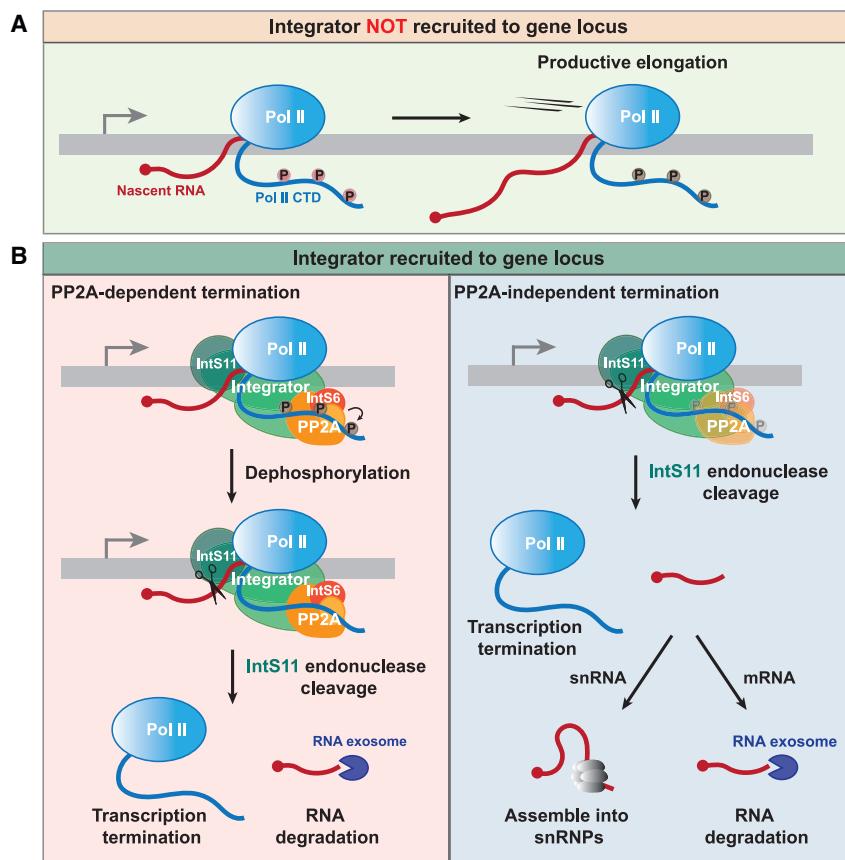
(B) DL1 cells were transfected with 500 ng of the indicated FLAG-tagged expression plasmid, and total protein was harvested after 48 h. A plasmid containing a multi-cloning site (MCS) was used as a control. Western blot analysis using an antibody that recognizes FLAG was used to confirm expression of each subunit. α-Tubulin was used as a loading control.

(C) DL1 cells were co-transfected with 400 ng of eGFP reporter plasmid and 100 ng of the indicated PP2A subunit over-expression plasmid (driven by the Ubi-p63e promoter). CuSO<sub>4</sub> was added for the last 14 h only when measuring eGFP production from the MtnA promoter. Northern blots (20 μg/lane) were used to quantify expression of each eGFP reporter mRNA. Representative blots are shown, and Rpl32 mRNA was used as a loading control. Data are shown as mean ± SD, n = 3. \*p < 0.05; n.s., not significant.

See also Tables S5 and S6.

PP2A activity and thus the IntS11 temporal window of opportunity represents an important way that the efficiency of premature termination events can be tuned. Here, we found that changes in protein levels can modulate the activity of the Integrator phosphatase module, and we propose that post-translational modifications (e.g., on IntS6 or the PP2A subunits) or the action of additional binding partners likely also tunes Integrator phosphatase activity.

At genes regulated by Integrator in a manner independent of the phosphatase module (Figure 7B, right), there are presumably alternative mechanisms that enable a sufficiently long temporal window of opportunity for IntS11 to cleave. The exact phosphorylation status of the Pol II CTD when Integrator is acting at these loci remains unclear, however. Ser7 phosphorylation has been implicated as being critical for Integrator recruitment to snRNAs,<sup>54–56</sup> and a conserved but relatively degenerate 3' box



**Figure 7. Integrator can catalyze transcription termination via PP2A-dependent and -independent mechanisms**

(A) In the absence of Integrator, Pol II is able to productively elongate.

(B) Integrator recruitment facilitates transcription termination. (Left) At some protein-coding genes, the Integrator phosphatase module must act prior to or, at minimum, simultaneously with the IntS11 endonuclease to enable cleavage of the nascent RNA, which is subsequently degraded by the RNA exosome. (Right) By contrast, the phosphatase module is dispensable for Integrator function at snRNA and many other protein-coding gene loci. Cleavage by IntS11 enables stable snRNAs to be produced and prematurely terminated mRNAs to be rapidly degraded. The exact CTD phosphorylation status at these genes when Integrator is acting remains to be determined.

sequence located 9–19 nt downstream of the 3' ends of mature snRNA transcripts is required for Integrator cleavage.<sup>57</sup> Details of the underlying molecular mechanism unfortunately remain poorly understood, and sequences resembling a 3' box are not present at most protein-coding genes attenuated by Integrator. Instead, nucleosome occupancy<sup>32</sup> or redundant phosphatases may act to stall or pause Pol II near protein-coding promoters, e.g., protein phosphatase 1 (PP1), which is known to help slow Pol II near polyadenylation signals.<sup>58–61</sup> Differences in Pol II speed/dynamics caused by histone modifications or RNA sequence/structure may also be at play.

Due to its high abundance, PP2A dephosphorylates many important substrates in eukaryotes and participates in many signaling cascades (for review, see Seshacharyulu et al.,<sup>45</sup> Xu et al.,<sup>46</sup> and Janssens and Goris<sup>47</sup>). We nonetheless found that the expression of the PP2A catalytic subunit (mts) can be limiting for Integrator function. Increasing the levels of IntS6 or canonical regulatory B subunits was sufficient to titrate PP2A subunits and block Integrator, pointing to a critical role for stoichiometry in controlling PP2A function and Integrator activity. Consistent with this idea, PP2A regulatory B subunits are differentially expressed across tissues and developmental time, and IntS6 is known to be lost or downregulated in several types of human cancers.<sup>62,63</sup> We thus propose that modulating the expression, localization, or modification status of B subunits (including IntS6) may be used by cells to couple rapid dephosphorylation of protein substrates to longer-term transcriptional changes via

the modulation of Integrator activity. It is also noteworthy that IntS6 underwent a gene duplication in many mammals, including humans (but not *Drosophila*), and it will be interesting in the future to determine whether IntS6 and IntS6L modulate PP2A activity in distinct ways and/or at distinct sets of Integrator-regulated genes.

In summary, the Integrator complex controls the fates of many nascent RNAs in metazoans, and our work has revealed

key insights into how its functional modules, especially the IntS6/PP2A phosphatase module, can be differentially employed across gene loci. Our results clarify that the Integrator RNA endonuclease and protein phosphatase activities do not always have to be coordinated with one another but instead can act independently (consistent with recent work<sup>38</sup>). Modulation of phosphatase activity by IntS6 is sufficient to disable Integrator at some loci, and we anticipate that there will be additional mechanisms that tune Integrator in a gene-specific manner to meet a cell's ever-changing transcriptional needs.

#### Limitations of the study

It remains unclear if all the Integrator complex modules are normally recruited as a group *in vivo* or if they can be selectively assembled at a gene locus. In addition, the exact characteristics that cause a gene locus to require the phosphatase module for Integrator activity need to be defined. Future work will be required to clarify if IntS6 over-expression affects PP2A activity beyond the Integrator complex and define how the IntS6-PP2A interaction is naturally modulated, including in species beyond *Drosophila*, to tune Integrator activity at specific loci *in vivo*.

#### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines
- METHOD DETAILS
  - Expression plasmids
  - Double-stranded RNAs
  - Generation of stable cell lines
  - Transfections, RNAi, and RNA isolation
  - Northern blotting
  - Antibody production
  - Western blotting
  - RT-qPCR
  - Immunoprecipitation
  - RNA-seq library generation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RNA-seq analysis
  - ChIP-seq analysis
  - PRO-seq analysis
  - Statistical tests

#### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2023.10.035>.

#### ACKNOWLEDGMENTS

We thank Veerle Janssens and members of the Wilusz and Yang labs for helpful discussions. This paper is dedicated to the memory of our colleague Deirdre Tatomer, who passed away while this work was ongoing. Supported by National Institutes of Health grant R35-GM119735 (to J.E.W.), Cancer Prevention & Research Institute of Texas grant RR210031 (to J.E.W.), and National Natural Science Foundation of China grant 31925011 (to L.Y.). J.E.W. is a CPRIT Scholar in Cancer Research.

#### AUTHOR CONTRIBUTIONS

D.C.T. and J.E.W. conceived the project. R.F., D.L., A.P.S., M.T., C.J.F., M.S.M.-F., M.C.M., and D.C.T. performed experiments and analyzed data. S.-N.Z., X.-K.M., and L.Y. analyzed the RNA-seq, PRO-seq, and ChIP-seq results. R.F., S.-N.Z., L.Y., and J.E.W. wrote the manuscript with input from the other authors.

#### DECLARATION OF INTERESTS

J.E.W. serves as a consultant for Laronde.

Received: March 6, 2023

Revised: October 12, 2023

Accepted: October 25, 2023

Published: November 22, 2023

#### REFERENCES

1. Schier, A.C., and Taatjes, D.J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* 34, 465–488. <https://doi.org/10.1101/gad.335679.119>.
2. Proudfoot, N.J. (2016). Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science* 352, aad9926. <https://doi.org/10.1126/science.aad9926>.
3. Cramer, P. (2019). Organization and regulation of gene transcription. *Nature* 573, 45–54. <https://doi.org/10.1038/s41586-019-1517-4>.
4. Mendoza-Figueroa, M.S., Tatomer, D.C., and Wilusz, J.E. (2020). The integrator complex in transcription and development. *Trends Biochem. Sci.* 45, 923–934. <https://doi.org/10.1016/j.tibs.2020.07.004>.
5. Welsh, S.A., and Gardini, A. (2023). Genomic regulation of transcription and RNA processing by the multitasking Integrator complex. *Nat. Rev. Mol. Cell Biol.* 24, 204–220. <https://doi.org/10.1038/s41580-022-00534-2>.
6. Sabath, K., and Jonas, S. (2022). Take a break: transcription regulation and RNA processing by the Integrator complex. *Curr. Opin. Struct. Biol.* 77, 102443. <https://doi.org/10.1016/j.sbi.2022.102443>.
7. Kirstein, N., Gomes dos Santos, H., Blumenthal, E., and Shiekhattar, R. (2021). The Integrator complex at the crossroad of coding and noncoding RNA. *Curr. Opin. Cell Biol.* 70, 37–43. <https://doi.org/10.1016/j.ceb.2020.11.003>.
8. Pfeiderer, M.M., and Galej, W.P. (2021). Emerging insights into the function and structure of the Integrator complex. *Transcription* 12, 251–265. <https://doi.org/10.1080/21541264.2022.2047583>.
9. Baillat, D., Hakimi, M.A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123, 265–276. <https://doi.org/10.1016/j.cell.2005.08.019>.
10. Shi, Y., and Manley, J.L. (2015). The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.* 29, 889–897. <https://doi.org/10.1101/gad.261974.115>.
11. Huang, K.L., Jee, D., Stein, C.B., Elrod, N.D., Henriques, T., Mascibroda, L.G., Baillat, D., Russell, W.K., Adelman, K., and Wagner, E.J. (2020). Integrator recruits protein phosphatase 2A to prevent pause release and facilitate transcription termination. *Mol. Cell* 80, 345–358.e9. <https://doi.org/10.1016/j.molcel.2020.08.016>.
12. Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., Chen, X., Li, Y., Liu, W., Li, J., et al. (2020). Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science* 370, abb5872. <https://doi.org/10.1126/science.abb5872>.
13. Vervoort, S.J., Welsh, S.A., Devlin, J.R., Barbieri, E., Knight, D.A., Offley, S., Bjelosevic, S., Costacurta, M., Todorovski, I., Kearney, C.J., et al. (2021). The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell* 184, 3143–3162.e32. <https://doi.org/10.1016/j.cell.2021.04.022>.
14. Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H., and Cramer, P. (2021). Structural basis of Integrator-mediated transcription regulation. *Science* 374, 883–887. <https://doi.org/10.1126/science.abk0154>.
15. Malovannaya, A., Li, Y., Bulynko, Y., Jung, S.Y., Wang, Y., Lanz, R.B., O’Malley, B.W., and Qin, J. (2010). Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc. Natl. Acad. Sci. USA* 107, 2431–2436. <https://doi.org/10.1073/pnas.0912599106>.
16. Yadav, L., Tamene, F., Göös, H., van Drogen, A., Katainen, R., Aebersold, R., Gstaiger, M., and Varjosalo, M. (2017). Systematic analysis of human protein phosphatase interactions and dynamics. *Cell Syst.* 4, 430–444.e5. <https://doi.org/10.1016/j.cels.2017.02.011>.
17. Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., Leitner, A., Beck, M., Hartl, F.U., Ban, N., et al. (2012). Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* 337, 1348–1352. <https://doi.org/10.1126/science.1221483>.
18. Zheng, H., Jin, Q., Wang, X., Qi, Y., Liu, W., Ren, Y., Zhao, D., Xavier Chen, F., Cheng, J., Chen, X., et al. (2023). Structural basis of INTAC-regulated

# Molecular Cell

## Article



- transcription. *Protein Cell* 14, 698–702. <https://doi.org/10.1093/procel/pwad010>.
19. Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* 15, 108–121. <https://doi.org/10.1038/nrm3742>.
20. Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399–403. <https://doi.org/10.1038/nature14906>.
21. Barbieri, E., Trizzino, M., Welsh, S.A., Owens, T.A., Calabretta, B., Carroll, M., Sarma, K., and Gardini, A. (2018). Targeted enhancer activation by a subunit of the integrator complex. *Mol. Cell* 71, 103–116.e7. <https://doi.org/10.1016/j.molcel.2018.05.031>.
22. Rubtsova, M.P., Vasilkova, D.P., Moshareva, M.A., Malyavko, A.N., Meerson, M.B., Zatsepin, T.S., Naraykina, Y.V., Beletsky, A.V., Ravin, N.V., and Dontsova, O.A. (2019). Integrator is a key component of human telomerase RNA biogenesis. *Sci. Rep.* 9, 1701. <https://doi.org/10.1038/s41598-018-38297-6>.
23. Xie, M., Zhang, W., Shu, M.D., Xu, A., Lenis, D.A., DiMaio, D., and Steitz, J.A. (2015). The host Integrator complex acts in transcription-independent maturation of herpesvirus microRNA 3' ends. *Genes Dev.* 29, 1552–1564. <https://doi.org/10.1101/gad.266973.115>.
24. Berk'yurek, A.C., Furlan, G., Lampersberger, L., Beltran, T., Weick, E.M., Nischwitz, E., Cunha Navarro, I., Braukmann, F., Akay, A., Price, J., et al. (2021). The RNA polymerase II subunit RPB-9 recruits the integrator complex to terminate *Caenorhabditis elegans* piRNA transcription. *EMBO J.* 40, e105565. <https://doi.org/10.15252/embj.2020105565>.
25. Beltran, T., Pahita, E., Ghosh, S., Lenhard, B., and Sarkies, P. (2021). Integrator is recruited to promoter-proximally paused RNA Pol II to generate *Caenorhabditis elegans* piRNA precursors. *EMBO J.* 40, e105564. <https://doi.org/10.15252/embj.2020105564>.
26. Skaar, J.R., Ferris, A.L., Wu, X., Saraf, A., Khanna, K.K., Florens, L., Washburn, M.P., Hughes, S.H., and Pagano, M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res.* 25, 288–305. <https://doi.org/10.1038/cr.2015.19>.
27. Elrod, N.D., Henriquez, T., Huang, K.L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., and Adelman, K. (2019). The integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol. Cell* 76, 738–752.e7. <https://doi.org/10.1016/j.molcel.2019.10.034>.
28. Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.S., Jiang, J.Z., Jonathan, M., Huang, K.L., Wagner, E.J., Cherry, S., and Wilusz, J.E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.* 33, 1525–1538. <https://doi.org/10.1101/gad.330167.119>.
29. Lykke-Andersen, S., Žumer, K., Molska, E.Š., Rouvière, J.O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., and Jensen, T.H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell* 81, 514–529.e6. <https://doi.org/10.1016/j.molcel.2020.12.014>.
30. Stein, C.B., Field, A.R., Mimoso, C.A., Zhao, C., Huang, K.L., Wagner, E.J., and Adelman, K. (2022). Integrator endonuclease drives promoter-proximal termination at all RNA polymerase II-transcribed loci. *Mol. Cell* 82, 4232–4245.e11. <https://doi.org/10.1016/j.molcel.2022.10.004>.
31. Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., and Steitz, J.A. (2021). Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Mol. Cell* 81, 502–513.e4. <https://doi.org/10.1016/j.molcel.2020.12.002>.
32. Beckedorff, F., Blumenthal, E., daSilva, L.F., Aoi, Y., Cingaram, P.R., Yue, J., Zhang, A., Dokaneheifard, S., Valencia, M.G., Gaidosh, G., et al. (2020). The human integrator complex facilitates transcriptional elongation by endonucleolytic cleavage of nascent transcripts. *Cell Rep.* 32, 107917. <https://doi.org/10.1016/j.celrep.2020.107917>.
33. Oegema, R., Baillat, D., Schot, R., van Unen, L.M., Brooks, A., Kia, S.K., Hoogeboom, A.J.M., Xia, Z., Li, W., Cesaroni, M., et al. (2017). Human mutations in integrator complex subunits link transcriptome integrity to brain development. *PLoS Genet.* 13, e1006809. <https://doi.org/10.1371/journal.pgen.1006809>.
34. Mascibroda, L.G., Shboul, M., Elrod, N.D., Colleaux, L., Hamamy, H., Huang, K.L., Peart, N., Singh, M.K., Lee, H., Merriman, B., et al. (2022). INTS13 variants causing a recessive developmental ciliopathy disrupt assembly of the Integrator complex. *Nat. Commun.* 13, 6054. <https://doi.org/10.1038/s41467-022-33547-8>.
35. Federico, A., Rienzo, M., Abbondanza, C., Costa, V., Ciccodicola, A., and Casamassimi, A. (2017). Pan-cancer mutational and transcriptional analysis of the integrator complex. *Int. J. Mol. Sci.* 18, 936. <https://doi.org/10.3390/ijms18050936>.
36. Yoshimi, A., Lin, K.T., Wiseman, D.H., Rahman, M.A., Pastore, A., Wang, B., Lee, S.C., Micl, J.B., Zhang, X.J., de Botton, S., et al. (2019). Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature* 574, 273–277. <https://doi.org/10.1038/s41586-019-1618-0>.
37. Tepe, B., Macke, E.L., Niceta, M., Weisz Hubshman, M., Kanca, O., Schultz-Rogers, L., Zarate, Y.A., Schaefer, G.B., Granadillo De Luque, J.L., Wegner, D.J., et al. (2023). Bi-allelic variants in INTS11 are associated with a complex neurological disorder. *Am. J. Hum. Genet.* 110, 774–789. <https://doi.org/10.1016/j.ajhg.2023.03.012>.
38. Hu, S., Peng, L., Song, A., Ji, Y.X., Cheng, J., Wang, M., and Chen, F.X. (2023). INTAC endonuclease and phosphatase modules differentially regulate transcription by RNA polymerase II. *Mol. Cell* 83, 1588–1604.e5. <https://doi.org/10.1016/j.molcel.2023.03.022>.
39. Ezzeddine, N., Chen, J., Waltenspiel, B., Burch, B., Albrecht, T., Zhuo, M., Warren, W.D., Marzluff, W.F., and Wagner, E.J. (2011). A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation. *Mol. Cell. Biol.* 31, 328–341. <https://doi.org/10.1128/MCB.00943-10>.
40. Replogle, J.M., Saunders, R.A., Pogson, A.N., Hussmann, J.A., Lenail, A., Gunja, A., Mascibroda, L., Wagner, E.J., Adelman, K., Lithwick-Yanai, G., et al. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* 185, 2559–2575.e28. <https://doi.org/10.1016/j.cell.2022.05.013>.
41. Zhang, F., Ma, T., and Yu, X. (2013). A core hSSB1-INTS complex participates in the DNA damage response. *J. Cell Sci.* 126, 4850–4855. <https://doi.org/10.1242/jcs.132514>.
42. Pfleiderer, M.M., and Galej, W.P. (2021). Structure of the catalytic core of the Integrator complex. *Mol. Cell* 81, 1246–1259.e8. <https://doi.org/10.1016/j.molcel.2021.01.005>.
43. Albrecht, T.R., Shevtsov, S.P., Wu, Y., Mascibroda, L.G., Peart, N.J., Huang, K.L., Sawyer, I.A., Tong, L., Dundr, M., and Wagner, E.J. (2018). Integrator subunit 4 is a ‘Symplekin-like’ scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res.* 46, 4241–4255. <https://doi.org/10.1093/nar/gky100>.
44. Chen, J., Waltenspiel, B., Warren, W.D., and Wagner, E.J. (2013). Functional analysis of the integrator subunit 12 identifies a microdomain that mediates activation of the Drosophila integrator complex. *J. Biol. Chem.* 288, 4867–4877. <https://doi.org/10.1074/jbc.M112.425892>.
45. Seshacharyulu, P., Pandey, P., Datta, K., and Batra, S.K. (2013). Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. *Cancer Lett.* 335, 9–18. <https://doi.org/10.1016/j.canlet.2013.02.036>.
46. Xu, Y., Xing, Y., Chen, Y., Chao, Y., Lin, Z., Fan, E., Yu, J.W., Strack, S., Jeffrey, P.D., and Shi, Y. (2006). Structure of the protein phosphatase 2A holoenzyme. *Cell* 127, 1239–1251. <https://doi.org/10.1016/j.cell.2006.11.033>.
47. Janssens, V., and Goris, J. (2001). Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. *Biochem. J.* 353, 417–439. <https://doi.org/10.1042/0264-6021:3530417>.
48. Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis

