

# Transcription directionality is licensed by Integrator at active human promoters

**Fan Lai**

fanlai@ynu.edu.cn

Yunnan University <https://orcid.org/0000-0002-0568-2616>

**Yunkun Dang**

Yunnan University

**Jiao Yang**

Yunnan University

**Jiangyang Li**

Yunnan University

**Langxi Miao**

Yunnan University

**Xu Gao**

Yunnan University

**Wenhao Sun**

Yunnan University

**Linghu Shuo**

Yunnan University

**Guiping Ren**

Yunnan University

**Bangya Peng**

Yunnan University

**Shunkai Chen**

Yunnan University

**Zhongqi Liu**

Yunnan University

**Bo Wang**

Yunnan University

**Ao Dong**

Yunnan