- of protein expression in yeast. *Nature* 425, 737–741. <https://doi.org/10.1038/nature02046>.
49. Ogris, E., Du, X., Nelson, K.C., Mak, E.K., Yu, X.X., Lane, W.S., and Pallas, D.C. (1999). A protein phosphatase methylesterase (PME-1) is one of several novel proteins stably associating with two inactive mutants of protein phosphatase 2A. *J. Biol. Chem.* 274, 14382–14391. <https://doi.org/10.1074/jbc.274.20.14382>.
50. Ogris, E., Mudrak, I., Mak, E., Gibson, D., and Pallas, D.C. (1999). Catalytically inactive protein phosphatase 2A can bind to polyomavirus middle tumor antigen and support complex formation with pp60(c-src). *J. Virol.* 73, 7390–7398. <https://doi.org/10.1128/JVI.73.9.7390-7398.1999>.
51. Sabath, K., Stäubli, M.L., Marti, S., Leitner, A., Moes, M., and Jonas, S. (2020). INTS10-INTS13-INTS14 form a functional module of Integrator that binds nucleic acids and the cleavage module. *Nat. Commun.* 11, 3422. <https://doi.org/10.1038/s41467-020-17232-2>.
52. Lin, M.H., Jensen, M.K., Elrod, N.D., Huang, K.L., Welle, K.A., Wagner, E.J., and Tong, L. (2022). Inositol hexakisphosphate is required for Integrator function. *Nat. Commun.* 13, 5742. <https://doi.org/10.1038/s41467-022-33506-3>.
53. Xu, C., Li, C., Chen, J., Xiong, Y., Qiao, Z., Fan, P., Li, C., Ma, S., Liu, J., Song, A., et al. (2023). R-loop-dependent promoter-proximal termination ensures genome stability. *Nature* 621, 610–619. <https://doi.org/10.1038/s41586-023-06515-5>.
54. Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* 318, 1777–1779. <https://doi.org/10.1126/science.1145989>.
55. Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and Murphy, S. (2010). The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J. Biol. Chem.* 285, 20564–20569. <https://doi.org/10.1074/jbc.M110.132530>.
56. Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., and Murphy, S. (2012). Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol. Cell* 45, 111–122. <https://doi.org/10.1016/j.molcel.2011.11.006>.
57. Hernandez, N. (1985). Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO J.* 4, 1827–1837.
58. Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D.L. (2019). Control of RNA Pol II Speed by PNTUS-PP1 and Spt5 dephosphorylation facilitates termination by a “sitting duck torpedo” mechanism. *Mol. Cell* 76, 896–908.e4. <https://doi.org/10.1016/j.molcel.2019.09.031>.
59. Kecman, T., Kuš, K., Heo, D.H., Duckett, K., Birot, A., Liberatori, S., Mohammed, S., Geis-Asteggiante, L., Robinson, C.V., and Vasiljeva, L. (2018). Elongation/termination factor exchange mediated by PP1 phosphatase orchestrates transcription termination. *Cell Rep.* 25, 259–269.e5. <https://doi.org/10.1016/j.celrep.2018.09.007>.
60. Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* 34, 132–145. <https://doi.org/10.1101/gad.332833.119>.
61. Parua, P.K., Booth, G.T., Sansó, M., Benjamin, B., Tanny, J.C., Lis, J.T., and Fisher, R.P. (2018). A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature* 558, 460–464. <https://doi.org/10.1038/s41586-018-0214-z>.
62. Filleur, S., Hirsch, J., Wille, A., Schön, M., Sell, C., Shearer, M.H., Nelius, T., and Wieland, I. (2009). INTS6/DICE1 inhibits growth of human androgen-independent prostate cancer cells by altering the cell cycle profile and Wnt signaling. *Cancer Cell Int.* 9, 28. <https://doi.org/10.1186/1475-2867-9-28>.
63. Röpke, A., Buhtz, P., Böhm, M., Seger, J., Wieland, I., Allhoff, E.P., and Wieacker, P.F. (2005). Promoter CpG hypermethylation and downregulation of DICE1 expression in prostate cancer. *Oncogene* 24, 6667–6675. <https://doi.org/10.1038/sj.onc.1208824>.
64. Kramer, M.C., Liang, D., Tatomer, D.C., Gold, B., March, Z.M., Cherry, S., and Wilusz, J.E. (2015). Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev.* 29, 2168–2182. <https://doi.org/10.1101/gad.270421.115>.
65. Chen, J., Ezzeddine, N., Waltenspiel, B., Albrecht, T.R., Warren, W.D., Marzluff, W.F., and Wagner, E.J. (2012). An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA* 18, 2148–2156. <https://doi.org/10.1261/rna.035725.112>.
66. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
67. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
68. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
69. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
70. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16, 284–287. <https://doi.org/10.1089/omi.2011.0118>.
71. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
72. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31, 2032–2034. <https://doi.org/10.1093/bioinformatics/btv098>.
73. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
74. Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>.
75. Lawrence, M., Huber, W., Pagès, H., Aboyou, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9, e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>.
76. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
77. Tatomer, D.C., Liang, D., and Wilusz, J.E. (2021). RNAi screening to identify factors that control circular RNA localization. *Methods Mol. Biol.* 2209, 321–332. [https://doi.org/10.1007/978-1-0716-0935-4\\_20](https://doi.org/10.1007/978-1-0716-0935-4_20).
78. Tatomer, D.C., Liang, D., and Wilusz, J.E. (2017). Inducible expression of eukaryotic circular RNAs from plasmids. *Methods Mol. Biol.* 1648, 143–154. [https://doi.org/10.1007/978-1-4939-7204-3\\_11](https://doi.org/10.1007/978-1-4939-7204-3_11).
79. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. <https://doi.org/10.1038/nbt.1621>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
<i>Drosophila</i> IntS6	This paper	UP2581 Bleed 3
<i>Drosophila</i> IntS8	This paper	UP2577 Bleed 3
<i>Drosophila</i> IntS11	This paper	BCM444 Bleed 2
<i>Drosophila</i> IntS12	This paper	BCM446 Bleed 2
FLAG	Sigma #F1804-1MG	AB_262044
α-tubulin	Sigma #T6074	AB_477582
PP2A C (mts)	Cell Signaling #2038	AB_2169495
RNA Pol II CTD (Clone 8WG16)	Sigma #05-952-I-100UG	AB_492629
RNA Pol II CTD phospho Ser2 (Clone 3E10)	Active Motif #61984	AB_2687450
RNA Pol II CTD phospho Ser5 (Clone 3E8)	Active Motif #61986	AB_2687451
RNA Pol II CTD phospho Ser7 (Clone 4E12)	Active Motif #61987	AB_2687452
Anti-FLAG beads	Sigma #A2220	AB_10063035
Amersham ECL Rabbit IgG, HRP-linked whole Ab	Cytiva #NA934	AB_772206
Amersham ECL Mouse IgG, HRP-linked whole Ab	Cytiva #NA931	AB_772210
Anti-rat IgG, HRP-linked Antibody	Cell Signaling #7077S	AB_10694715
<b>Chemicals, peptides, and recombinant proteins</b>		
Effectene	Qiagen	Cat #301427
Copper sulfate	Fisher BioReagents	Cat #BP346-500
Trizol	ThermoFisher Scientific	Cat #15596018
Protease inhibitors	Roche	Cat #11836170001
<b>Critical commercial assays</b>		
MEGAscript Kit	ThermoFisher Scientific	Cat #AMB13345
NorthernMax Formaldehyde Load Dye	ThermoFisher Scientific	Cat #AM8552
NorthernMax 10X Running Buffer	ThermoFisher Scientific	Cat #AM8671
NorthernMax 10X Denaturing Gel Buffer	ThermoFisher Scientific	Cat #AM8676
ULTRAhyb-Oligo	ThermoFisher Scientific	Cat #AM8663
NuPAGE 4-12% Bis-Tris gels	ThermoFisher Scientific	Cat #NP0323
PVDF membrane	Bio-Rad	Cat #1620177
NuPAGE 3-8% Tris-Acetate gels	ThermoFisher Scientific	Cat #EA03755
SuperSignal West Pico PLUS Chemiluminescent Substrate	ThermoFisher Scientific	Cat #PI34080
TURBO DNase	ThermoFisher Scientific	Cat #AM2238
TaqMan Reverse Transcription Reagents	ThermoFisher Scientific	Cat #N8080234
2x Power SYBR Green PCR Master Mix	ThermoFisher Scientific	Cat #4368708
<b>Deposited data</b>		
RNA-seq from DL1 cells, DL1 cells over-expressing IntS6, and DL1 cells over-expressing IntS12	This paper	GEO: GSE223973
RNA-seq from DL1 cells depleted of IntS4, IntS6, mts, or Pp2A-29B	This paper	GEO: GSE223974
PRO-seq from DL1 cells treated with β-gal dsRNA	Elrod et al. <sup>27</sup>	GEO: GSE114467
RNA-seq from DL1 cells depleted of IntS11 and rescued with WT IntS11 or E203Q mutant IntS11	Elrod et al. <sup>27</sup>	GEO: GSE114467
RNA-seq from DL1 cells depleted of IntS8 and rescued with WT IntS8 or WFEF/A mutant IntS8	Huang et al. <sup>11</sup>	GEO: GSE150844

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IntS1 and IntS12 ChIP-seq from DL1 cells	Elrod et al. <sup>27</sup>	GEO: GSE114467
Raw image files	Mendeley	<a href="https://data.mendeley.com/datasets/r46b2sgkpb/1">https://data.mendeley.com/datasets/r46b2sgkpb/1</a>
<b>Experimental models: Cell lines</b>		
DL1	Sara Cherry, Univ of Pennsylvania	CVCL_Z231
DL1 pMtnA IntS6	This paper	N/A
DL1 pMtnA IntS12	This paper	N/A
DL1 pMtnA IntS7	This paper	N/A
DL1 pMtnA IntS13	This paper	N/A
DL1 pMtnA IntS14	This paper	N/A
<b>Oligonucleotides</b>		
Table S5	This paper	N/A
<b>Recombinant DNA</b>		
Hy_pMT eGFP SV40	Kramer et al. <sup>64</sup>	Addgene #69911
Hy_pHml eGFP SV40	Tatomer et al. <sup>28</sup>	Addgene #132645
Hy_pCG8620 eGFP SV40	Tatomer et al. <sup>28</sup>	Addgene #132646
Hy_pUbi-p63e eGFP SV40	Tatomer et al. <sup>28</sup>	Addgene #132650
Hy_pRnrS eGFP SV40	This paper	Addgene #195062
Hy_pAna eGFP SV40	This paper	Addgene #195063
pUC U4:39B eGFP	Chen et al. <sup>44</sup>	N/A
Hy_U5:34A eGFP SV40	This paper	Addgene #195064
pUb 3xFLAG MCS	Chen et al. <sup>65</sup>	N/A
pUb FLAG-Mts	This paper	Addgene #195065
pUb FLAG-Mts H59Q	This paper	Addgene #208402
pUb FLAG-Mts D85N	This paper	Addgene #208403
pUb FLAG-Mts H241A	This paper	Addgene #208404
pUb FLAG-Pp2A-29B	This paper	Addgene #195066
pUb FLAG-Cka	This paper	Addgene #195067
pUb FLAG-tws	This paper	Addgene #195068
pUb FLAG-wdb	This paper	Addgene #195069
pUb FLAG-wrd	This paper	Addgene #195070
pUb FLAG-IntS6 AA 1-400	This paper	Addgene #195071
pUb FLAG-IntS6 AA 1-600	This paper	Addgene #195072
pUb FLAG-IntS6 AA 101-1284	This paper	Addgene #195073
pUb FLAG-IntS6 AA 1197-1284	This paper	Addgene #195074
pUb FLAG-IntS6 AA 101-1200	This paper	Addgene #195075
pUb FLAG-Human IntS6	This paper	Addgene #198408
pUb FLAG-Human IntS6L	This paper	Addgene #198409
pUb FLAG-Zebrafish IntS6	This paper	Addgene #198410
pUb FLAG-Zebrafish IntS6L	This paper	Addgene #198411
pUb FLAG-IntS1	Tatomer et al. <sup>28</sup>	N/A
pUb FLAG-IntS2	Eric Wagner, Univ of Rochester	N/A
pUb FLAG-IntS3	Eric Wagner, Univ of Rochester	N/A
pUb FLAG-IntS4	Eric Wagner, Univ of Rochester	N/A
pUb FLAG-IntS5	Tatomer et al. <sup>28</sup>	N/A
pUb FLAG-IntS6	Eric Wagner, Univ of Rochester	N/A
pUb FLAG-IntS7	Eric Wagner, Univ of Rochester	N/A
pUb FLAG-IntS8	Huang et al. <sup>11</sup>	N/A
pUb FLAG-IntS9	Eric Wagner, Univ of Rochester	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pUb FLAG-IntS10	Eric Wagner, Univ of Rochester	N/A
pUb FLAG-IntS11	Tatomer et al. <sup>28</sup>	N/A
pUb FLAG-IntS12	Chen et al. <sup>44</sup>	N/A
pUb FLAG-IntS13	Mascibroda et al. <sup>34</sup>	N/A
pUb FLAG-IntS14	Eric Wagner, Univ of Rochester	N/A
pMT FLAG MCS puro	Elrod et al. <sup>27</sup>	N/A
pMtnA FLAG-IntS6 puro	This paper	Addgene #195076
pMtnA FLAG-IntS7 puro	This paper	Addgene #208405
pMtnA FLAG-IntS12 puro	This paper	Addgene #195077
pMtnA FLAG-IntS13 puro	This paper	Addgene #208406
pMtnA FLAG-IntS14 puro	This paper	Addgene #208407
pRSFDuet-1	Novagen	Cat #71341
pRSFDuet-1 IntS6 AA 1035-1284	This paper	Addgene #196904
pRSFDuet-1 IntS8 AA 1-308	This paper	Addgene #196905
pRSFDuet-1 IntS11 AA 300-597	This paper	Addgene #199329
pGEX-6P-1	Amersham	Cat #27-4597-01
pGEX-6P-1 IntS12	This paper	Addgene #210521

**Software and algorithms**

Trimmomatic	Bolger et al. <sup>66</sup>	N/A
ImageQuant	Cytiva Life Sciences	N/A
HISAT2	Kim et al. <sup>67</sup>	N/A
featureCounts	Liao et al. <sup>68</sup>	N/A
DESeq2	Love et al. <sup>69</sup>	N/A
R 3.6.3	<a href="https://www.r-project.org">https://www.r-project.org</a>	N/A
clusterProfiler	Yu et al. <sup>70</sup>	N/A
bowtie2	Langmead and Salzberg <sup>71</sup>	N/A
Sambamba	Tarasov et al. <sup>72</sup>	N/A
MACS2	Zhang et al. <sup>73</sup>	N/A
ChIPseeker	Yu et al. <sup>74</sup>	N/A
GenomicFeatures	Lawrence et al. <sup>75</sup>	N/A
Prism v10.0.0	GraphPad	N/A
Bowtie	Langmead et al. <sup>76</sup>	N/A

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jeremy E. Wilusz ([jeremy.wilusz@bcm.edu](mailto:jeremy.wilusz@bcm.edu)).

**Materials availability**

All unique materials generated in this study are available upon request. Plasmids generated in this study have been deposited to Addgene.

**Data and code availability**

- RNA-seq data have been deposited at GEO and original Northern and western blot images have been deposited at Mendeley. All data are publicly available as of the date of publication. Accession numbers and DOIs are listed in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Cell lines

*Drosophila* DL1 cells were cultured at 25°C in Schneider's *Drosophila* medium (Thermo Fisher Scientific 21720024), supplemented with 10% (v/v) fetal bovine serum (HyClone SH30396.03), 1% (v/v) penicillin-streptomycin (Thermo Fisher Scientific 15140122), and 1% (v/v) L-glutamine (Thermo Fisher Scientific 35050061).

### METHOD DETAILS

#### Expression plasmids

*Drosophila* reporter plasmids expressing eGFP under the control of the inducible Metallothionein A promoter (Hy\_pMT eGFP SV40; Addgene #69911), Hml promoter (Hy\_pHml eGFP SV40; Addgene #132645), CG8620 promoter (Hy\_pCG8620 eGFP SV40; Addgene #132646), or Ubi-p63e promoter (Hy\_pUbi-p63e eGFP SV40; Addgene #132650) were described previously.<sup>28,64</sup> Reporter plasmids expressing eGFP under the control of the RnrS promoter (Hy\_pRnrS eGFP SV40, Addgene #195062) or Ana promoter (Hy\_pAna eGFP SV40, Addgene #195063) were generated from Hy\_pPepck1 eGFP SV40 (Addgene #132644) as detailed in [Methods S1](#). Plasmids expressing individual Integrator or PP2A subunits under the control of the Ubi-p63e promoter were generated using the previously described pUb 3xFLAG MCS plasmid<sup>65</sup> as detailed in [Methods S1](#). Full-length IntS1-14 expression plasmids were obtained from Eric J. Wagner, University of Rochester. Plasmids expressing IntS6 (pMtnA FLAG-IntS6 puro, Addgene #195076), IntS7 (pMtnA FLAG-IntS7 puro, Addgene #208405), IntS12 (pMtnA FLAG-IntS12 puro, Addgene #195077), IntS13 (pMtnA FLAG-IntS13 puro, Addgene #208406), or IntS14 (pMtnA FLAG-IntS14 puro, Addgene #208407) under the control of the Metallothionein A promoter were generated using the previously described pMT FLAG MCS puro plasmid.<sup>27</sup> The reporter plasmid expressing eGFP downstream of the U4:39B snRNA gene was previously described.<sup>44</sup> An analogous reporter expressing eGFP downstream of U5:34A snRNA (Hy\_U5:34A eGFP SV40, Addgene #195064) was generated from Hy\_pPepck1 eGFP SV40. Details of all cloning, including full plasmid sequences, are described in [Methods S1](#).

#### Double-stranded RNAs

Double-stranded RNAs (dsRNAs) from the DRSC (*Drosophila* RNAi Screening Center) were generated by *in vitro* transcription (MEGAscript kit, Thermo Fisher Scientific AMB13345) of PCR templates containing the T7 promoter sequence on both ends as previously described in detail.<sup>77</sup> Primer sequences are provided in [Table S5](#).

#### Generation of stable cell lines

To generate DL1 cells stably maintaining the inducible IntS6, IntS7, IntS12, IntS13, or IntS14 transgenes, 2 x 10<sup>6</sup> cells were first plated in complete media in 6-well dishes. After 24 h, 2 µg of pMtnA FLAG-IntS6 puro (Addgene #195076), pMtnA FLAG-IntS7 puro (Addgene #208405), pMtnA FLAG-IntS12 puro (Addgene #195077), pMtnA FLAG-IntS13 puro (Addgene #208406), or pMtnA FLAG-IntS14 puro (Addgene #208407) plasmid was transfected using Effectene (Qiagen 301427; 16 µL Enhancer and 30 µL Effectene Reagent). On the following day, 5 µg/mL puromycin was added to the media to select and maintain the cell population.

#### Transfections, RNAi, and RNA isolation

To determine the effect of Integrator subunit over-expression on the output of the eGFP reporters, DL1 cells were seeded in 12-well plates (5 x 10<sup>5</sup> cells per well) in complete Schneider's *Drosophila* medium and cultured overnight. On the following day, 500 ng of plasmid DNA was transfected into each well using Effectene (4 µL of enhancer and 5 µL of Effectene reagent; Qiagen 301427). Unless otherwise noted, 400 ng of eGFP plasmid was transfected along with 100 ng of the Integrator/PP2A sub-unit expression plasmid. Total RNA was isolated ~48 h later using TRIzol (Thermo Fisher Scientific 15596018) according to the manufacturer's instructions.

To determine the effect of Integrator subunit down-regulation on the output of the eGFP reporters, 1 x 10<sup>6</sup> DL1 cells were first bathed in 2 µg of dsRNAs in 12-well plates in 0.5 mL Schneider's *Drosophila* medium without FBS, penicillin/streptomycin, or L-glutamine. After incubating cells at 25°C for 45 min, 1 mL of complete *Drosophila* medium was added and cells were grown at 25°C. After 24 h, Effectene (4 µL of enhancer and 5 µL of Effectene reagent; Qiagen 301427) was used to transfect 500 ng of the indicated eGFP plasmid. Total RNA was isolated ~48 h later using TRIzol (Thermo Fisher Scientific 15596018) according to the manufacturer's instructions. When examining reporters driven by the Metallothionein A promoter, a final concentration of 500 µM copper sulfate (Fisher BioReagents BP346-500) was added to cells for the last 14 h prior to RNA isolation.

#### Northern blotting

Northern blots using 1.2% denaturing formaldehyde agarose gels, NorthernMax reagents (Thermo Fisher Scientific), and oligonucleotide probes were performed as previously described in detail.<sup>78</sup> Oligonucleotide probe sequences are provided in [Table S5](#). Blots were hybridized overnight at 42°C with oligonucleotide probes in ULTRAhyb-Oligo (Thermo Fisher Scientific AM8663), washed two times with 2x SSC, 0.5% SDS, and viewed with the Amersham Typhoon scanner (Cytiva) followed by quantification using ImageQuant (Cytiva).

### Antibody production

The C-terminal region of *Drosophila* IntS6 (amino acids 1035–1284) and the N-terminal region of *Drosophila* IntS8 (amino acids 1–308) were individually cloned into pRSFDuet-1 using the BamHI and HindIII restriction enzyme sites to generate pRSFDuet-1 IntS6 AA 1035–1284 (Addgene #196904) and pRSFDuet-1 IntS8 AA 1–308 (Addgene #196905), respectively. The C-terminal region of *Drosophila* IntS11 (amino acids 300–597) was cloned into pRSFDuet-1 using the Sall and HindIII restriction enzyme sites to generate pRSFDuet-1 IntS11 AA 300–597 (Addgene #199329). Full length *Drosophila* IntS12 was cloned into pGEX-6P-1 using EcoRI and NotI to generate pGEX-6P-1 IntS12 (Addgene #210521). Details of cloning, including full plasmid sequences, are described in [Methods S1](#). Plasmids were transformed into BL21 Star (DE3) *E. coli* and grown in terrific broth media supplemented with 50 µg/mL kanamycin for IntS6, IntS8, and IntS11, and with 100 µg/mL ampicillin for IntS12. Expression of the His-tagged proteins was induced at OD<sub>600</sub> ~0.8 by addition of 0.3 mM IPTG. IntS6, IntS8, IntS11, and IntS12 cultures were incubated at 16°C for 20 h, 25°C for 7 h, 16°C for 20 h, and 16°C for 16 h, respectively, before cells were harvested. All purification steps were carried out at 4°C unless otherwise noted.

For purification of His-tagged IntS6, the cell pellet was lysed by sonication in lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 0.5 mM DTT, 25 mM imidazole, 1 mM PMSF, 100 µM leupeptin, 10 µM pepstatin A, and 1 mM benzamidine). Cell debris was removed by centrifugation at 20,000 × g for 15 min. The supernatant was filtered and loaded onto a Ni-column (Cytiva 17524801). The column was then washed with wash buffer (30 mM Tris pH 8.0, 300 mM NaCl, 50 mM imidazole) and eluted by gradient of imidazole to 400 mM. Fractions containing His-tagged IntS6 were pooled, concentrated, and loaded onto a Superdex 200 column in PBS. Fractions containing His-tagged IntS6 were pooled, concentrated, flash frozen, and stored at -80°C.

His-tagged IntS8 and IntS11 were purified under denaturing conditions. Cells were resuspended and incubated in lysis buffer (30 mM Tris pH 7.5, 300 mM NaCl, 1 mM DTT, 1 mM EDTA, 100 µg/mL lysozyme, 1 mM PMSF, 100 µM leupeptin, 10 µM pepstatin A, and 1 mM benzamidine) on ice for 30 min, and then sonicated for a total of 5 min. Sample was pelleted by centrifugation at 20,000 × g for 15 min and then resuspended in denaturing buffer (30 mM Tris pH 7.5, 300 mM NaCl, and 8 M urea) at 4°C for 30 min while stirring. Cell debris was removed by centrifugation at 20,000 × g for 15 min. The solubilized IntS8 protein was diluted with buffer to lower the urea concentration to 1 M, loaded onto a Ni-column (Cytiva 29051021), and eluted by adding elution buffer (30 mM Tris pH 7.5, 300 mM NaCl, and 500 mM imidazole). Fractions containing His-tagged IntS8 were pooled, dialyzed against PBS, concentrated, flash frozen, and stored at -80°C. The solubilized IntS11 protein in denaturing buffer was flash frozen and stored at -80°C.

For purification of IntS12, the cell pellet was lysed by sonication in lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 5 mM DTT, 0.1% Triton X-100, 1 mM PMSF, 100 µM leupeptin, 10 µM pepstatin A, and 1 mM benzamidine). Cell debris was removed by centrifugation at 12,000 × g for 1 h. The supernatant was incubated with pre-equilibrated Glutathione Sepharose 4B resin for 40 min while rotating. The protein bound resin was washed with lysis buffer, then with buffer A (50 mM Tris pH 7.5, 1 M NaCl, 3 mM DTT, 1 mM PMSF, 100 µM leupeptin, 10 µM pepstatin A, and 1 mM benzamidine), and then with buffer B (50 mM Tris pH 7.5, 100 mM NaCl, 3 mM DTT, and 5% glycerol). IntS12 was released from the resin by incubating the resin with PreScission protease for 17 h at 4°C. IntS12 was eluted with buffer B and fractions containing IntS12 were passed through a heparin column to remove contamination. Flow through was collected, dialyzed in PBS, concentrated, flash frozen, and stored at -80°C.

Purified His-tagged IntS6, IntS8, and IntS12 proteins in PBS were shipped to a commercial vendor (Cocalico Biologicals) and used to inoculate rabbits. His-tagged IntS11 was purified from SDS-PAGE gels and the cut bands were likewise used to inoculate rabbits. The reactivity and specificity of antisera was confirmed with Western blots using whole cell extracts from DL1 cells treated for 3 d with dsRNAs targeting either IntS6, IntS8, IntS11, or IntS12.

### Western blotting

In experiments where Pol II levels were analyzed using western blots, cells were resuspended in RIPA buffer (150 mM NaCl, 1% Triton X-100, 50 mM Tris pH 7.5, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40) supplemented with protease/phosphatase inhibitor (Cell Signaling 5872S) and benzonase nuclease (Sigma E1014-25KU) and then incubated on ice for 1 h. For western blot experiments where Pol II levels were not measured, cells were resuspended in RIPA buffer supplemented with protease inhibitors (Roche 11836170001) and incubated on ice for 20 min. Cell lysates were cleared at 21,000 × g for 20 min at 4°C and protein concentrations were measured using a standard Bradford assay (Bio-Rad 5000006). Lysates containing 20 µg protein were then resolved on NuPAGE 3–8% Tris-Acetate gels (Thermo Fisher Scientific EA03755) when the Rbp1 CTD antibody was used to resolve hyper- and hypophosphorylated isoforms. For Westerns using all other antibodies, lysates containing 20 µg protein were resolved on NuPAGE 4–12% Bis-Tris gels (Thermo Fisher Scientific NP0323). All gels were transferred to PVDF membranes (Bio-Rad 1620177). Membranes were blocked with 10% nonfat milk in TBST for 1 h before incubation in primary antibody (diluted in 1x TBST) overnight at 4°C. Membranes were then washed with 1x TBST (4 × 10 min) followed by incubation in secondary antibody (diluted in 1x TBST) at room temperature for 1 h. Antibody incubation conditions are summarized in [Table S6](#). Membranes were washed with 1x TBST (4 × 10 min) and processed using SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific PI34080). The intensity of protein bands were quantified using ImageQuantTL (Cytiva).

### RT-qPCR

5 µg of total RNA (quantified by Nanodrop) was digested with TURBO DNase (Thermo Fisher Scientific AM2238) in a 20 µL reaction following the manufacturer's protocol. Samples were then incubated at 75°C for 10 min in the presence of 15 mM EDTA. 1 µg of the

digested RNA was reverse transcribed to cDNA in a 20  $\mu$ L reaction using TaqMan Reverse Transcription Reagents (Thermo Fisher Scientific N8080234) with random hexamers following the manufacturer's protocol except that 4  $\mu$ L of 25 mM MgCl<sub>2</sub> instead of 1.4  $\mu$ L was used. RT-qPCR reactions were performed in 15  $\mu$ L reactions that contained 1.5  $\mu$ L of cDNA (diluted up to 10-fold in H<sub>2</sub>O), 7.5  $\mu$ L 2x Power SYBR Green PCR Master Mix (Thermo Fisher Scientific 4368708), and 6  $\mu$ L 1.5  $\mu$ M gene-specific primer pairs. Primer sequences are provided in [Table S5](#).

Using the QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific A28566) and clear plates (Thermo Fisher Scientific 4346907), the following cycling conditions were used: 95°C for 10 min, 40 amplification cycles of 95°C for 15 s followed by 60°C for 1 min, and a final melting cycle of 95°C for 10 s, 65°C for 1 min, and 97°C for 1 s. Subsequently, a melt curve was performed to verify that amplified products were a single discrete species. Threshold cycle (CT) values were calculated by the QuantStudio 3 system and relative transcript levels (compared to RpL32) were calculated using the 2<sup>-ΔΔCT</sup> method. RT-qPCR reactions were performed using at least three independent biological replicates, with each replicate having two technical replicates.

### Immunoprecipitation

Parental DL1 cells and DL1 cells stably maintaining inducible FLAG-tagged IntS6 or IntS12 transgenes (DL1 pMtnA IntS6 and DL1 pMtnA IntS12, respectively) were grown in T75 flasks (4 flasks/each) for 3 d and 500  $\mu$ M copper sulfate (Fisher BioReagents BP346-500) was added for the last 24 h. Cells were harvested by centrifugation at 340  $\times g$  for 5 min and then washed with ice-cold PBS. Cells were resuspended in hypotonic buffer (50 mM Tris pH 7.5, 10 mM KCl, 1 mM DTT, and protease inhibitor mix containing 1 mM PMSF, 100  $\mu$ M leupeptin, 10  $\mu$ M pepstatin A, and 1 mM benzamidine) and incubated for 20 min on ice. Cell suspension was then homogenized using a glass Dounce homogenizer. Nuclei were pelleted at 1000  $\times g$  for 10 min at 4°C and washed once with hypotonic buffer. Nuclei were incubated in lysis buffer (40 mM Tris pH 7.5, 300 mM NaCl, 1 mM DTT, 10% glycerol, 0.75% Triton X-100 and protease inhibitors) for 20 min on ice. Insoluble proteins and cell debris were removed by centrifugation at 21,000  $\times g$  for 30 min and the supernatant was passed through a 0.45  $\mu$ m filter (Thermo Fisher Scientific F2513-14). The supernatant was diluted with dilution buffer (20 mM Tris pH 7.5, 10 mM NaCl, 1 mM DTT, 10% glycerol, and protease inhibitors) such that the final NaCl and Triton X-100 concentrations became 200 mM and 0.5%, respectively. The diluted supernatant was then incubated with pre-equilibrated anti-FLAG beads (Sigma A2220) while rotating for 2 h at 4°C. The beads were washed five times with wash buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT, 10% glycerol, 0.25% Triton X-100, and protease inhibitors). Bound proteins were eluted with 0.1 M glycine pH 3.5, TCA precipitated, and resuspended in 1x loading dye containing 5 mM DTT.

### RNA-seq library generation

To determine the effect of Integrator subunit over-expression on the endogenous transcriptome, DL1, DL1 pMtnA IntS6, and DL1 pMtnA IntS12 cells were seeded in 6-well plates (2  $\times$  10<sup>6</sup> cells per well) in 2 mL complete Schneider's *Drosophila* medium (with FBS, penicillin/streptomycin, and L-glutamine) and grown for 3 d. As indicated, a final concentration of 500  $\mu$ M copper sulfate (Fisher BioReagents BP346-500) was added to cells for the last 24 h prior to RNA isolation using TRIzol (Thermo Fisher Scientific 15596018). Total RNA was isolated from three biological replicates, treated with DNase I, depleted of ribosomal RNAs, and strand-specific RNA-seq libraries were generated by Genewiz/Azenta Life Sciences. Libraries were sequenced using Illumina HiSeq, 2  $\times$  150 bp configuration.

To determine the effect of Integrator/PP2A subunit down-regulation on the endogenous transcriptome, 5  $\times$  10<sup>5</sup> DL1 cells were bathed in 2  $\mu$ g of dsRNAs in 12-well plates in 0.5 mL Schneider's *Drosophila* medium without FBS, penicillin/streptomycin, or L-glutamine. After incubating cells at 25°C for 45 min, 1 mL of complete *Drosophila* medium was added and cells were grown at 25°C for 3 d. Total RNA was isolated using TRIzol (Thermo Fisher Scientific 15596018) according to the manufacturer's instructions. RNAs were treated with DNase I, depleted of ribosomal RNAs, and strand-specific RNA-seq libraries were generated by Genewiz/Azenta Life Sciences. Libraries were sequenced using Illumina HiSeq, 2  $\times$  150 bp configuration.

The GEO accession numbers for the RNA-seq datasets generated in this study are GSE223973 and GSE223974.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### RNA-seq analysis

For paired-end RNA-seq samples generated in this study, Trimmomatic<sup>66</sup> (version 0.39; parameters: PE -threads 3 -phred33 TruSeq3PE-2.fa:2:30:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30) was used for quality control of RNA-seq datasets, including removal of adaptor sequences and low-quality bases at both ends of reads. Reads were next aligned to the *D. melanogaster* dm6/BDGP6.22 reference genome using HISAT2<sup>67</sup> (version 2.1.0, parameters: -no-softclip -score-min L,-16,0 -mp 7,7 -rfg 0,7 -rdg 0,7 -dta -k 1 -max-seeds 20). Mapping summaries of RNA-seq datasets are described in [Table S1](#). Fragment counts were calculated per gene using featureCounts<sup>68</sup> (version 2.0.1, parameters: -s 2 -p -fraction -O -T 16 -t exon -g gene\_id) in a strand-specific manner based on the *D. melanogaster* Ensembl gene annotation (BDGP6.22.96). Differentially expressed genes (DEGs) were then identified using DESeq2<sup>69</sup> (version: 1.26.0) under R 3.6.3 with a threshold of an adjusted *P* value < 0.001 and |log<sub>2</sub>(fold change, FC)| > 0.585. Pearson correlation coefficient (PCC) of FC of 96 up-regulated genes between IntS6 over-expression, IntS4 depletion, and IntS6 depletion were calculated under R 3.6.3. GO enrichment analysis (biological process) for the 107 genes up-regulated with IntS6 over-expression was performed and visualized with clusterProfiler<sup>70</sup> (version 3.14.3, parameters: pvalueCutoff = 0.05, qvalueCutoff = 0.2).

Fragment counts mapped to 3 kb downstream of snRNA gene bodies were calculated using featureCounts<sup>68</sup> (version 2.0.1, parameters: -s 2 -p -fraction -O -T 16 -t exon -g gene\_id) in a strand-specific manner based on the *D. melanogaster* Ensembl gene annotation (BDGP6.22.96). DESeq2<sup>69</sup> (version: 1.26.0) was then used to compare differences for each snRNA between treatments under R 3.6.3. Next, to compare the overall endogenous snRNA readthrough levels between treatments, readthrough transcription from a given snRNA was calculated by dividing RNA-seq fragment counts aligned to 3 kb downstream of the snRNA gene body with the fragment counts aligned to the mature snRNA. Statistical difference between treatments was assessed using Wilcoxon signed-rank test under R 3.6.3.

For single-end RNA-seq samples downloaded from GEO (GSE114467),<sup>27</sup> Trimmomatic<sup>66</sup> (version 0.39; parameters: SE -threads 5 -phred33 TruSeq3PE-2.fa:2:30:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30) was used for quality control, including removal of adaptor sequences and low-quality bases. Reads were then aligned to *D. melanogaster* dm6/BDGP6.22 reference genome using HISAT2<sup>67</sup> (version 2.1.0, parameters: -k 1 --max-seeds 2). The mapping summaries of RNA-seq datasets are described in Table S1. Fragment counts were calculated per gene using featureCounts<sup>68</sup> (version 2.0.1, parameters: -s 1 -p -fraction -O -T 16 -t exon -g gene\_id) and then normalized to FPKM (fragments per kilobase of gene per million fragments mapped).<sup>79</sup>

Lengths of all genes were extracted from the *D. melanogaster* Ensembl gene annotation (BDGP6.22.96).

### ChIP-seq analysis

Published ChIP-seq data for IntS1 and IntS12 in DL1 cells (3 replicates/each) were downloaded from GEO (GSE114467).<sup>27</sup> Raw sequences were filtered and trimmed using Trimmomatic<sup>66</sup> (version 0.39; parameters: PE -threads 3 -phred33 TruSeq3PE-2.fa:2:30:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30). Reads were then aligned to the *D. melanogaster* dm6/BDGP6.22 reference genome using bowtie2<sup>71</sup> (version 2.3.5) with default parameters. Multi-mapped reads were removed using sambamba<sup>72</sup> (version 0.8.0, parameter: view -F "[XS] == null and not unmapped"). Mapping summaries of ChIP-seq datasets are described in Table S1. IntS1 and IntS12 ChIP-seq peaks were next called using MACS2<sup>73</sup> (version 2.2.7.1, parameters: -g 142573017 -q 0.001). ChIP-seq peaks were annotated by ChIPseeker<sup>74</sup> (version 1.22.1), with the TxDb object of *D. melanogaster* Ensembl gene annotation (BDGP6.22.96) generated using the makeTxDbFromGFF function in GenomicFeatures package (version 1.38.0).<sup>75</sup> Genes closest to peaks and with peaks located within  $\pm 1$  kb of gene bodies in all 3 replicates were defined as IntS1 or IntS12 binding genes. Finally, IntS1 binding genes and IntS12 binding genes were merged to define a set of 3,932 Integrator bound genes.

### PRO-seq analysis

Published paired-end PRO-seq data for DL1 cells (3 replicates of control ( $\beta$ -gal dsRNA) samples, spiked with mouse embryonic stem cells) were downloaded from GEO (GSE114467).<sup>27</sup> Cutadapter (version v1.18, parameters: -a TGGAATTCTCGGGTGCCAGG -A NNNNNNNNGATCGTCGGACTGTAGAACTCTGAAC -q 10 -m 20) was used to remove adapter sequences, low quality bases at 3' ends, and reads that are <20 nt after trimming. Remaining fragments were first aligned to the mouse mm10 reference genome. Unmapped fragments were then aligned to the *D. melanogaster* dm6/BDGP6.22 reference genome. Bowtie<sup>76</sup> was used in above two alignment steps with same parameters (version 1.12, parameters: -m1 -v2 -un). Mapping summaries of PRO-seq datasets are described in Table S1.

Fragments mapped to the promoter and gene body for each gene were calculated using featureCounts<sup>59</sup> (version 2.0.1, parameters: -s 2 -F SAF -Q 20 -p -fraction -O -a). Transcription start sites (TSSs) were extracted from *D. melanogaster* Ensembl gene annotation (BDGP6.22.96). TSSs  $\pm$  250 nt were selected to represent the promoter region and +250 to +1250 nt from TSSs were selected to represent gene body region for each gene. Fragment counts were then normalized to FPKM (fragments per kilobase of gene per million fragments mapped)<sup>79</sup> and FPKM values for 3 replicates were averaged to denote the PRO-seq signal.

### Statistical tests

For Northern blots, Western blots, and RT-qPCRs, statistical significance for comparison of means was assessed by one-way ANOVA using GraphPad Prism. Statistical analyses and error bars are explained in the corresponding figure legends, when applicable.

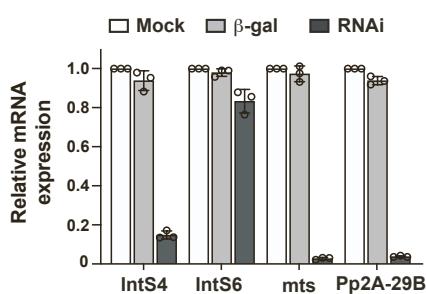
**Supplemental information**

**IntS6 and the Integrator phosphatase module  
tune the efficiency of select premature  
transcription termination events**

Rina Fujiwara, Si-Nan Zhai, Dongming Liang, Aayushi P. Shah, Matthew Tracey, Xu-Kai Ma, Christopher J. Fields, María Saraí Mendoza-Figueroa, Michele C. Meline, Deirdre C. Tatomer, Li Yang, and Jeremy E. Wilusz

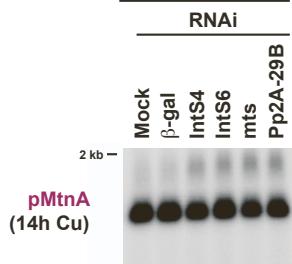
# Supplemental Figure S1

**A**

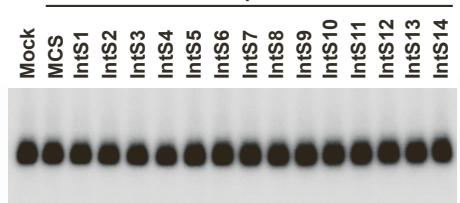


**C**

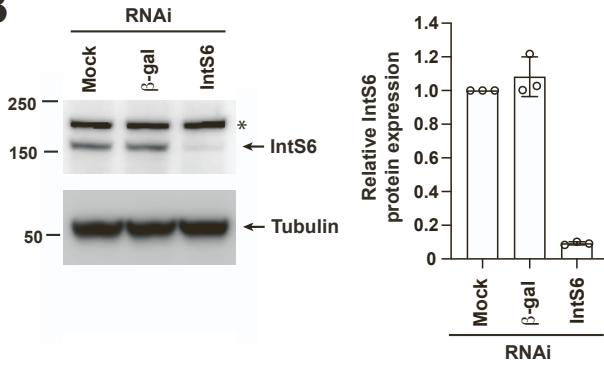
Promoter / snRNA  
eGFP Reporter  
RNAi



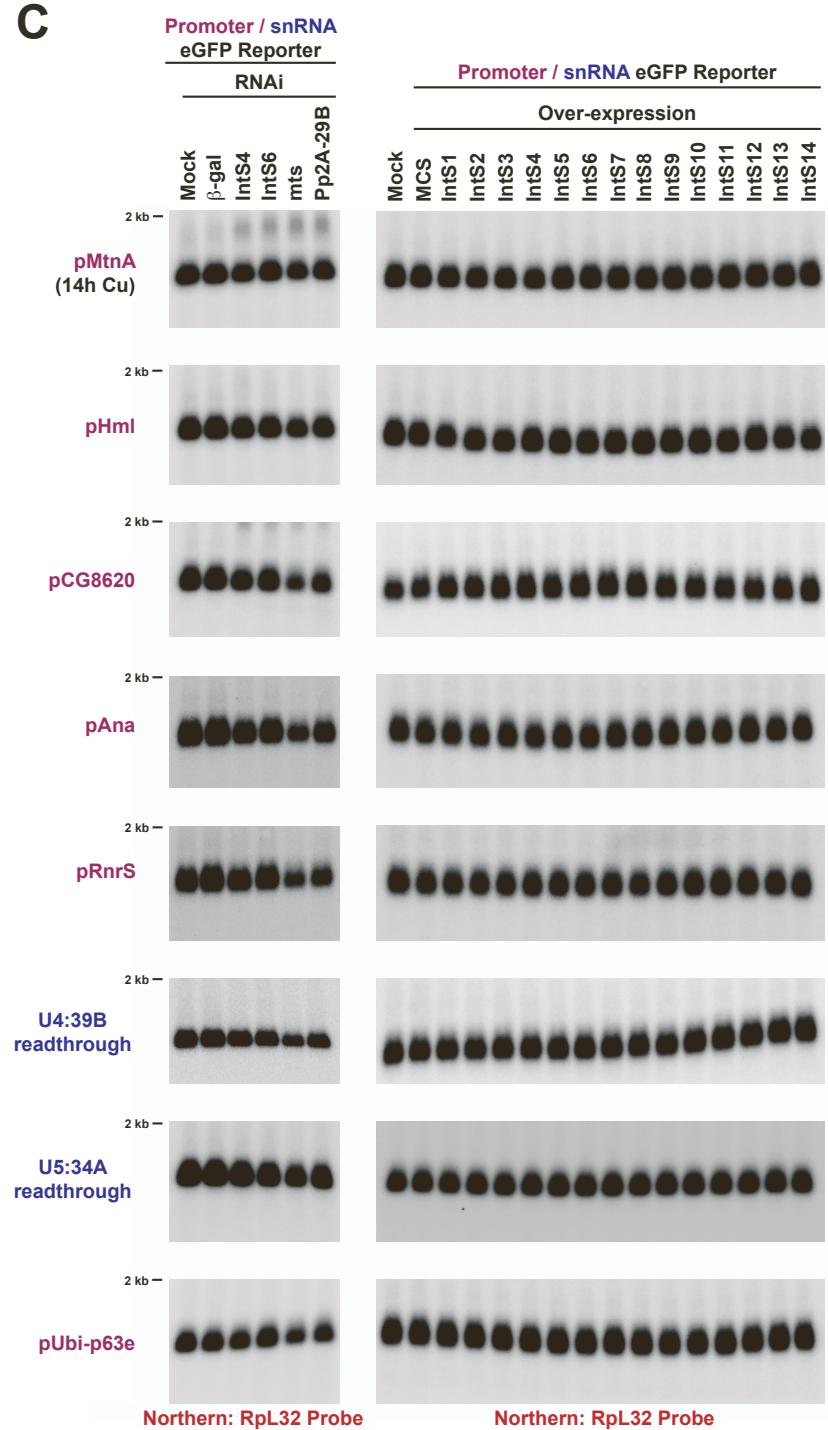
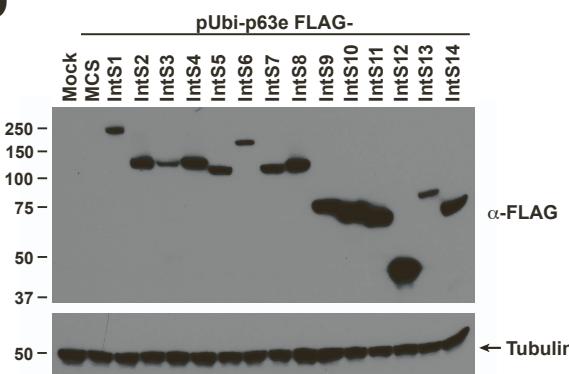
Promoter / snRNA  
eGFP Reporter  
Over-expression



**B**



**D**



**Figure S1. Validation of Integrator subunit depletion and over-expression effects in *Drosophila* DL1 cells. Related to Figure 1.**

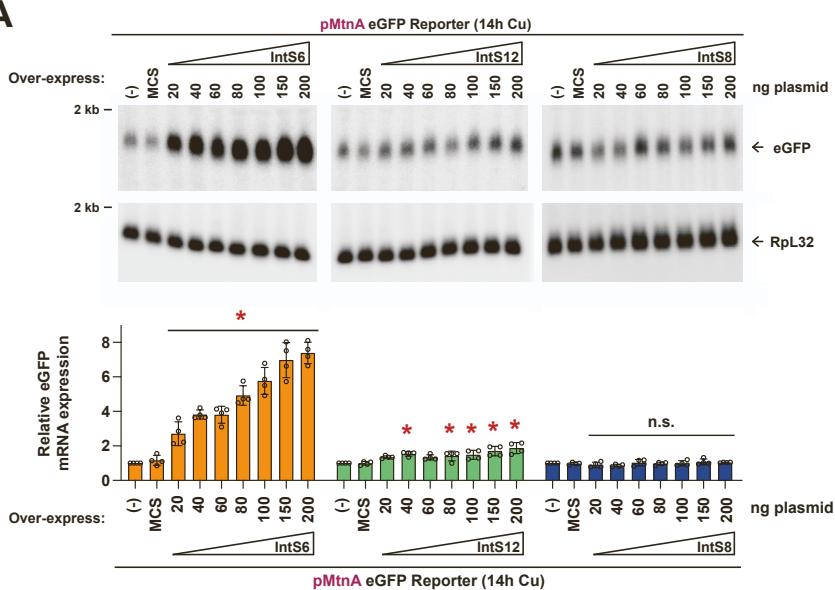
**(A, B)** To quantify the efficiency of dsRNA-mediated depletion of target gene expression, *Drosophila* DL1 cells were treated with dsRNAs for 3 d and then total RNA or protein was isolated. **(A)** RT-qPCR was used to quantify expression of IntS4, IntS6, mts, or Pp2A-29B mRNAs and data were normalized to expression of RpL32 mRNA. Data are shown as mean ± SD,  $N = 3$ . **(B)** Western blotting was performed using an antibody that recognizes the C-terminal region of *Drosophila* IntS6 (amino acids 1035-1284). \* denotes non-specific band.  $\alpha$ -tubulin was used as a loading control. Data are shown as mean ± SD,  $N = 3$ .

**(C)** Equal loading of RNA was verified for the Northern blots shown in **Figure 1B** by re-probing the membranes for RpL32 mRNA.

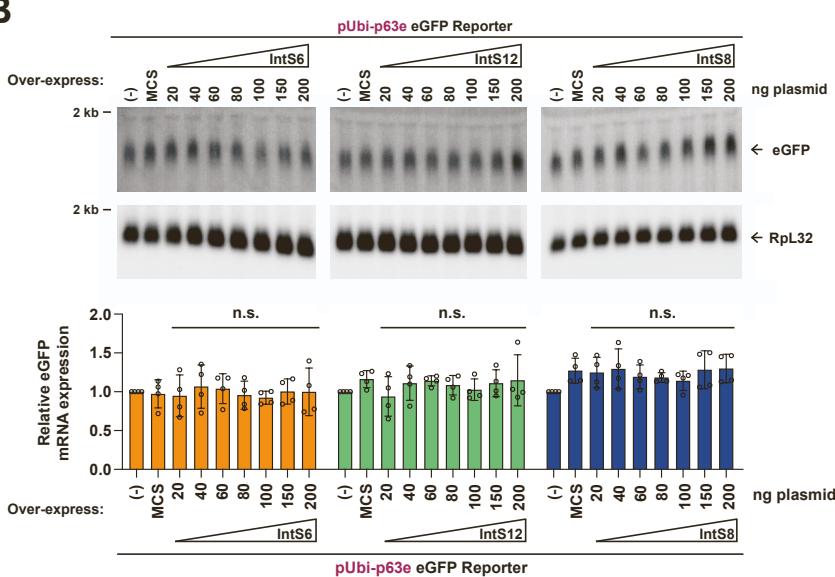
**(D)** DL1 cells were transfected with 500 ng of the indicated FLAG-tagged expression plasmids and total protein was harvested after 48 h. A plasmid containing a multi-cloning site (MCS) was used as a control. Western blot analysis using an antibody that recognizes FLAG was used to confirm expression of individual Integrator subunits.  $\alpha$ -tubulin was used as a loading control.

# Supplemental Figure S2

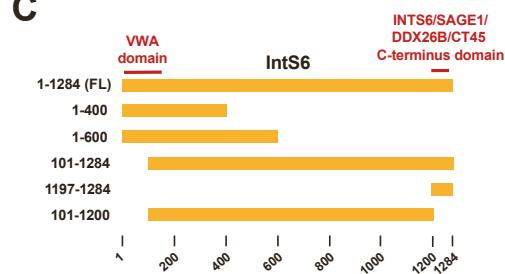
**A**



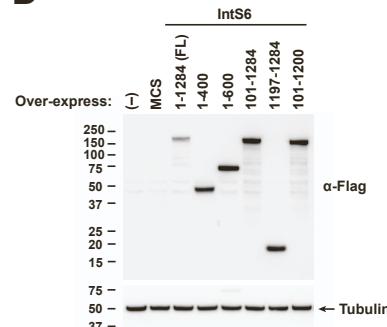
**B**



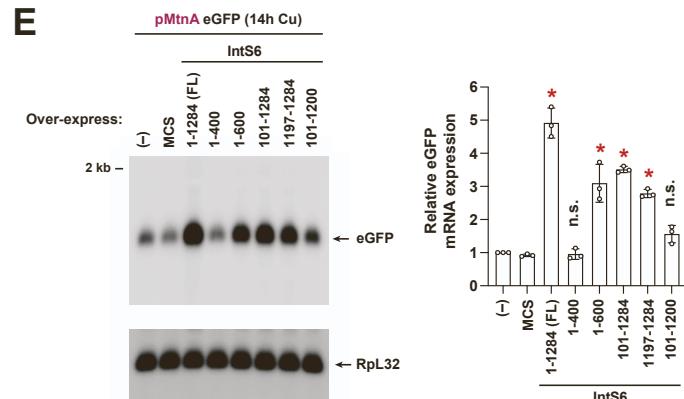
**C**



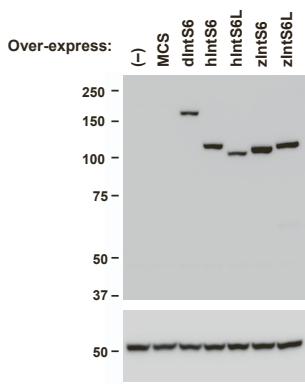
**D**



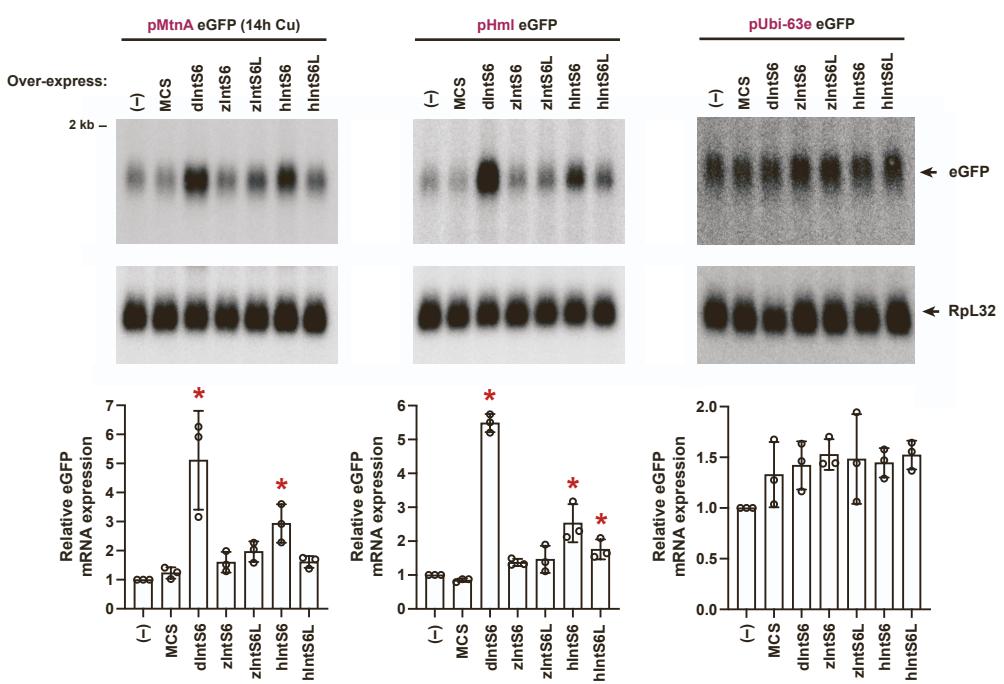
**E**



**F**



**G**



**Figure S2. IntS6 over-expression is sufficient to inhibit Integrator activity at select protein-coding gene promoters. Related to Figure 1.**

**(A, B)** *Drosophila* DL1 cells ( $5 \times 10^5$  cells per well) were seeded in 12-well dishes and co-transfected the next day with (i) a constant amount (300 ng) of pMtnA eGFP (**A**) or pUbi-p63e eGFP (**B**) reporter plasmid and (ii) variable amounts of plasmids that express FLAG-tagged IntS6, IntS12, or IntS8 from the Ubi-p63e promoter (0, 20, 40, 60, 80, 100, 150, or 200 ng). Empty vector (pUb 3xFLAG MCS) was added as needed so that 500 ng DNA was transfected in all samples. Total RNA was isolated ~40 h after transfection (500  $\mu$ M CuSO<sub>4</sub> was added for the last 14 h in **A**) and eGFP mRNA expression levels were analyzed by Northern blotting (20  $\mu$ g/lane). Data were normalized to the (-) samples where only the eGFP plasmid was transfected and data are shown as mean  $\pm$  SD,  $N = 4$ . (\*) P < 0.05; n.s., not significant.

**(C)** Schematics of plasmids that express FLAG-tagged full-length (FL) or shortened isoforms of the 1284 amino acid *Drosophila* IntS6 open reading frame from the Ubi-p63e promoter. Characterized protein domains in IntS6 are noted in red.

**(D)** 500 ng of each *Drosophila* IntS6 expression plasmid was transfected into DL1 cells and total protein was harvested after 48 h. A plasmid containing a multi-cloning site (MCS) was used as a control. Western blot analysis using an antibody that recognizes FLAG was used to confirm expression of each IntS6 protein isoform.  $\alpha$ -tubulin was used as a loading control.

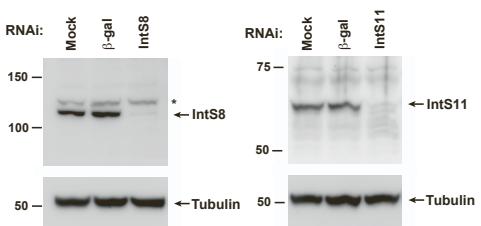
**(E)** DL1 cells ( $5 \times 10^5$  cells per well) were seeded in 12-well dishes and co-transfected the next day with 400 ng of pMtnA eGFP plasmid and 100 ng of the indicated IntS6 expression plasmid. Total RNA was isolated ~48 h after transfection (500  $\mu$ M CuSO<sub>4</sub> was added for the last 14 h) and eGFP mRNA expression levels were analyzed by Northern blotting (20  $\mu$ g/lane). Data were normalized to the (-) samples where only the pMtnA eGFP plasmid was transfected and data are shown as mean  $\pm$  SD,  $N = 3$ . (\*) P < 0.05; n.s., not significant.

**(F)** 500 ng of plasmid that expresses FLAG-tagged *Drosophila* (d), human (h), or zebrafish (z) IntS6/IntS6-like (IntS6L) protein was transfected into DL1 cells and total protein was harvested after 48 h. A plasmid containing a multi-cloning site (MCS) was used as a control. Western blot analysis using an antibody that recognizes FLAG was used to confirm expression of each isoform.  $\alpha$ -tubulin was used as a loading control.

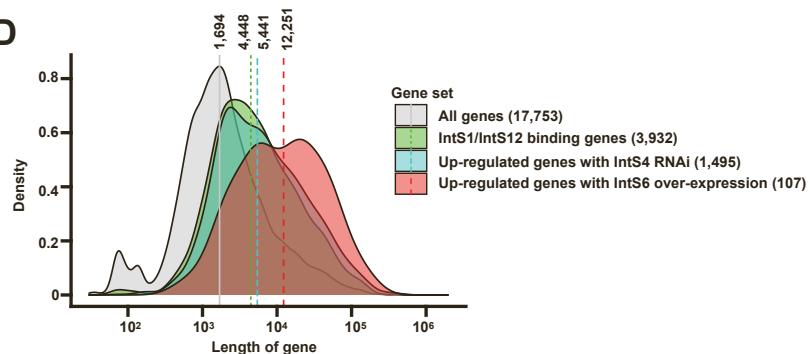
**(G)** DL1 cells ( $5 \times 10^5$  cells per well) were seeded in 12-well dishes and co-transfected the next day with 400 ng of the indicated eGFP plasmid and 100 ng of the indicated IntS6/IntS6L expression plasmid. Total RNA was isolated ~48 h after transfection (500  $\mu$ M CuSO<sub>4</sub> was added for the last 14 h only when measuring eGFP production from the MtnA promoter) and eGFP mRNA expression levels were analyzed by Northern blotting (20  $\mu$ g/lane). Data were normalized to the (-) samples where only the eGFP plasmid was transfected and data are shown as mean  $\pm$  SD, N = 3. (\*) P < 0.05; n.s., not significant.

# Supplemental Figure S3

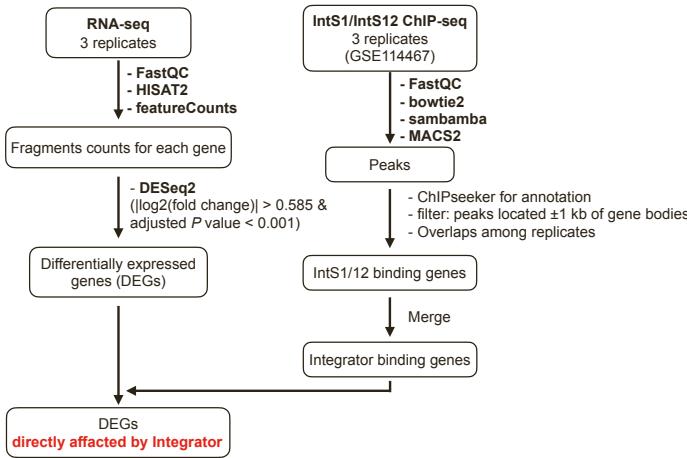
**A**



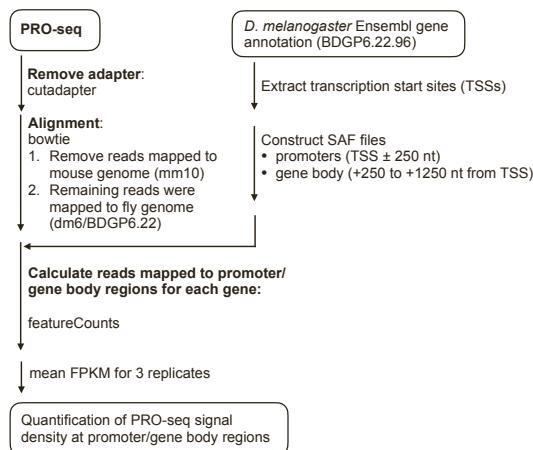
**D**



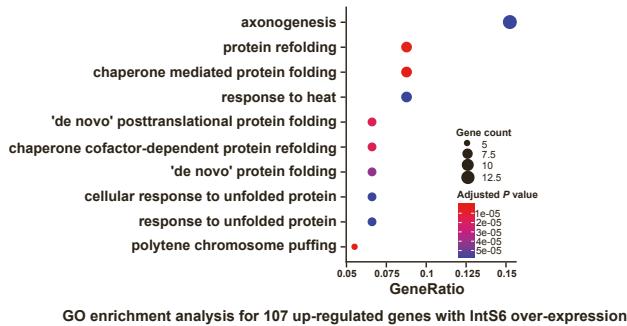
**B**



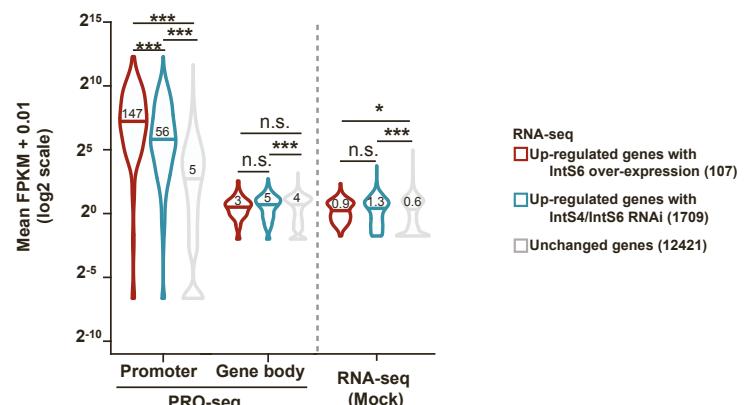
**E**



**C**



**F**



**G**

**IntS1/IntS12 ChIP-seq  
3 replicates**

Quality control and read trimming: FastQC, Trimmomatic  
Read mapping: bowtie2  
Remove multi-mapped reads: sambamba  
Peak calling: MACS2 (FDR < 0.001)

Peaks

Peak annotation: ChIPseeker

Sample	Peak counts	Closest genes
IntS1 rep1	3,515	2,388
IntS1 rep2	3,814	2,545
IntS1 rep3	4,118	2,699
IntS12 rep1	7,106	5,093
IntS12 rep2	6,971	5,056
IntS12 rep3	6,626	4,803

Keep peaks located ± 1 kb of gene bodies

Sample	Peak counts	Closest genes
IntS1 rep1	2,899	2,096
IntS1 rep2	3,132	2,231
IntS1 rep3	3,348	2,350
IntS12 rep1	5,891	4,535
IntS12 rep2	5,788	4,496
IntS12 rep3	5,509	4,270

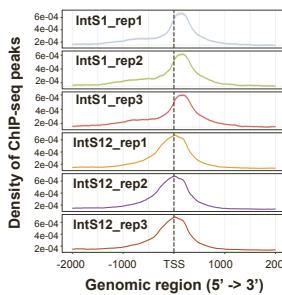
Gene overlaps among replicates

IntS1 or IntS12 binding genes

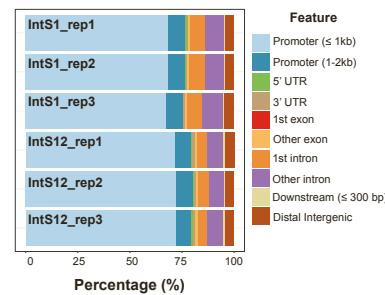
Merge

Integrator binding genes

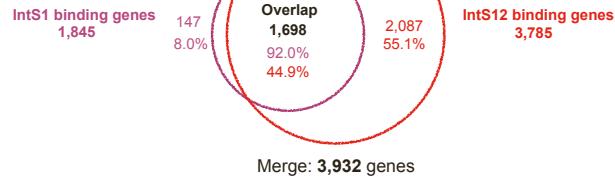
**H**



**I**



**J**



**Figure S3. Identification of endogenous differentially expressed genes upon IntS6 over-expression. Related to Figures 2 and 3.**

(A) To confirm the specificity of the IntS8, IntS11, and IntS12 antibodies, DL1 cells were treated with dsRNAs for 3 d and Western blots performed. \* denotes non-specific band.

(B) Parental DL1 cells or DL1 cells stably maintaining IntS6 or IntS12 transgenes driven by the copper inducible MtnA promoter were grown for 3 d. 500  $\mu$ M CuSO<sub>4</sub> was added for the last 24 h prior to total RNA isolation. rRNA depleted RNA-seq libraries were then generated, sequenced, and analyzed. Analysis pipeline used to identify genes that are both differentially expressed upon IntS6 over-expression and have IntS1/12 ChIP-seq signal  $\pm$  1 kb of the gene body is shown (also see **Figure S3G**).

(C) GO enrichment analysis (biological process) for the 107 genes up-regulated upon IntS6 over-expression was performed and visualized with clusterProfiler<sup>70</sup> (parameters: pvalueCutoff = 0.05, qvalueCutoff = 0.2).

(D) Length distribution of different gene sets as per the *D. melanogaster* Ensembl gene annotation (BDGP6.22.96). The vertical line indicates the median gene length of each gene set.

(E) Analysis pipeline used to quantify PRO-seq signal density, which denotes activity of engaged Pol II at promoter and gene body regions, in DL1 cells that had been treated with control ( $\beta$ -gal) dsRNA (GSE114467). <sup>27</sup>

(F) PRO-seq density in DL1 cells at promoter (TSS  $\pm$  250 nt) and gene body (+250 to +1250 nt from TSS) regions of different gene sets (left). Normalized RNA-seq counts from mock treated DL1 cells (3 replicates) (**Table S1**) for different gene sets are also shown (right). Violin plots represent range of values, with line indicating median. P values were calculated by Wilcoxon signed-rank test. (\*) P < 0.05; (\*\*\*\*) P < 0.001; n.s., not significant.

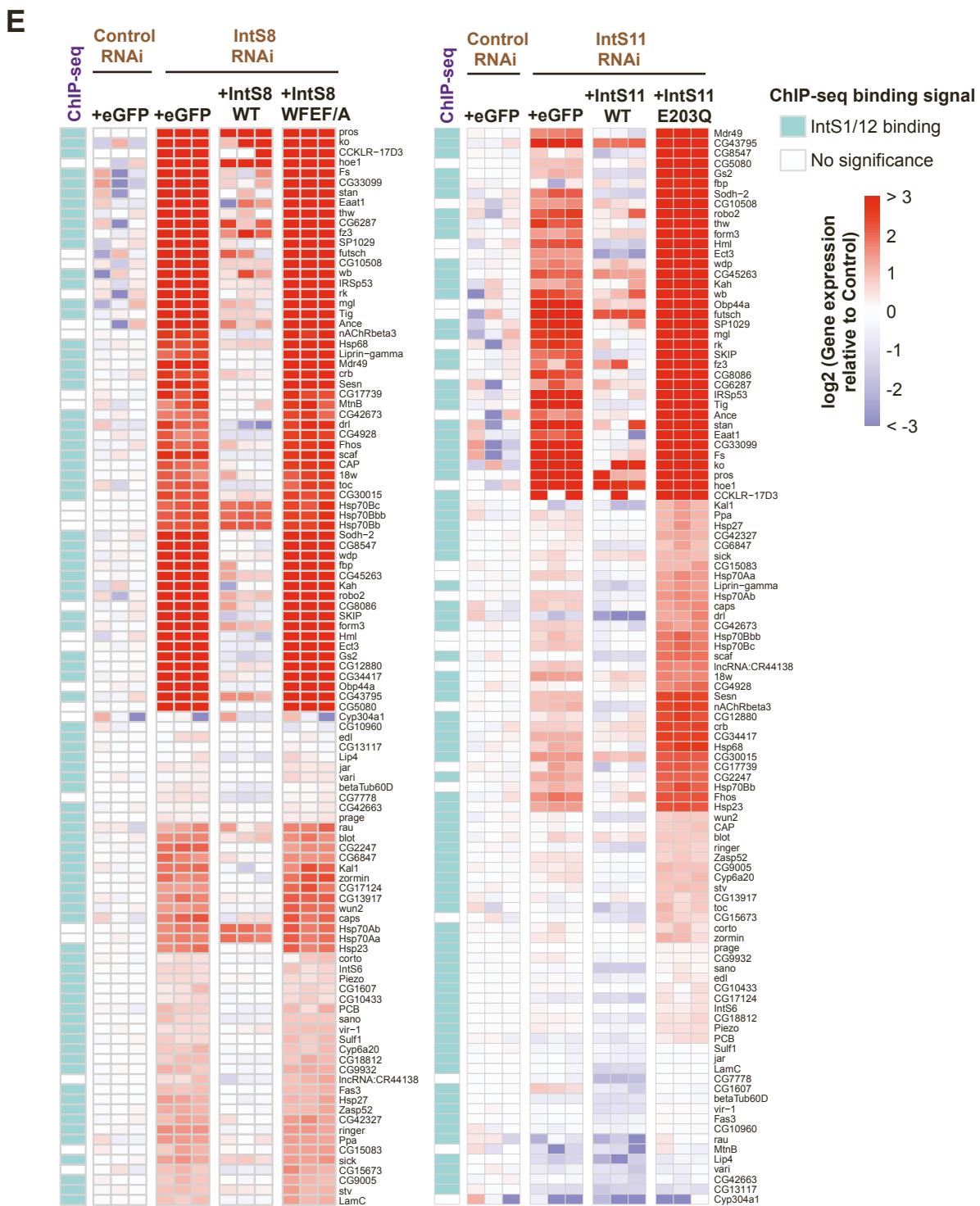
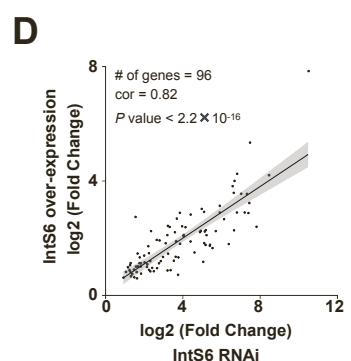
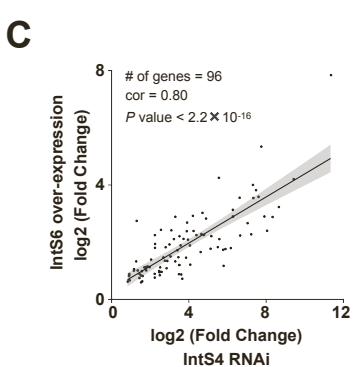
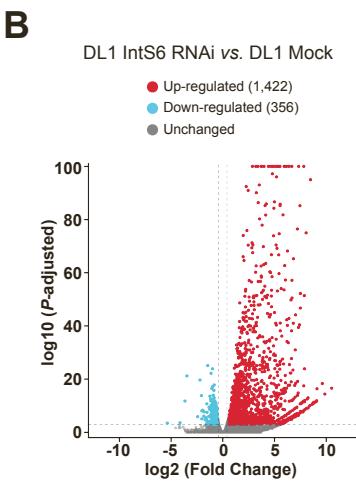
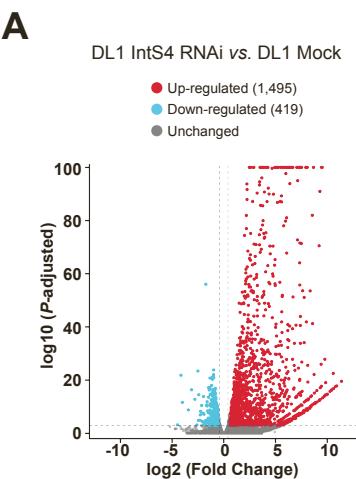
(G) Analysis pipeline used to identify genes that have peaks of IntS1 or IntS12 ChIP-seq signal  $\pm$  1 kb of the gene body in DL1 cells (GSE114467). <sup>27</sup>

(H) Densities of ChIP-seq peaks  $\pm$  2 kb of transcription start sites (TSS).

**(I)** Proportions of gene elements with mapped ChIP-seq peaks.

**(J)** Comparison of genes bound by IntS1 and IntS12 identified 1,698 genes that are bound by both subunits along with a total merged list of 3,932 genes bound by Integrator.

## Supplemental Figure S4



**Figure S4. Identification of differentially expressed genes upon Integrator subunit depletion. Related to Figure 3.**

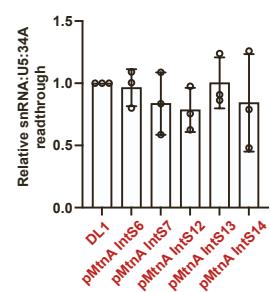
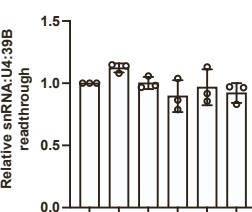
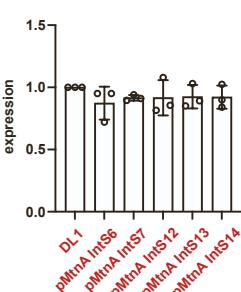
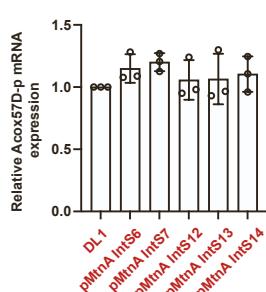
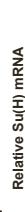
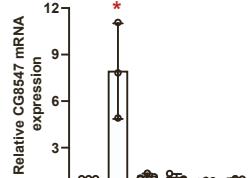
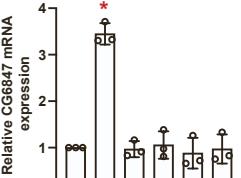
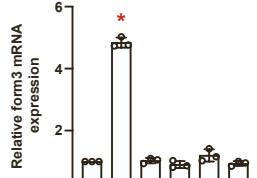
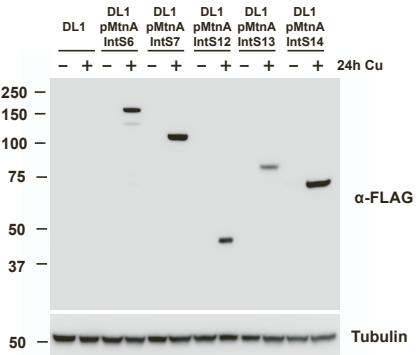
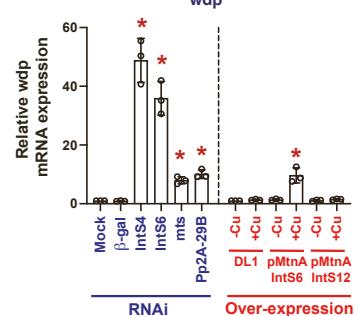
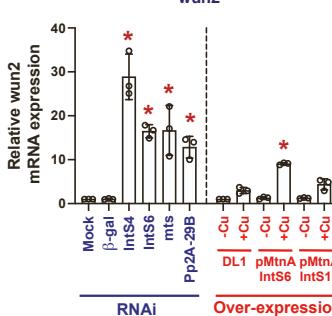
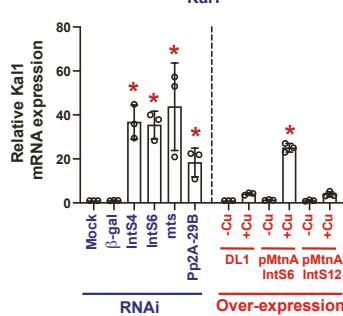
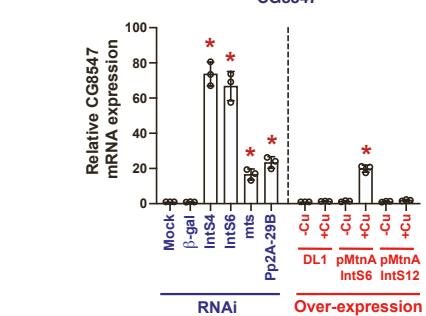
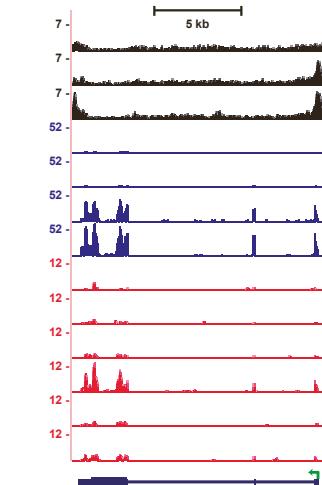
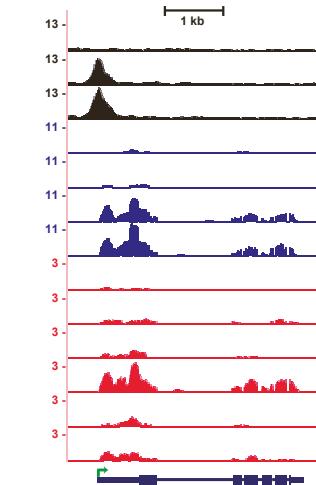
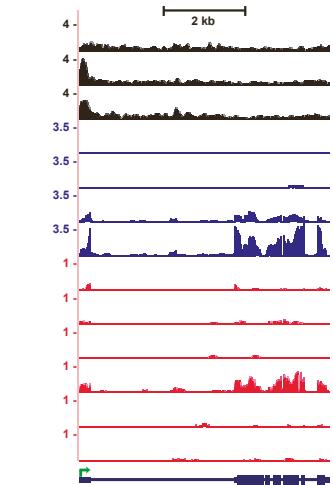
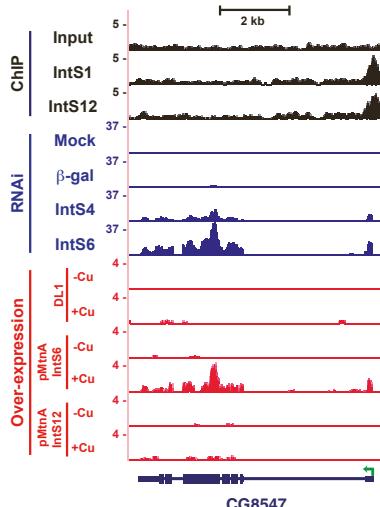
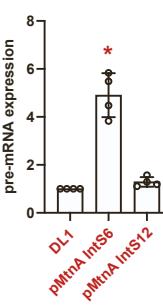
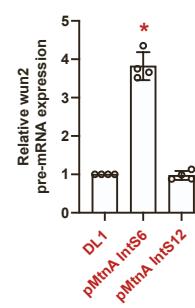
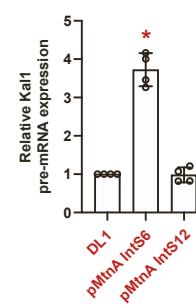
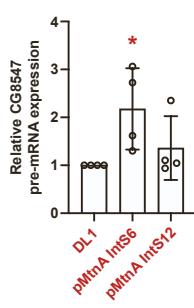
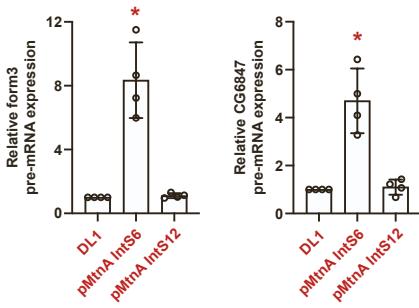
**(A)** DL1 cells were mock treated or treated with IntS4 dsRNAs for 3 d. Total RNA was isolated and rRNA depleted RNA-seq libraries were then generated, sequenced, and analyzed. The magnitude of change in mRNA expression compared with statistical significance (*P*-value) is shown as a volcano plot. Threshold used to define IntS4 affected mRNAs was ( $|\log_2(\text{fold change})| > 0.585$  and adjusted *P* value  $< 0.001$ ).

**(B)** DL1 cells were mock treated or treated with IntS6 dsRNAs for 3 d. Total RNA was isolated and rRNA depleted RNA-seq libraries were then generated, sequenced, and analyzed. The magnitude of change in mRNA expression compared with statistical significance (*P*-value) is shown as a volcano plot. Threshold used to define IntS6 affected mRNAs was ( $|\log_2(\text{fold change})| > 0.585$  and adjusted *P* value  $< 0.001$ ).

**(C, D)** The correlation of fold change for 96 up-regulated genes between IntS6 overexpression, IntS4 depletion **(C)** and IntS6 depletion **(D)**. The grey area indicates the confidence interval of linear regression. cor: Pearson correlation coefficient.

**(E)** Heat maps show relative expression levels of the 107 genes that are up-regulated upon IntS6 over-expression in published RNA-seq samples. On the **left**, DL1 cells were treated for 60 h with control ( $\beta$ -gal) or IntS8 dsRNAs and rescued using a stably integrated transgene expressing eGFP, wild-type (WT) IntS8, or IntS8 with a mutation (WFEF/A) that disrupts the interaction of IntS8 with PP2A (GSE150844).<sup>11</sup> Of the 107 genes that were up-regulated upon IntS6 over-expression, 89 (83%) were also significantly up-regulated upon IntS8 depletion, with 82 (85%) of these changes rescued by the WT IntS8 transgene. None of the gene expression changes were rescued with the WFEF/A mutant IntS8 transgene. On the **right**, DL1 cells were treated for 60 h with control ( $\beta$ -gal) or IntS11 dsRNAs and rescued using a stably integrated transgene expressing eGFP, wild-type (WT) IntS11, or IntS11 with a mutation that disrupts endonuclease activity (E203Q) (GSE114467).<sup>27</sup> Of the 107 genes that were up-regulated upon IntS6 over-expression, 33 (31%) were also significantly up-regulated upon IntS11 depletion, with 28 (85%) of these changes rescued by the WT IntS11 transgene. Only 1 (3%) of the gene expression changes were rescued with the E203Q mutant IntS11 transgene.

## Supplemental Figure S5



**Figure S5. Changes in endogenous pre-mRNA and mRNA levels upon individual over-expression of Integrator subunits. Related to Figure 3.**

**(A)** Parental DL1 cells or DL1 cells stably maintaining IntS6 or IntS12 transgenes driven by the copper inducible MtnA promoter were grown for 3 d. 500 µM CuSO<sub>4</sub> was added to each for the last 24 h prior to total RNA isolation from four independent biological replicates. Expression of the indicated pre-mRNAs (using primers that amplify across an exon-intron boundary) for genes whose mRNAs were up-regulated upon IntS6 over-expression was quantified using RT-qPCR. Data are shown as mean ± SD, N = 4. (\*) P < 0.05.

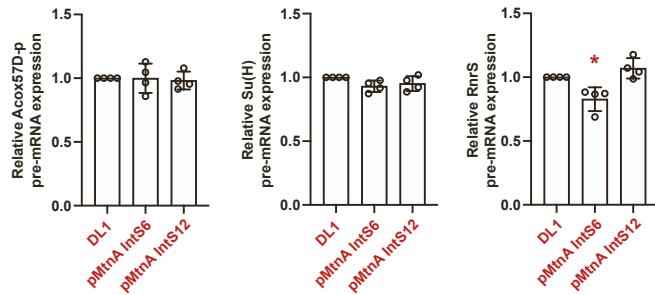
**(B-E)** UCSC genome browser tracks (top) and validation RT-qPCR experiments (bottom) for protein-coding loci that are affected by IntS6 over-expression. IntS1 and IntS12 ChIP-seq profiles in DL1 cells (GSE114467) are shown in black. RNA-seq data generated from DL1 cells treated for 3 d with control ( $\beta$ -gal), IntS4, or IntS6 dsRNAs are shown in blue. RNA-seq data generated from parental DL1 cells or DL1 cells stably maintaining inducible IntS6 or IntS12 transgenes are shown in red. 500 µM CuSO<sub>4</sub> was added for the last 24 h as indicated. Green arrow, transcription start site (TSS). (Bottom) Expression of the indicated mRNAs was quantified using RT-qPCR. Data are shown as mean ± SD, N = 3. (\*) P < 0.05.

**(F)** The parental DL1 cells as well as stable cell lines that express FLAG-tagged IntS6, IntS7, IntS12, IntS13, or IntS14 from the copper-inducible MtnA promoter were seeded in 12-well plates (5 x 10<sup>5</sup> cells per well) and grown for 3 d. As indicated, a final concentration of 500 µM CuSO<sub>4</sub> was added to cells for the last 24 h prior to harvesting total protein. Western blot analysis was then performed using antibodies that recognize FLAG.  $\alpha$ -tubulin was used as a loading control.

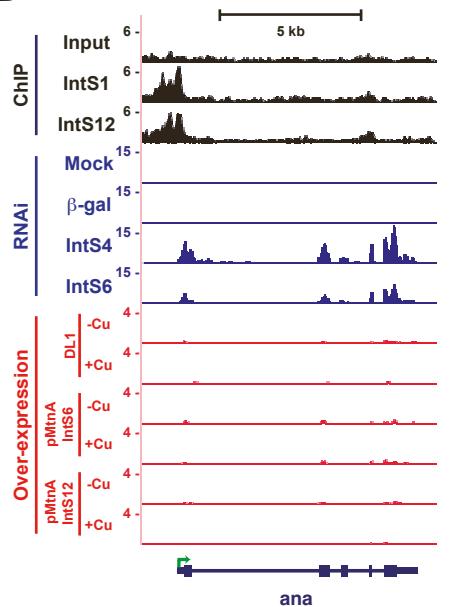
**(G)** The parental DL1 cells as well as stable cell lines that express FLAG-tagged IntS6, IntS7, IntS12, IntS13, or IntS14 from the copper-inducible MtnA promoter were seeded in 12-well plates (5 x 10<sup>5</sup> cells per well) and grown for 3 d. A final concentration of 500 µM CuSO<sub>4</sub> was added to cells for the last 24 h prior to harvesting total RNA. Expression of the indicated transcripts was quantified using RT-qPCR. Data are shown as mean ± SD, N = 3. (\*) P < 0.05.

# Supplemental Figure S6

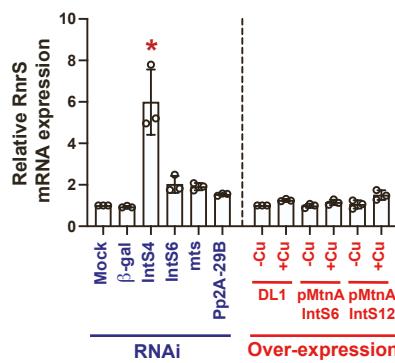
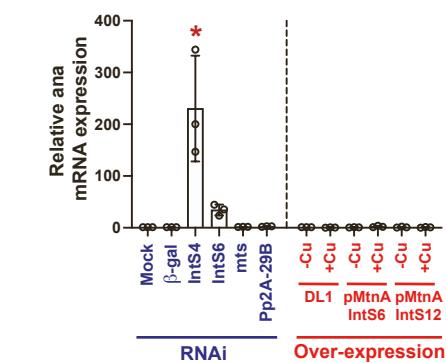
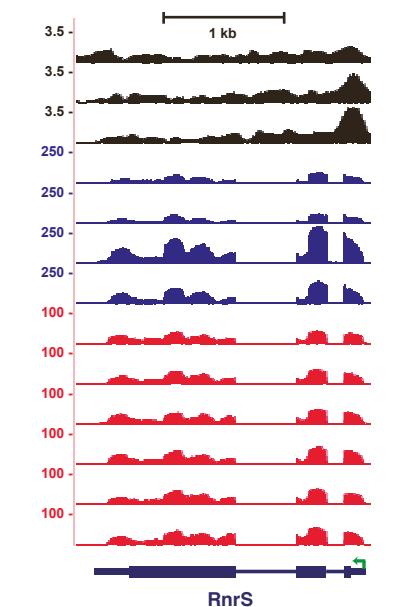
**A**



**B**



**C**



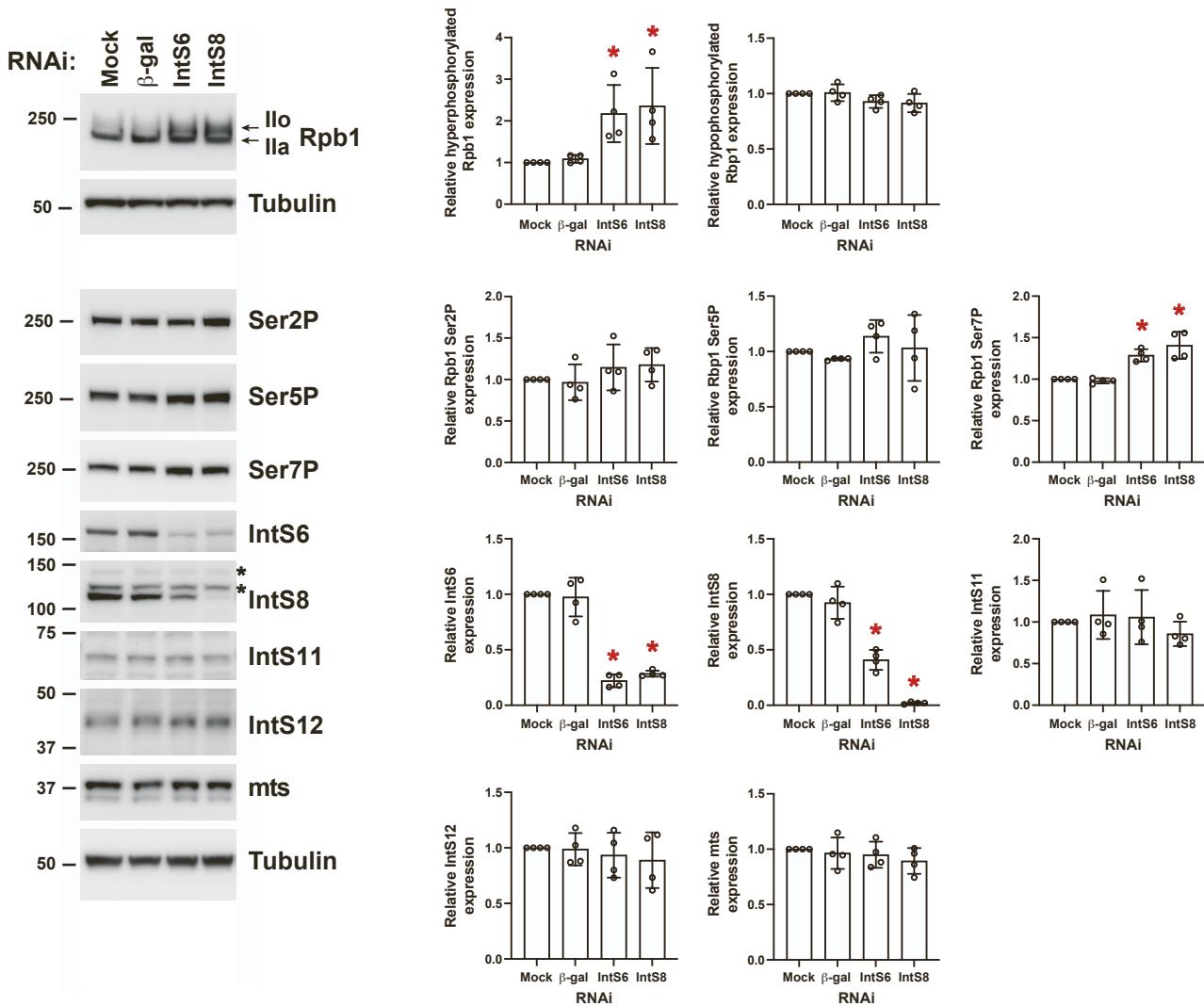
**Figure S6. Examples of protein-coding genes that were unaffected by IntS6 over-expression. Related to Figure 3.**

**(A)** Parental DL1 cells or DL1 cells stably maintaining IntS6 or IntS12 transgenes driven by the copper inducible MtnA promoter were grown for 3 d. 500  $\mu$ M CuSO<sub>4</sub> was added to each for the last 24 h prior to total RNA isolation from four independent biological replicates. Expression of the indicated pre-mRNAs (using primers that amplify across an exon-intron boundary) for genes whose mRNAs were unchanged upon IntS6 over-expression was quantified using RT-qPCR. Data are shown as mean  $\pm$  SD, N = 4. (\*) P < 0.05.

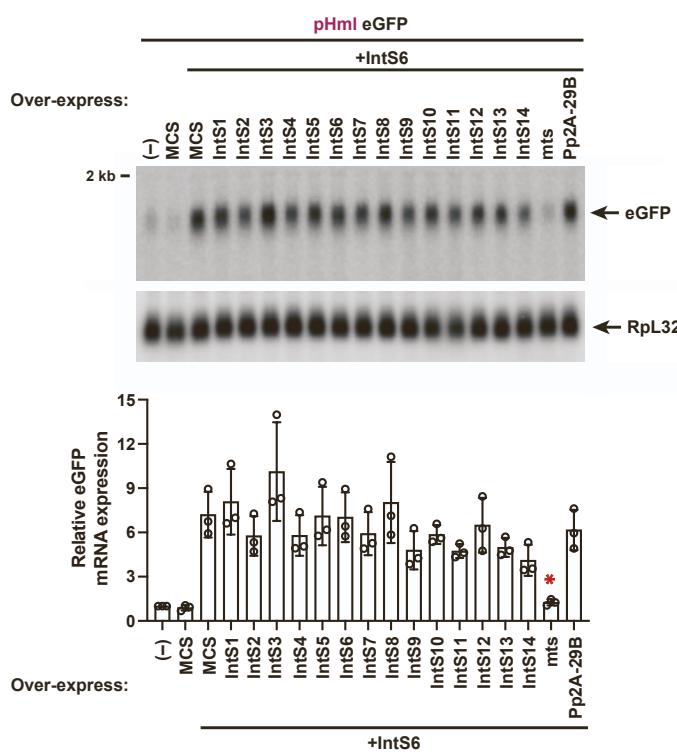
**(B, C)** UCSC genome browser tracks (top) and validation RT-qPCR experiments (bottom) for protein-coding loci that are not affected by IntS6 over-expression. IntS1 and IntS12 ChIP-seq profiles in DL1 cells (GSE114467) are shown in black. RNA-seq data generated from DL1 cells treated for 3 d with control ( $\beta$ -gal), IntS4, or IntS6 dsRNAs are shown in blue. RNA-seq data generated from parental DL1 cells or DL1 cells stably maintaining inducible IntS6 or IntS12 transgenes are shown in red. 500  $\mu$ M CuSO<sub>4</sub> was added for the last 24 h as indicated. Green arrow, transcription start site (TSS). (Bottom) Expression of the indicated mRNAs was quantified using RT-qPCR. Data are shown as mean  $\pm$  SD, N = 3. (\*) P < 0.05.

# Supplemental Figure S7

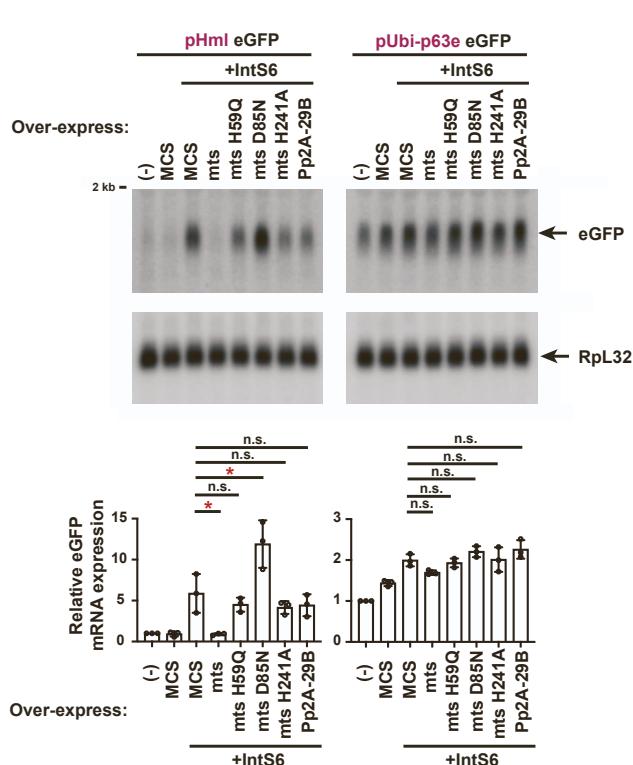
A



B



C



**Figure S7. Changes in CTD phosphorylation upon Integrator subunit depletion and a critical role for the mts phosphatase activity. Related to Figures 2 and 4.**

**(A)** DL1 cells were treated with dsRNAs for 3 d and Western blots performed using antibodies that recognize Rbp1 (C-terminal domain), Rbp1 phosphorylated at Ser2, Ser5, or Ser7, IntS6 (amino acids 1035-1284), IntS8 (amino acids 1-308), IntS11 (amino acids 300-597), IntS12, or mts. Ilo denotes hyperphosphorylated Rbp1, while Ila denotes hypophosphorylated Rbp1.  $\alpha$ -tubulin was used as a loading control. \* denotes non-specific band. Data were normalized to mock treated cells and are shown as mean  $\pm$  SD,  $N = 4$ . (\*)  $P < 0.05$ .

**(B)** DL1 cells were co-transfected with 300 ng of eGFP reporter plasmid, 100 ng of IntS6 over-expression plasmid (driven by the Ubi-p63e promoter), and 100 ng of the indicated Integrator/PP2A subunit over-expression plasmid (driven by the Ubi-p63e promoter). Empty vector (pUb 3xFLAG MCS) was added as needed so that 500 ng DNA was transfected in all samples. Northern blots (20  $\mu$ g/lane) were used to quantify expression of the eGFP reporter mRNA. Representative blots are shown and RpL32 mRNA was used as a loading control. Data are shown as mean  $\pm$  SD,  $N = 3$ . (\*)  $P < 0.05$ .

**(C)** DL1 cells were co-transfected with 300 ng of eGFP reporter plasmid, 100 ng of IntS6 over-expression plasmid (driven by the Ubi-p63e promoter), and 100 ng of the indicated PP2A subunit over-expression plasmid (driven by the Ubi-p63e promoter). Northern blots (20  $\mu$ g/lane) were used to quantify expression of each eGFP reporter mRNA. Representative blots are shown and RpL32 mRNA was used as a loading control. Data are shown as mean  $\pm$  SD,  $N = 3$ .  
(\*)  $P < 0.05$ ; n.s., not significant.

## Methods S1: Expression plasmid construction, Related to the STAR Methods.

The following *Drosophila* expression plasmids were generated from the previously published **Hy\_pPepck1 eGFP SV40** plasmid (Addgene #132644; Tatomer et al. (2019) *Genes Dev*, 33, 1525-1538):

<b>Hy_pRnrS eGFP SV40</b>	<b>(Addgene #195062)</b>
<b>Hy_pAna eGFP SV40</b>	<b>(Addgene #195063)</b>
<b>Hy_U5:34A eGFP SV40</b>	<b>(Addgene #195064)</b>

The Pepck1 promoter (marked in gray) drives expression of eGFP (marked in green) followed by the SV40 polyadenylation signal (marked in yellow). The full sequence of **Hy\_pPepck1 eGFP SV40** is as follows:

GGTTAATAAGGAATATTGTAGTACTGCCTTGACTAGAGATCATAATCAGCCATACCACATTGTAGAGGTTTACTTGCTTAAAAAACCTCCC  
ACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTAACTTGTATTGCACTTATAATGGTACAAATAAGCAATAGCATC  
ACAATTTCACAAAATAAGCATTTCACTGCATTCTAGTTGTGTTGTCACATCAATGATCTTATCATGTCTGGATTAAATTAAAAAA  
ACAGAGCCATATTACGTCAACTTCCCAGCTGCTCTGTTGAACCTTTCTCGAGTACCCAGAGGATTCTGGTCTCTGGTAAACTGGATTG  
GGTACAAAGTCACACCACCCGAGCACCTCATTCTCCTTCCCAAAGACAAAACAAGAGTGGCCACATGACAACCATCAGGCCATGCTATT  
TTGGTGCCCCCCAGAGTCACATGAAATGTAATCTGGCCTCAAGCGCCATGGCAATTGGGAGGCCCGGCTCATCTGGATGCCGAGGG  
GGCGGATTAGCAGCAAAATTTCGTCACCGATTCAAAATAAGCAGAAAGTTTTCTGATTGCAATGTCACAAACGAGAGGCCAACACAT  
GTGGTCAACAGGCTAAGTGGCGGGCGCTGTGATGTGGCCACATTGATGTTGCAACATGTAGTATATTGTGGCAATGAGCTATAGA  
CACATCCAGAATTACCCAGCTGAGGGACAACACTGCTGTTGCTCAGCTGCTCTGAACTGGTTCTATATTGCAAGCGGCCACATTGATC  
GCATAAAATCATAAATAAGATAATTAAATAACAGCGCAGGAAAGTGGCAAATATCAGGGTGTCAAATAACAAACAGCTTATCAATGCAACAGCA  
CTAGAAAACGCTCTCAGGCCACATGAAAACAACACTCTGCTGTTGCCATGTGGGACATACCCATATACTGATCTGATGCCATGCC  
CCGTAGACCCCCCCCCCACCACAAACCAACTCATCGTAGCTGCCCTGTGAAAACCAAACAATGCCAGGCCCTGGCAGGCCAACATCAG  
CTCGTCCGGATTGACCCCTCACTCGCGAGGCGATCGTAAACCCGATTGCGTGGCAGACTCGTGGCAGAGTAACCGAGGTGACGGTATCGATAAGCTG  
CGTGCAGGGCTGGTAAAGGGCCGGATCGGAGCTGGCATGTTGCGCAGACTCGTGGCAGAGTAACCGAGGTGACGGTATCGATAAGCTG  
TATCACCATGGTGAACAGGGCAGGGAGCTGTTCACGGGGTGTGGCCATCCTGGTCAAGCTGGCAGCTGGACGGCAGCTAACGGCACAAGTTCAGCGTG  
TCGGCGAGGGGAGGGCGATGCCACCTACGGCAACGTCAGCTGAATTCTGCAACCGGCAAGCTGCCGTGCCCTGGCCACCTCGTGA  
CCACCTGACCTACGGCGTGCAGTGCCTCAGCGCTACCCGACCACATGAAGCAGCACGACTCTTCAAGTCCGCCATGCCGAAGGCTACGCTCA  
GGAGCGCACCATTCTCAAGGACGAGCAACTACAAGACCCGCCAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAG  
GGCATCGACTCAAGGAGGACGGCACATCTGGGCAAAAGCTGGAGTACAACACTACAGCACAACGTATATCATGCCGACAAGCAGAAGA  
ACGGCATCAAGGTAACCTCAAGATGCCACAAACATCGAGGACGGCAGCTGCAGCTGCCGACACTACCAGCAGAAACCCCCATGCCGACGG  
CCCCGTGCTGCTGCCGACAAACCAACTCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCAAACGAGAAGGCCGATCACGTTCTGCTGAGTT  
GTGACCGCCGCCGGATCAGTCAGGCACTGAGCTACAAGTAACGGCCGAATTGTTATTGCAAGCTTATAATGTTACAATAAGCAAA  
TAGCATCACAAATTCAAAATAAGCATTTCAGTCATTGAGTGTGGTTGTCACATCAATGTTACTTCAGCTGAGCTGTTGCTGAGCTTAA  
GGGGCCGCCAGCGAGGCTGGATGGCCTCCCCATTATGATTCTTCGCTCCGGCGATCGGATGCCCGTGTGCAAGGCCATGCTGCCAGG  
CAGGTAGATGACGACCATCAGGGACAGCTCAAGGATCGCTCGCGCTTACAGCTTAACCTCGATATTGACCGCTGATCGTCACGGCATT  
ATGCCGCCTCGCGAGCACATGAAACGGGTGGCATGGATTGAGGCCGCCCTATACTTGTCTGCCCTCCCGCGTTGCGCTGCGGTGATGGAG  
CCGGGCCACCTCGACCTGAATGGAAGGCCGCCACCTCGCTAACCGATTACCAAGAATTGAGGCCATCAATTCTGCGGAGAACCTGTA  
ATGCGCAACACCAACCTTGGCAAGAACATCCATCGCTCCGGCATCTCCAGCAGCCAGCAGCGCATTCCGGCAGCGTGGCTCTGGCCACG  
GGTGCAGTATGCTGCTCTGTCGTTGAGGACCCGGCTAGGCTGGGGGTTGCCCTACTGGTAGCAGAAATGACCGATAACGCGAGCAGAAG  
TGAACGCACTGCTGCAAAACGCTGCACTGAGCAACAAACATGAATGGTTCTGCTTCCGTTGCTAAAGTCTGAAACGCGGAAGTCAG  
CGCTCTCGCTTCTCGCTACTGACTCGCTGCCGCTGGTGTGGCTGCCGAGCGGTATAGCTCACTAAAGCGGTAAATCGGTTACCA  
CAGAATCAGGGATAACGAGGAAAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGCCCGTGTGGCGTTTCCATAG  
GCTCGCCCCCTGACGAGCATCACAAATGAGCCCTCAAGTCAAGTCAAGGCTGAGGACTATAAGATAACAGGGCTTCCCTGG  
AGCTCCCTCGTGCCTCTCGTGTGGAGGACCCGGCTACCGGATACCTGTCGCTTCTCCCTCTGGAGCGCTGGCTTCTCATGCTCAC  
GCTGTAGGTATCTGAGTCTGGTGTAGGTCTCGCTCAAGCTGGGTGTGTCAGCAGACCCCGGTTAGCCGACCGCTGCCCTTACCGTAA  
CTATCGTCTGAGTCAACCCGGTAAGACAGACTTACGCCACTGCACTGCCAGCAGCACTGTAACAGGATTAGCAGAGCAGGTATGTTAGGCG  
ACAGAGTTCTGAGTGGCTAACTACGGCTACACTAGAAGGACAGTATTGTTGATCTGCGCTCTGTAAGCCAGTACCTTCGGAAAAAGAG  
TTGGTAGCTCTGATCCGGCAAACAAACACCACCGCTGGTAGCGGTGGTTTTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA  
AGATCCTTGTACTTCTACGGGTCTGACGCTCAGTGGAAACGAAACTCACGTTAAGGGATTGTTGCTAGAGATTATCAAAAGGATCTCACC  
TAGATCCTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATGAGTAAACTGCTGACAGTACCAATGCTTAATCAGTGAGGCAC  
CTATCTCAGCGATCTGCTATTCTCATCCACTGCTGACTCCCCGTCGTAGATAACTACGATACGGAGGGCTTACCATCTGCC  
TGTGCAATGATGACCGCGAGACCCACGCTACCCGGCTCAGGATTTATCAGCAATAAACAGCCAGCGGAGGGCCAGCGAGAAGTGGCCTGCA  
ACTTATCCGCCATCCAGCTTAAATTGTTGCCGGAGGCTAGAGTAAGTAGTGTGCAAGCTTAAAGTGTGCAACGTTGCTGCCATTGCTG  
CAGGCATCGTGTGTCAGCTCGTGTGGTATGCCCTCATCGCTGCCGTTCCAACGATCAAGGAGGTACATGATCCCCATGTTGCAA  
AAAAGCGGTAGCTCTCGCTCCGATCGTGTGAGAAGTGGCTGAGGAGCTTACCTGAGCTTACGTTACGTTGCAACGTTGAGGAA  
ACTGTCATGCCATCCGTAAGATGCTTTCTGTGACTGGTAGACTCAACCAAGTCATTCTGAGAATAGTGTATGCCGACCGAGTTGCTTG  
CGCGCTCAACACGGATAATACCGGCCACATAGCAGAACTTTAAAGTGTCACTATTGAAACGTTCTCGGGGCAAACACTCTCAAGGATCTT  
ACCGCTGTTGAGATCCAGTGTGAGTAAACCAACTCGTGCACCCAACTGATCTCAGCATCTTTACTTCAACGAGCTTCTGGGTGAGCAAAACA  
GGAAGGCAAAATGCCGCAAAAAGGGATAAGGGCGACCGGAATGTTGAATACTCATCTCCCTTCAATATTATTGAGCATTATCAGG

ACCGCTGTTGAGATCCAGTCGATGTAACCCACTCGTCGACCCAACGTGATCTTCAGCATCTTACTTCAACAGCCTGGGTGAGCAAAAACA  
 GGAAGGCAAAATGCCGAAAAAGGAATAAGGGCAGACGGAAATGTGAATACTCATCTCCCTTTCAATATTGAAGCATTATCAGG  
 GTTATTGTCATGAGCGGATACATATTGAATGTATTAGAAAATAACAAATAGGGCTCCGCACATTCCCCGAAAGTGCCACCTGACGT  
 CTAAGAACCATATTATCATGACATTAACCTATAAAATAGGCATCACGAGGCCCTTCGCTTCAAGAATTACATTTGACGATTTTT  
 TTATCAAAAGTCGAGTTTCAACATTCTCATCAACCGAGCAAGGCAACGGCTTGAATAATATGTTATATACATATCAATCGC  
 TGCTGACTGCGTATTGATGCCCAAGATTACATATTGAAATCAGGATTCAAAGGAGATCAATGTCAAATGCCGACAGGAACATGAAAGACG  
 CTGTTATGCGCAATTAAAATTGGTTAATTGCTGAACTGTTGGCATCTTAAGTCTGTTAACAAACATCAACTACTATGTAC  
 GTAGAAGCGTTAACCCATTGACATACAGATGAGAATCTGGTTGCTTAATCAGTCAGAATGACTCCGATGATGACTCATTACCTGACCAGT  
 TTCGCTGCTTCTTCACAAACTACTGTATGTTGATCTGAAATAGCAATACATTGAAATTCACATGGTCTAGTCAGCTTATTATCATTAAATT  
 GACACCAAGTCGTTATTGAGCTATCGAGCTCAACATTATCCATGAATAAGCCGAAACATATGCAATCTATGAAAGTT  
 ATATAAGCACACCTACTTGAACATACAACTTGTGCAATAGGTTCAACGAAATTGAGGCAATTTCGATAACCCAAGCGATTGACT  
 GTTCCGTTGATTCCAATGAAATTGAAATGACATAGTTGCTATATGTCAGTCAGTCAGCTTATCTCTGGGTTTCTCAGAGT  
 TTCGAAACGCTTCTTTGGATCTGGTATTTGGAACTCTGGTATTTGGAAAGGGCTCCCTGGAAATTGTTACACTGCTTATCATT  
 GCGAACAGCGCCGAAGCTATCGCAGCTTAACATTACAATGCACTTTTACGACCAATAATGACATTTCGCTTCTCGCCGTTGAT  
 AACCGAACGCGATGTCGCGCAGGAATGTTGCTTGCACACAAACGCAATCAAATGGATCAATTTCGCTTCCAGTGAACAGGAGAAC  
 GAAACGACCATATGATATGCTCTGATGTCATGGTATTGAAATCAATGCAATTAAAGCCCGCTGTCATGGCTTCTCGCC  
 GAAATGCTGATAACGCTGCTCCTCAACTGCTTGATGGACAACTTCCATTATTTGGATCGTTAAATTAAAGCG  
 CCTGTTACGCAATTACGTTGTTCCGGTGCCTGGTTACGTTCTGGAAAGCGAAATGGTTTAGGATTGGAAACCCCTCATCATGTT  
 GGAATATACTATTCAACCTACAAAGTAACGTTAACACACTACTTTATATTGATATGAATGCCACACCTTATGCCATAAAACATATTGAA  
 GAGAATACCACCTTTTATTCCTCTTCTGTACGTTTGTGTAAGTAGGCTGTTGCTGGTGTGAGTTGAAATAACTAAATA  
 TAAATCATAAAACATAACACTTGTACATTATTATTAAAGAAAGGAAATAAAATTAAACAGGTTATGGACCTGCGACCCA  
 AGCTGGCGCTGTCGGGGCAATGAGATATGAAAAAGCCTGACCTACCGCGACGCTGTCGAGAAGTCTGATGAAAGTTGACACCGCT  
 CCGACCTGATGCGAGCTCTCGAGGGCGAAGAATCTGCTGCTTCACTGCTGATGAGGAGGGCTGGATATGCTCTGCGGGTAATAGCTGCGCGA  
 TGGTTCTACAAAGATGTTATGTTATCGCAGCTTCCGGCCTGGCGCTCCCGATTCCGGAAAGTGTGACATTGGGAAATTGAGCAGGCTG  
 ACCTATTGCACTCCCGCGTGCACAGGTTGCACTGGCAAGACCTGCTGAAACCGAACCTGCGCTGTCAGGCGAGGCGATGG  
 ATCGGATCGCTGCCGATCTAGCAGACGGGTTGGCCATTGGACCGAAGGAATCGTCATAACTACATGGCGTATTGATATG  
 CGCAGATTGCTGATCCCCATGTTGATCTGGCAACTGCTGACGGGATTCGGCTCAACAATGCTCTGACGGACAATGGCCGATAACAGCGGT  
 CATTGACTGGCAGGACTGCCGAAGTCCGGCACCTCGTGCACGGGATTCGGCTCAACAATGCTCTGACGGACAATGGCCGATAACAGCGGT  
 CGAGCAGGCGATGTCGGGATTCCAATACGAGGTTGCCAACATCTTCTGGAGGCCGTGTTGCTGATGGAGCAGACGCGCTACTT  
 CGAGCAGGAGGCACTCGGAGCTTGCGAGGCTGCGCCTCCGGCTATATGCTCGCATTGGCTTGTGACCAACTCTAICAGAGCTGGITGACGGC  
 AATTTCGATGATGCGCTGGCGCAGGGTCATGCGACGCAACTCGCCGAGGGACTGTCGGGCTGACAAATCGCCGAGAAGCG  
 CGGGCTCTGGGAGCTGGCTGTGAGAAGTACTCGCCGATAGTGGAAACGCGCCAGCAGCTGTCAGGCGAAAGGAATAGAGT  
 ACCGAAACAAGAGCTGATTGAGAACGCCCTCAGGCCAACACTCGCGGAGCCTAGCAAGGCAAATGCGAGAGAACGGCTTACGTTGG  
 TTCTCGTCCACAGTCGCTAAGCTCGCTGGCTGGGTCGCGGGAGGGCGGTGCACTGATTGAGGCTTCTGGATTGTTGG  
 GATTGTCATGCCCAACGCACTGGGTGACTGACTGATCCCAGATTGGAGATGCCCGCTGCGATTGGGTGCAAGATCTTGTAAGGAA  
 CCTTACTCTGTGGTGTGACATAATTGACAAACTACCTACAGAGATTAAAGCTAAAGTAAATAAAATTGTTAAGTGTATAATGTTAAAC  
 TACTGATTCTAATTGTTGTTAGATTCAACCTATGGAACTGATGAATGGGAGCAGTGTGTTGAAATGCTTAAATGAGGAAACCTGTT  
 CTCAGAAGAAATGCCATCTAGTGTGAGGACTACTGCTGACTCTAACATTCTACCTCCAAAAGAAGAGAAGGTTAGAAGACCCAAAGAC  
 TTCTCTGCTTCAAGTTTTGAGCTGATGCTGTTAGTAATAGAATCTGCTTGTGCTTAACTACACCAAAAGGAAAAGCTGAC  
 TGCTATACAAGAAAATTATGAAATATTCTGAACTTATAAGTAGGCATAACACTTAAATACATACACTGTTTCTTACTCCACACAG  
 GCATAGAGTGTGCTATTAAATAACTATGCTAAAATTGTTAGCTTAAAGG

**To generate Hy\_pRnrS eGFP SV40, the following sequence was inserted between the Pacl and Xhol sites:**

CGATTAAACGTACTTATCGATAAAATCAACAGCCAGCTTATGTTATTGACTGTTATTACATGATTCGCTGTAATGAAATTCAAAAGTTG  
 TTGAAATTGAGTATGTTAGCACCTGTCATTCACAAATTGAGATTCCAAAAAAATTAAATTGAAATTCAATGTCCTGGAAAAAAATT  
 AACTTGTTCACCTGATTTTATCTAATTGAGCAAGAGGACCAATTTCACTCAAAACTAAATTGAGCTTCACTGCTGTTCTG  
 GATTAGTCTAGGGTGTCTCATGCAAAATCGATGATAAAATTCAAAATTGTTAGTATTGAAACTATTAAATTAAATTGTTCA  
 GTGACAAGCTGGAAGCTAACATAAAATTGCACTAAGGATTGATTGTTAAGAAAAGAAAACCTACCCACCCATAATTGCAATT  
 CCTAAATTATAAAAGTGAATTGACGCACTCGACAGCCCTGATTCCCATAGTTCCCATACCAAAATGGCGCAATCGAAACAGTTGC  
 CGCCGGCATAAAACCAATGAGCTGATTCCCATGCACTTGCACACAACCTGACACTTAAACACGTCGTTAAGTGAATT  
 TTACCAAGAGAATCAGAGAACGCCCTACTGCTAATAATCATTCAAAACATCTCAA

**To generate Hy\_pAna eGFP SV40, the following sequence was inserted between the Pacl and Xhol sites:**

ATATATCTGTTTCATATCCCCCACACCACACCACACCCACCTACGGGGGATTATGAAATCTGGGGCATTGATGCGATT  
 GCAATTTCACATAACACATATCACTCGACGTCTGGAGAAGCCGCCACTCACCACCTATTCAAGACTAAACGGTGTGAGAC  
 ACATCACATGAATGGCAGCGGGAAAGACCGAATCCAGACTTAAAGTGTGCAATTACAAATTGAAATTCAATTGGTCCAGACGT  
 CGCGGAGTGTGAAAAGATGTTATGTCGAAATAAAATAAAATCAGGAACGCTCTGAAAGTCTGCTGTTGGATTCCGTATGCGAA  
 TGGAAATGAAATAATTGAGGCTTGTCTCATGCGTAGTACCGAGAAGTGGCCAAAGGATGGGCTTCTGGGTTAAGCGCCAAATGCCCA  
 AAGCCGTGCCAGGACCGCACGCACTTAACTAGGATATCCGTCATTACCTTCAATTGATGAGCAGTACCGACTTCCGGCCTATCTCGTGT  
 CAGATGATTAGTCCGACGCAAAATGCCGGCAGACGCTGAAAGTCAGAGACAAAGGCATCTGCCCATCGTCCCCAAAAATAACAGCAACCGA  
 TAACGGACATACTGGGTATCAAAACATTGATTAGGTCAACTAATTGTCGCTTGCAGGAAAACACTAGGTTAATCTAATGCCACGCA  
 CGTAGTAATTGAAATAATTGAGGCTTACGGACATTGAAACGAATCGATGGGTGAAAGCGTCAAAGGATTGACACAGGTTGTTCA  
 GTTGAAGTGCCTGGAGTGTGTTAGGAAATAGATAGATGATTGATGCTAGCAGTTCTGGTTGATGCCCTGTTGATAGCAGCACTGAAA  
 GGAAGTCGCCAAACAAACCTAGGTTAGGGCGAGTATCGCTTCCGGCTAAAGATAAGCCCATCTCAGGCCGAGGCCAACAGATA

GATACACCCATAACAGCGCGTTCTGTACGTATCGCACGCACGCCACTGTAGAGCACCACGTCAGCAGCGCCAGATCCTCAAGCGCG  
 TTGATTGGTGCACCGTGTATCCGAGCGCCAGCCAGCCAGGCCAAGGCCAACAGCGCTACACTGCAAACACTTAGTGGGGGTTTGAGTGG  
 CGAACAAATCGATAATAATATAATGTAAAAAAACTAGTATAAAGAAACTAAGCTAAGATTATGTACGCTTCAGCTCAGGTACCTGAATA  
 GTACCTTCATCTCAATATTATATTATATAATATTACAGTAGTAAAGATTTTGTCCAATTATTAAAGTAAGCCTTTATTATAGAT  
 AAATGTTATGTTAAGATATTACATGTATTAAATGTATTAAAGATATTAAAGTATTGCAATCAGCTAAAGTAAGTTGTTCTTTAAGG  
 CGGTCGCACTTCTCACAGTGCCGGCGAACGCAAGCAGCTTACGACCATAACAAAGTAGATATGCCGGCACAAACCAATTTCATTAAAGACT  
 AAAAACCGACTCGGCAAATCACTCATCTTAGGGTTAGAGCGAACACATCGAGAACAGCTACAAAATTAACAAATTCG  
 GTCGTATCCAGCTGAAGTGTAAAAGATAGCATTAAATTGCTGGCAAAATGCAAGTGAACAAACGTTGAATTAGTGAAGAAGAAA  
 ACTACAATTCCGTGTGAAAACATTAGTTGGCAACAGACCAAGAGCAGAACACAAAA

To generate **Hy\_U5:34A eGFP SV40**, the following sequence was inserted between the **Pacl** and **Xhol** sites:

AACTTGAAACGGAAAACGAAAAGGGTCTCGATGGGCAAACGGCCAGAACCATTTGGAATTATAATGTGAAAGTAGAAAGTTATCGTACTTCTAT  
 TGCATATTAAAGGGTTTATGGTCATATACTACAGACATTCCGTTCCAGGTAGGCCACATGATATAATATGACTAAGATCCCTACCAAAACATAC  
 TTTAGGGTTGTTGCGTTTACAATCATTATGATTCCAACATGTTCAAGCTCGTTCAAATGATCGGGCTGATACTTTGTTGAAATTACTC  
 TGTTCTCTCAATTGTCGAATAATCTTCGCTTACTAAAGATTCCGTGAGAGGAACACTCTAAATGAGTCTAAATAATCTTTGTTGAGTGC  
 CCCGGCGACTTCGTTAGCTGGGCCATAAGGAATTCTGAATGTAAGTTTGCTTATTATAAGGCTGTTGCTGTCATACATGGACTACTCTACT  
 AAGGTTACCAAGAGTGTGAAATTCTGAATCGAGTTGTACCGTAGACGGCTAAATGTCGAATAATCAGCTTATTATAATGTGCAAGTC  
 AATTATATATTCTTAATGTCTATAGTGACTG

The following *Drosophila* expression plasmids were generated from the previously published **pUb 3xFLAG MCS** plasmid (Chen et al. (2012) *RNA*, 18, 2148-2156):

<b>pUb FLAG-Mts</b>	<b>(Addgene #195065)</b>
<b>pUB Flag-dMts H59Q</b>	<b>(Addgene #208402)</b>
<b>pUB Flag-dMts D85N</b>	<b>(Addgene #208403)</b>
<b>pUB Flag-dMts H241A</b>	<b>(Addgene #208404)</b>
<b>pUb FLAG-Pp2A-29B</b>	<b>(Addgene #195066)</b>
<b>pUb FLAG-Cka</b>	<b>(Addgene #195067)</b>
<b>pUb FLAG-tws</b>	<b>(Addgene #195068)</b>
<b>pUb FLAG-wdb</b>	<b>(Addgene #195069)</b>
<b>pUb FLAG-wrd</b>	<b>(Addgene #195070)</b>
<b>pUb FLAG-IntS6 AA 1-400</b>	<b>(Addgene #195071)</b>
<b>pUb FLAG-IntS6 AA 1-600</b>	<b>(Addgene #195072)</b>
<b>pUb FLAG-IntS6 AA 101-1284</b>	<b>(Addgene #195073)</b>
<b>pUb FLAG-IntS6 AA 1197-1284</b>	<b>(Addgene #195074)</b>
<b>pUb FLAG-IntS6 AA 101-1200</b>	<b>(Addgene #195075)</b>
<b>pUb FLAG-Human IntS6</b>	<b>(Addgene #198408)</b>
<b>pUb FLAG-Human IntS6L</b>	<b>(Addgene #198409)</b>
<b>pUb FLAG-Zebrafish IntS6</b>	<b>(Addgene #198410)</b>
<b>pUb FLAG-Zebrafish IntS6L</b>	<b>(Addgene #198411)</b>

The Ubi-p63e promoter (pUb; marked in gray) drives expression of an N-terminal 3x FLAG tag (marked in blue). Start codon is marked in green. The full sequence of **pUb 3xFLAG MCS** is as follows:

GACGAAAGGGCCTCGTGTACGCCATTTTATAGTTAATGTCATGATAATAATGGTTCTTAGCAGTCAGGGCACTTTGGGGAAATGTGCG  
 CGGAACCCCTATTGTTATTTCTAAATACATTCAAATATGTAACCGCTCATGAGACAATAACCCGATAAAATGCTCAATAATATTGAAAAGG  
 AAGAGTATGAGTATTCAACATTCGGTGCGCCCTTATCCCTTTGGCGCATTTCGCTTCTGTTTGCTCACCCAGAAACGCTGGTGAAG  
 TAAAAGATGCTGAAGATCAGTTGGGTGACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAAGATCCTGAGAGTTTCGCCCGAAGAACG  
 TTTCCAATGATGAGCACTTTAAAGTTCTGATGCGCGGTATTATCCGTATTGACGCCGGAAAGAGCAACTCGTCGCCGCATAACTAT  
 TCTCAGAATGACTGGTGAGTACTCACCAAGTCACAGAAAAGCATCTACGGATGGCATGACAGTAAGAGAATTATGCACTGCTGCCATAACCAGTA  
 GTGATAACACTCGGCCAACTTACTCTGACAACGATCGGAGGACCGAAGGAGCTAACGCTTTTGACACATGGGGATCATGTAACCTGCCT  
 TGATCCTGGAAACGGAGCTGAATGAAGGCCATACCAAACGACGAGCGTACACCCAGATGCCCTGAGCATGGCAACACGCTGCGCAAACACTATT  
 ACTGGCGAACTACTACTCTAGCTCCCGCAACAAATTAAATAGACTGGATGAGGGCGATAAAAGTTGAGGCCACTCTGCCCTGCCCG  
 CTGGCTGGTTATGCTGATAAAATCTGGAGCCGGTGGCTGGTATCATGGCAGCAGTGGGCCAGATGTTAAGGCCCTCCCGTATCGT  
 AGTTATCTACACGACGGGAGTCAGCAACTATGGATGAGCAAATAGACAGATCGCTGAGTAGGTGCTCACTGATTAACGATTGGTAATGTC  
 GACCAAGTTACTCATATATACTTTAGATTAAACTTCATTTAAATTTAAAGGATCTAGGTGAAGATCCTTTGATAATCTCATGACCA  
 AAATCCCTAACGTGAGTTCTGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTCTGCGCGTAATCTG  
 CTGCTGCAAACAAAAACACCAGCTACCAGCGTGGTTGTTGCCGGATCAAGAGCTACCAACTCTTCCGAAGGTAACGGCTCAGCAGA  
 GCGCAGATACCAAATACTGTTCTTAGTGTAGCCTAGTTAGGCCACCACTTCAAGAACCTCTGAGCACGCCCTACACCGCTCTGCTAATCC  
 TGTACCGTAGCTGCCAGCGCAAGTGTCTACCGGGTGGACTAACAGGATGTTACCGGATAAGGCCAGCGCTGGCTGAAC  
 GGGGGGTTCTGACACAGCCAGCTGGAGCGAAGCAGTACACCGAAGTGGAGATACCTACAGCGTGAAGCTGAGATAAGCGCCACGCTCCGAA  
 GGGAGAAAGCGGAGCAGGTATCCGGTAAGCGCAGGGTGGCAAGAGGAGAGCGCAGCAGGGAGCTTCCAGGGGAAACGCTGGTATCTTAGTC  
 CTGCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTGATGCTGTCAGGGGGGGAGCCTATGAAAACGCCAGCAACGCCCTTTT  
 ACGGTTCTGGCCTTTGCTGCCCTTGCTCACATGTTCTCTCGTTATCCCTGATTCTGTTGAGATAACCGTATTACGCCCTTGAGTGAGCT  
 GATACCGCTGCCAGCCGAACGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGGCCAATACGCAAACCGCTCTCCCGCGCTT  
 GCGGATTCAATTATGCACTGGCAGCAGGTTCCGACTGGAAACGGGGCAGTGAAGCAGCAACGAAATTATGAGTTAGCTCACTCATTAGC  
 ACCCAGGTTTACACTTTGCTCCGGCTATGTTGAGCTGTTGAGGATAACACAGGAAACAGCTATGACCATGATTAC  
 GCCAAAGCTTGTGGCGAACGCGACAGAGATTCCATGTTGAGGTTGAGCTGTTGAGGATAACACAGGAAACAGCTATGACCATGATTAC  
 CGCGTGGGGTCTCTCAGGGCTTGTGAATTAGCGCGCAGATGCCGATGGCGCAGGCCGAGCTTCACTTGCACCGTACGGCTT  
 CTGTTCTCGGTTCAAAATCTCAGCTCATTTCGTTGCAATCAGTACTGTTCAAAATCGAAAATCGCGAACCGTAGTGTGAC  
 CGTGGGGGCTCGCAAATTTAGGTTATGGCCACACAGGGGAAAGCACAGTGGATTATGTTAATATTATAATATCGAGGTT  
 TTCATTACTATCCAGATGTAAGCCACTAAAGCATTAAACATTATTGCGGAAAGAGTATAAACAAATTTCACTTAAAGGATTAAGAAAA  
 GCTAGTTGTGAATTATGCGCAGCGTGCAGTAGCTCCATTAAACACATAAGTTGAAAATATACATAAAATAGCAGT  
 CGTGGCGAACGCTCAACACATCACACTTAAACACCCCTTACACAGAAATTACTTTAAATTCAGTCAACGCTGAGTTCAAATT  
 ATAGCCGGTAGAGAACAGCTGCTTCAAAAGCAAACAAAGGGCTTCAAAACACCAATCTTCAACAGCTGGACTTTGTAAGT  
 TTAGATCAAAGGTGGCATTGCAATTGATGTTGAGGTTAGGTGCTTAGGTTAAGGTTACCTCATTAGCCGCTAACGTACAGGAGGTT  
 CTCATTGAAATTGCTTAAATATTGTTGACTGTGGTAGTTAACGAAAACACAAAAAAAGTAGTACACAGAAATCATAAA  
 AAATTAAATACAAGGTATTGCTACGTTACAAACATTCCGGACAATTCTCTGACTAAAGTGTACGAACACTACGGTATTGTTAGT  
 GATTTCACCGACACCGAAGGTATATAACAGCGTTCGCAACGGTCCTCAAAACCAATTGACATTGCACTGAGCAAGTACAAGTAGAAAGTA  
 AACCGCAATCGGAAAATTTACTTAATTGTTGTTGAGTAAAGTACAATTAAAGAACATTCTCGAAAGTCACAAGAACGTAAGTTTAACT  
 CGCTGTTACCAATTAGTAATAGAGCAACAGACGTTGAGTAACTTCAAGAAAACCTGATTTCAGGGCTTGGCCATTTCATTCAA  
 CGCTCTACGTAATTACAAAATAAGAAATTGGCAGCCGCATCTGTTTCCAAATGAAATTGGCATCAAACGAAACAAATCTATAAAACATT  
 CGCTGTTGATTTCGCAAAGATTATTGCAAAATTGTAATTGCACTGACGATTGAGGAAATTGAGAATTCGAGAACCGACTCGATCGACGATT  
 TGTGCTGATGTTATTAGTTAGTTAGGTTATTGAGTATTGCAACGAAATTCCACTTATTGAAATAGTGTACGCTGTTGATATACAA  
 AGCCACGCTGAAAATTCTTATTGCTTAGTTGAGCTGACGTCACCATATACACACAAAATAATGTTGATGCACTGCTGTTGATATACAA  
 TGCACACACTCGAACACGAAAACGATGACGAAGCAACGGAACAAAGGTTCTCAACTACCCCTGTTCCCTGTTCTCGCTTCTGTTCCAA  
 ATTGCTAGAGGGTTAATAGGGTTCTCAACAAAGTGGCTGCTGATAATAAGTTCCCATTTCCTCCAGGCCAGGAAGTTAGTTCAATAGT  
**TTGTAATTCAACGAAACTATTGATTGCTACTAATTCCACACTCTATTCTCCCGCAGAATA** TCCAACACTGCAAGGTCAGCTAGCT  
 AGAGGAAGCTT **ATG**ACTACAAAGGACATGACGGTGATAAAAGATCATGATATCGATTACAAGGATGACGATGACAAGGATCCACTAGTCAGTG  
 TGGTGGATTCTGCGAGTATCCAGCACAGTGGCCGCCCTCGAGTCTAGAGGGCCCGGGTCAAGGTTAGCTTACCTTAACCCCTCTCCCGTC  
 TCGATTCTACCGTACCGGTATCATCACCATCACCATGAGTTATGACTAAATCTTAGTTGATGTTGATGTTAATACAATATGTTATGT  
 TAAATATGTTTAATAAATTTAAATTCACATTGTAACACATTGCTCATTACACACTCTTCAAGCGCTGGGATCGATG  
 CTCACTCAAAGCGTAATACGGTTATCCACAGAACGAGGGATAACGCAAGGAAAGAACATGTGAGCCATATGGGCCATGTGAGCCATATGGTGCA  
 CTCTCAGTACAATCTGCTGATGCCGATAGTTAACCGAGCCCCGCAACACCGCTGACGCCCTGACGGCTGTCTGCTCCCGC  
 ATCCGCTTACAGACAAGCTGTGACCGTCTCGGGAGCTGCATGTCAGAGGTTTCCACCGTACACCAGAAACGCGC  
 acccaggcgccgggagcctatgttacgcgaagaacaccggattttcccttaa

To generate pUb FLAG-Mts, the following sequence was inserted between the Spel and XbaI sites:

Cgaggataaagcaacaacaaaagatctgtatcaatggattgaggcattgaaatggcgttgcacatcacttgcacccctctgcgcac  
 aaggccaaggagattctctcaaggagtccatgtcaggagtagggccatggccgttgcacccacttgcgttgcacccatggccgttgcac  
 tggagcttcggataggccgttgcacccacttgcgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcac  
 cttctgggtggccctgaaagggtcgatccatcgccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcac  
 gacgagtgccgtccgtcaagtatggcaatgcacccatggccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcac  
 tctctgcctgcacggggccctcgatccctcgatccatggccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcac  
 gtgcgtatctgtctggccatggccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcac  
 ttacaacacacaaacccgttgcacactgggtgtccgtccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcac  
 tctcgtccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcacccatggccgttgcac  
 acccaggcgccgggagcctatgttacgcgaagaacaccggattttcccttaa

To generate pUb FLAG-Mts H59Q, the following sequence was inserted between the Spel and XbaI sites (mutation from WT sequence marked in yellow):

Cgaggataaagcaacaacaaaagatcttgcattgtggattgagcagttgaacgaatgcacatcagttgcacagagacacaagttcgacccttcgac  
aaggccaaggagattctctcaaggagtccaatgtgcaggaggtaaaatgcccggacactacccatgtttcatgggcactacgtggaccgtggataactccctggagacccgtgac  
tggagcttccggataggccgcaagtctccggacaccaactaccatgtttcatgggcactacgtggaccgtggataactccctggagacccgtgac  
cctctggggccctgaagggtcgatccctgcgaggcgcatcaccatctgcgcggtaaccacgaggatctggactacttgcactacgtggaccgtggataactccctggagacccgtgac  
gacgagtgcctgcgcaagtatggcaatgccaacgttggaaagtacttcacggatctggactacttgcactacgtggaccgtggataactccctggagacccgtgac  
tctctgcctgcacggaggcctgatccctgcgacatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgac  
gtgcgatctgtctggatccctgcgacatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgac  
ttaacaacacaaaacggcctgacactgggtcgcgcgcccattacatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgac  
tctcagcgcacactattgttgcacccgtggcaaccacggcttcatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgac  
accaggcgcggggagcctatgttacgcgaagaacacccgattttctttaa

**To generate pUb FLAG-Mts D85N,** the following sequence was inserted between the Spel and XbaI sites (mutation from WT sequence marked in yellow):

Cgaggataaagcaacaacaaaagatcttgcattgtggattgagcagttgaacgaatgcacatcagttgcacagagacacaagttcgacccttcgac  
aaggccaaggagattctctcaaggagtccaatgtgcaggaggtaaaatgcccggacactacccatgtttcatgggcactacgtggaccgtggataactccctggagacccgtgac  
tggagcttccggataggccgcaagtctccggacaccaactaccatgtttcatgggcactacgtggaccgtggataactccctggagacccgtgac  
cctctggggccctgaagggtcgatccctgcgaggcgcatcaccatctgcgcggtaaccacgaggatctggactacttgcactacgtggaccgtggataactccctggagacccgtgac  
gacgagtgcctgcgcaagtatggcaatgccaacgttggaaagtacttcacggatctggactacttgcactacgtggaccgtggataactccctggagacccgtgac  
tctctgcctgcacggaggcctgatccctgcgacatgcacagggggtggctggggatctgcctctggccggtaacccttggccaggatatttcgaaacc  
ttaacaacacaaaacggcctgacactgggtcgcgcgcccattacatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgac  
tctcagcgcacactattgttgcacccgtggcaaccacggcttcatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgac  
accaggcgcggggagcctatgttacgcgaagaacacccgattttctttaa

**To generate pUb FLAG-Mts H241A,** the following sequence was inserted between the Spel and XbaI sites (mutation from WT sequence marked in yellow):

Cgaggataaagcaacaacaaaagatcttgcattgtggattgagcagttgaacgaatgcacatcagttgcacagagacacaagttcgacccttcgac  
aaggccaaggagattctctcaaggagtccaatgtgcaggaggtaaaatgcccggacactacccatgtttcatgggcactacgtggaccgtggataactccctggagacccgtgac  
tggagcttccggataggccgcaagtctccggacaccaactaccatgtttcatgggcactacgtggaccgtggataactccctggagacccgtgac  
cctctggggccctgaagggtcgatccctgcgaggcgcatcaccatctgcgcggtaaccacgaggatctggactacttgcactacgtggaccgtggataactccctggagacccgtgac  
gacgagtgcctgcgcaagtatggcaatgccaacgttggaaagtacttcacggatctggactacttgcactacgtggaccgtggataactccctggagacccgtgac  
tctctgcctgcacggaggcctgatccctgcgacatgcacagggggtggctggggatctgcctctggccggtaacccttggccaggatatttcgaaacc  
ttaacaacacaaaacggcctgacactgggtcgcgcgcccattacatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgac  
tctcagcgcacactattgttgcacccgtggcaaccacggcttcatgtggatccctgcgacatgtggatccctgcgacatgtggatccctgcgac  
accaggcgcggggagcctatgttacgcgaagaacacccgattttctttaa

**To generate pUb FLAG-Pp2A-29B,** the following sequence was inserted between the BamHI and XbaI sites:

Agcagcaagcacaatcggtgcacgattcactatccattgcggttctaattcgatgaactaaaaacgaggacgttcagttcggttaactcc  
atcaaaactgttccaccattgcacgcgttggcggaggagcgcacacggcgttgattcccttcacccgagaccatatacgatggagac  
aggactgtggccctggccgaccaactggcaacttactgttgcggcggatcttcacccatgtttcccttcacccgatgttgcggccatgttgc  
ggccaccgttagggaaacccgtggcgacacaactggcgtggatcttcacccatgtttcccttcacccgatgttgcggccatgttgc  
gtggccacactgtggccggatcttcacccatgtttcccttcacccgatgttgcggccatgttgcggccatgttgcggccatgttgc  
cagtggccggatcttcacccatgtttcccttcacccatgtttcccttcacccgatgttgcggccatgttgcggccatgttgcggccatgttgc  
caagggtcgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
caagggtcgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
gaggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
gggtcggttacatgttgcggccatgttgcggccatgttgcggccatgttgcggccatgttgcggccatgttgcggccatgttgcggccatgttgc  
gttcaaggatggccggggggatcttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
ctttagtccatgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
tgcggccatgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
cttggggccctatgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
tgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
cttggggccctatgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
cgttccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
acgatcacgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
aatgattctgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
accaccaactgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
ccttccctggggccggccatgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgcggccatgttgc  
ggccattggccggatgttgcggccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgtttcccttcacccatgttgc

**To generate pUb FLAG-Cka,** the following sequence was inserted between the Spel and NotI sites:

To generate **pUb FLAG-tws**, the following sequence was inserted between the *SpeI* and *NotI* sites:

Cgcggtaatggagaggcgtctgggtcttgcacagattaaaggcgccctagacgatgtcagggatgcagacatcatatcctgcgttggaaatc  
aatcacatggcgagctgtggccactgggacaaggcggtcgctgtcatcttcagctgtatctgcctcaaaagccgccaatccgagacgcg  
gcaatacatgtctatcgacattcaatcgacgagccgaattcgactacctaagttccggagattgaggagaagataaacaagatccgg  
gtctggcaaaaagaatcccggtcacttgcgtccacaagacagaatgtcaatttggaaaggctagtgcagctgtcagaaatgcgttgc  
tacaacacaaaggaggagaacggactgtccggatccacaagatgtaaacggccctccggatccatccgtgaagcagataccactgttgg  
ccctcccccacggcgacacttcgccaatgcacacacctatcatcaattcaataagcgttaatccgtatccggacacttgcgtccgg  
gcggatcaatctgtggacttggagggtgtcaaccagagctacaacattgtcgacatcaagccaccaatggaggagctaacagggtgatc  
gcggccgaattccatccgaccggatgtcaatgtttgtctactcgagctctaagggcacaatacgttatgttatgcgttgcgcgc  
accggcacagcaaacagttcgaggagccgagaatccaacgaatcgcagctttcagcgaaaataatcagctccatcagcgatgt  
gaagctaaacacttcccgctgtcgatcgaactcgaaaggatgtgcgttagggcgtccaggagcatcatcaaaacggatcg  
taacgcacggaaagtggacttgcggcaaacgaaagaaggatgagatcagctgttgactgttagatttcaacaagaagatc  
ctgcgtcccgatccatcgccgtggctgcgaccaataaccttcatatttcaggataatttttag

To generate **pUb FLAG-wdb**, the following sequence was inserted between the *Sph*I and *Not*I sites:

To generate **pUb FLAG-wrd**, the following sequence was inserted between the Spel and NotI sites:

Gcataacgaggcgtagatccaacaataaaaagcagtacgtcgccagcgcacggcaacagcagcatcagaacaaacgacaacaacagcagcatcttca  
gttgtggaaaccacaaacaatcgccgcccacagcctcagccgagaactaaaaacgaaacaacgcccacaaacaatagaacacacaaggcg  
gcagcatcagcagtagtagccgcaacaatatagtcataccggcatcgccactaaccgttatcaaagagagaacttaacttaatgacaaacaaac  
agcagcagcagtagccgccccccatgttagaaggagtagtcccgacataaactgtttccgacataatccggccgactccatttgacgt  
ttggctaacaaactcaaggacaacacacccaccgtacgtacgtccggccacccggatcagaacgtctgaatcaccggcaccggatcg  
gcaaggagaagcgccagaccagcgcgggtacaatgcctcaagaactgcgaactgcgcgttccgttacacgagaagaccgcgcgc  
acgtgaggagctgttcatacagaagatcccgacatgtgcacactgttcgacttctccgagccgcgtcagccgcactcaagttcaaggagg  
gcggctctgcacaaatggtcatttccaccaacccatggcgtataaccgaagtatttacccggaggcgatcaatatgtttctgtcaacc  
tttccgaacttgcgcgcattgtccaaacccgaatgttgcgcatttgcgcggaggatgagccacgttgcgttgcgcattctgt  
actcggttacgagctgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
ttggattttatcgatttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
tttagaaagcagatcaaaatgtctttacagattttatgaaacggagcatataatgcataatgcgtatccgcgttgcgttgcgttgc  
aatggcttgcgtcgcttaaggaggaggcataaaacaaatttacttaaggatgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
ccgcagctcacctattgtgtggcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
acagtccaaaggaggatgttttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
ccaaatggcaactgtcgctcttcgcctatttcgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
aactcgccgttatattaccattatgttcccgccgtcataatgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
tcaagctgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
ggaggatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
ccggccctgtcgatcaatatagtgagaacagcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
aacaggcgcaccaggagccccgagaggtgagacaggacttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
aacaggcgcaccaggagccccgagaggtgagacaggacttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc  
aacactaaacaactactaa

To generate **pUb FLAG-IntS6 AA 1-400**, the following sequence was inserted between the SpeI and XbaI sites:

cacaatcatactcttcgtggacacctcgccatgtgcagaaggcgtatgtgaatgggtacagaaaacgtatctggacattgcaaggga  
gcccgtggagacgtttctaagtatcgccagcgtacgccaggattgcggagatcgctacatgtgcacattcgaggagccaccggcaaaacgtga  
aaggatggaaaggagaaccatgcgcacccatgcataacaacgtcgaaagactcgacgtacggccctcacccatgtggatcgatcgccatgtc  
gttcgattgtttaaatctgtatcgatcgatcgccatgcatacgatcgccggcagggtggcccggttttatctggagccatcggtatcattgtg  
attacggcgcggcgcgtcattcgatccggaaatgtgttcattcgagatcatactgcgcataaataccggcgcacaaaatgtcaagg  
agccatctcgctggatcagcgcttgcgttgcgcatacgccggcaacaagattgacgagcgactggatgcaagggtgcgcacatgc  
ttctcccatcgaacggatgtgcgcaggattaccggccgcgttcatatcgatgcgcactacgtactcaaccagtgcattgaaagtcttgc  
aagggttcagccgtgtgtgtcgacttcgcagccatgtgcgcacaggacttccgcgtccgcgcacgcacatcat  
ccggcatggaaacatccatcaacgtccgcggccgcgtcccgatattgttccatccgttaaaaggatgtatcgatcgccatgc  
cgacggaaaacttccatccgttccgcgtccgcgtcccgatattgttccatccgttaaaaggatgtatcgatcgccatgc  
ctgacggatgttccatccgttccgcgtccgcgtcccgatattgttccatccgttaaaaggatgtatcgatcgccatgc  
ttctcaacaacgcgcgacatgttccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtcc  
caactttcgccgttccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtcc

To generate **pUb FLAG-IntS6 AA 1-600**, the following sequence was inserted between the SpeI and XbaI sites:

To generate **pUb FLAG-IntS6 AA 101-1284**, the following sequence was inserted between the Spel and Xbal sites:







aggacagagaagacgcgtcgccaccggacggagaggagatactacgtggaaaattgtctgtacagatcccgtgaccacacgcacaaactgt  
 tgaggagctctccggccaggagggggatggaaagtcaatgaaaggagacactccacggcacttgcgtcatgatccccctggaggg  
 agtaacgcagacgtcgccactcgggatatacaaaggatggcccaagctggcccaattatgaggctgtggaggaggtgaaagg  
 ctgtatcgtccagaggatcttcatgcataaaaggccaggtttaaaaggctatgtccatgcacatgtccacaggat  
 ggagattgaagacagacatgtacccgcacaaatcaaatgtccacaggatagt

The following *Drosophila* expression plasmids were obtained from Eric Wagner's group (University of Rochester) and were generated from the previously published **pUb 3xFLAG MCS** plasmid (Chen et al. (2012) RNA, 18, 2148-2156):

**pUb FLAG-IntS1**

**pUb FLAG-IntS2**

**pUb FLAG-IntS3**

**pUb FLAG-IntS4**

**pUb FLAG-IntS5**

**pUb FLAG-IntS6**

**pUb FLAG-IntS7**

**pUb FLAG-IntS8**

**pUb FLAG-IntS9**

**pUb FLAG-IntS10**

**pUb FLAG-IntS11**

**pUb FLAG-IntS12**

**pUb FLAG-IntS13**

**pUb FLAG-IntS14**

To generate **pUb FLAG-IntS1**, the following sequence was inserted between the **SpeI** and **XbaI** sites:

cgatcgccggaaaggaaacgcgcgtccaaaccgatcgcagaagaagggtccgtggccgagaactcttgcgtcgcaagagcgtacgcgtactcc  
 aagtcaagatctccatcaaaggcatgtcctcgccaccgcacgcgtgaaggctccaccgcactggccagctcgagcaagcggttcccgccca  
 acctaaggacgcggccggccggacatgtcctccggcagttagtcgtcgagacgtggagcaattcccgctcgactgcgtatggataccgttgt  
 cgagaccatatacgccgttggagcagaacgcacgcgaaactgtgggtcggttataaaggcagacaacgcgagctcccgcc  
 tccaagggtggacaacattgccttctggccctgtactacgtggccaaagggtcagccgactatatttgcacggacatagtggctgcgcgttgt  
 cttctctgcggccggaggccaaacgtgaaaatgaggtaacacaccaacctgcacattctgtttgcacatctgtgaccgggggttcatggagatttc  
 ccagtggccggagggtgtccgcattacatagacatgcgttaacgcgtactggccggacaatgaactgtgtcccgctgtggaaac  
 atatgcgcagccattaaaacccgcacccacatcgttccctccgtgggatgtgagctgtccctgccttctggtaacgcacacagagac  
 tgaccgtggacacgttccgggataatccacacagagctggatcagaacttgcataccgcacatccgcacggataactacacc  
 cactaaatcccgatccgcgtgtggcagaacatgtcagcgatgccattcgcatcgcataacaaggcgcacgcgcaggataactacacc  
 agaaacttccctaaagggttccctgtgcaccacttctggcattgcagaggatcgtgtggcatacttccgcctggagctgtggatccacaacgg  
 ttggtaagtttgcggcccgatgtctctgtacatctgtccatcaacatcaaggacgcacacacaggacaacacgaaactgtcttggtgt  
 gggctcaagacgaagccctaataaccactacatgtcctgtccatgcagccgaaatcttgcgttgcacatgc









ggtcgagggtgcctgcagcaacttagcaagcagaagaaacactccgtcgctggagtccgttaattaagcctcacctcacagctaattacgcggcc  
ccaccgcgatccgcgaagagactgcccacgcggccacgatactgtgaccctcaatcagattgtaaagccccagcgcgatccctcccgca  
gcttcctgttgcacatctccaaatggccggacttccaggtgacgctggaaacattcgttggtggacgagaatggcattactgtgcactggcccaa  
atccctcgatgggtgcgggtgctggaggatcctagcaagcaggcgctccagcggcggacttcccaggcagtgggacagacgaggagtttaa

To generate pUb FLAG-IntS8, the following sequence was inserted between the BamHI and BstBI sites:

agacgatcccttaagccaaagccagttccactggctgccaaacggtaactgtgggtcgagttccgtcgatccgcacaagattacgcggcatctg  
cagcgcggcgcaccccgagccagcgcgcattggagactgtatcgatcgatccgtcgatccgtcgatccgcacaacccgcggcggactggggacgcgc  
gtgatttgcagaatctaaaccacgcggcatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgt  
cgagttggcatccccatagccccaaacggccggagaacggccgcgtcaatggggcacttccaggtgacgctggaaacattcgttggtggac  
gccaattggctgaaatggggatctggatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgt  
gctgtcccttaagtattccctgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc  
ccgcattgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
tattactggccacaccactcaaccatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
cgctgcgtatgtatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
ccatattacccgtacccaaacactatgtttggccggaaacggccgcgcgttaactaaccacccaggattggagagccaaactgcactac  
tattactggccacaccactcaaccatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
cgctactaacaactacacagacatccgtatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
cattagggttccatctccacccgcatactcaaggagacagttcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
ggcagcatcttggcgcgttgagttcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
gtacttcaaaaacaggccgcactcaaggacttccatcgactgtttcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
atttggactgttctcgatccaggatattcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
gactggatgttgcactccaaatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
tagtgaatattatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
gcccgttagctttcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
agtgtactgaatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
cactgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
aatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
cccttgcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
taacatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
tttggactcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
accaggatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
tttcatggctgaagatggcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
ttatgtacatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
acgggtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
actattccgtatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
gatgggaaccctcgactcaaggccaaataacgaagaaattaacgcgactcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
caatacctgtctaatctagaggcccgccg

To generate pUb FLAG-IntS9, the following sequence was inserted between the BamHI and XbaI sites:

aatgcgattgtattgtctcagcggggacctggccaaaggccatgttatattcattacccatgttgcggattatgtggactgcggctcacggag  
cagacgtccgtatccctccgtccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgtt  
tggatggcgaggatgttgcggacttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgtt  
ggatgttgcggacttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgttccgtt  
acgctcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
atccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
ttccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
accaggatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
tttcatggctgaagatggcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
ttatgtacatcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
acgggtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
actattccgtatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
gatgggaaccctcgactcaaggccaaataacgaagaaattaacgcgactcgatccgtcgatccgtcgatccgtcgatccgtcgatccgtcgatcc  
caatacctgtctaatctagaggcccgccg







To generate **pMtnA FLAG-IntS7 puro**, the following sequence was inserted between the SphI and XbaI sites:

To generate **pMtnA FLAG-IntS12 puro**, the following sequence was inserted between the SpeI and EcoRI sites:

To generate **pMtnA FLAG-IntS13 puro**, the following sequence was inserted between the SpeI and Xhol sites:

To generate **pMtnA FLAG-IntS14 puro**, the following sequence was inserted between the SpeI and Xhol sites:

ccccacaccttaatagcgcgtggatgcatcgctatcgatgctgcgcggcggccggaaagaaaatgagcacacccatccaggctgctggccaccaaggccatc  
cagcatctgctggacaatctcacggcggccggcaattggagcacgtgctgtctccattcgacgacggcgaactgaaggtgacttcacgc  
gggactacgaccaggctgcccggcgtcaagaaggtgagccgtggacaaggcgtgcgtatgagcatgtctccaaaggcgtgttgcataatgtc  
ggcgtggggcaacccagaacatctgcagggtgttcttacggactcggcgtccggcttggcaaacactcgatcaggcgttccctggaggcctac  
ggcggaaaaggagtggaggcggagtccggcttccgtaaagactctggcaactaacactgtttcatctgcgttgcgcgtcgcaggcattact  
tcaccagggattggcgttaccagcagctgtggacaaggctcactgaaggccagctttcatgacaaagccccaagagcagcgtacgc  
cgagggaaatccaaatccaaacccgaaccccgacccacaaaagcgaactggccgcactacggcttcgaactaattgaacgtctcgagggc  
tacaagaggtcgaggtgaccctgaaatggcggcagctacttcccatggaggcatctgttgcgttgcgtccgtatgagcaaaagt  
ccccatattcgccgagaacccaccattcgccacacagaccagaagattgggtgtgtgttccctgtggacattggatcaccggcc  
cctgagttcgacactgggtgtcccaaagtggagccggagaagagcggcagactgggtgtggaaatctgagcgcagccaaagccacccaa  
ctcaatctgatagataccaggcaatccaaatccaaactcgagctggagaagctggagccgatataaaggatgttacgcaaggattcaaa  
ggaccccgcccccacagacacggcaaaaaaaaacttgcgttgcgtcccttcacggccctcaaaa  
ggagaacatggcagctgtgggtgacaagtggtagccgtatcatatgcattcaccgcacagaagaagtcacccatgtgt  
atacttccaccggcacaacgtgattccctggcttagagacccatggatcgtgggttcccgaggatttgcgtccggagagacggcc  
cggtgcgtccgaccgcagatcttgcagagcagttgtgggtgtggatcgcacaggctagctgcagttccgcacccagaaggta  
ctgcacacgc  
caagaagatggccgacaagacgcacacttctcaaaaggactcaatcgatccgcgcgtccctggcgtggattctgcagttgt  
tgcgtggccatgtcggtggaaaggactcgcccatctcgcgttgcacggccgcacacgcactgcacccctgcagctgc  
cagtcgcacccgcgtccacggatcgttgcgtccctggatcgcgtccggc  
tgcgttgc  
tgcatacatgtattgagccggcc

The following protein expression plasmids used for antibody production were generated from pRSFDuet-1 (Novagen):

**pRSFDuet-1 IntS6 AA 1035-1284 (Addgene #196904)**

**pRSFDuet-1 IntS8 AA 1-308** (Addgene #196905)

pRSFDuet-1 IntS11 AA 300-597 (Addgene #199329)

To generate **pRSFDuet-1 IntS6 AA 1035-1284**, the following sequence was inserted between the BamHI and HindIII sites:

To generate **pRSFDuet-1 IntS8 AA 1-308**, the following sequence was inserted between the BamHI and HindIII sites:

atggacgatcctttaagccaaaggccagttccactggctccgaaacgggtactgtggttcgagttctgtcgatccgcacaagattacgcagcatgcagcgcgcgcatcccggccagccagcgccatggagctgtatcgtcagttcatcagcatgcacgcgcgaaacacggcgcaggagtcaagtggggacgccttgcagtgatttcgagaatctaaccagacgcacatcgaaatccggaccatccccggctgggttgtggccctgcgcgacaacgcacactgccttgcgcgagttggcatgcggccatagccacaaaggccgcggagaaggcctgcaattgaaccgaagcagggtggactgaaaatcctcgagctgaaggtagccactggctgaatggatctggatgcattggagaagaacactgcgtgtgattatgcagctggcctgtgcgcgatctgtgcaccataagctacggctgcctttaagtattccctggccaaatgatttgcgcggagaatttcgcgtgtggaaacgcggaaagaggcacaagattgcgccttgcattacacccgcatgtgcgaatgcaactgataaaaggcgcggcgtaaaagcgcacgcgtccctaaaacacacatgttaccaactgtggaccagctccacgcacatgtgcacaccatcatcgatgttgcgttgcaccaactgcgtgcctgcacaaaaaccttctatatttcatactacacagacttgcgtgcgttatgcaccaataacacccaggattgagagccaaactgcactacgagctqcggccatattacatgtgcacccaaacatgtttqgcccggaaagcggctaa

To generate pRSFDuet-1 IntS11 AA 300-597, the following sequence was inserted between the Sall and HindIII sites:

Ttcacttcaaggcacatcaagcccttgacaaggcctacatcgacaatccggcgcgtggtagtgttgcaccgcaggcatgctgcacgcagggccccccctgcagatcttcaagaagtggcgcgaatgagaacaacatgggtattatgcccgcgtactgtgtgcagggaaacctgtggcaacaagattctcggtggcccaagaagggtggagttcgagaaccgcgcagggtggcaggtcaaatggccgtggagttacatgagttctcgcccatgcagatgc当地ggcattatgcagtaatcagaactgtgagccaaaaatgtcatgctcgtccatggtaagcgggaaagatgaaagtctcgctcgagatcaaaagacgaaatttcaacctgtggaaacctatcgccggcaacggcggaaacctgtgtgatttctacgcctgtgaagataccaggatgtcatcgatccgttactgttaaaagcagggcccgctctacaaccccgccgatcccaagaggcggcgcataatccatggcgtccgttaatgagaacaatcgaataatgtctgcaaaaacctaaaccgtgcctcaaaggaaatcgaattaatcgcacatgttacgtccaaagtgtcaaatggacgactccggccgggtatctgcacaaggcgaagactgtgactgtgcagaagaaaagctagccggctgcagggtgcacgtcaggagaacggccatagccatcgatgtccgttgagggtgaagggtggaggaggacgagaaggatcccaagcagaagaacattttgtatatcgtggaccaaccaggacgaggacattggagccatcataccgaatgtgtcgagaatatgtgtctcg

The following protein expression plasmid used for antibody production were generated from pGEX-6P-1 (Amersham):

pGEX-6P-1 IntS12 (Addgene #210521)

To generate **pGEX-6P-1 IntS12**, the following sequence was inserted between the EcoRI and NotI sites:

ggccaaatatacgccgcgtctgagctgcccagaagaatggatcccggtgtcaagaaggccatcaagtgtcactcgagtaatcccacctcg  
cgccgaactcgctgtcttggacgggtttgaaggcgcgtttggccggagaaaagttgaccaacaatgcgcacggatgtggagga  
tgaagcgaactttccggacgtgtccacgcccccatcaatgcgcacggatcatcaatttgcaccaactcaccggacaaggagccc  
tccgattcggtggacacggatcggtgtacggcgtctaagcgcgttggaaatgttaacaccgggtacacccggcactttgggcacctaatt  
gtcgctgtcgccggatgttttcacggccacaatcggtgttgcgttgcacatcggtgtccaaatgcgttgcgttgcacatcggtgtccaaacggcc  
cataccaaggaggaggccgcgtatggccacggacaaatgcgcgtgcacatcggtgtccaaacagccaaacggcgcggggaggacacatcct  
gcagcgcgcgttaacgcacactgtcttagccgacgacccatgcgttgcacagccaaatgcgtcggtcgctgcgtcatcgaaact  
cttccaaactcctcgtaacccttctaccggcccgagcccgacgttccacgacgacgatggccacaatgcgtcggtcgctgcgtcatcgaaact  
ttcgctgtcgccaaatcgacaaaggagaacgcgttcttaagtccacggccgtttcttgcgcacatcgccgaaatggagaagcacaacagc  
agtggcacctcatcgacgcagcggccgtccacggccaaatccacgttccaaagcgttccacgcacatcacaaggcggcagcagcagcaagc  
gatccaaacggat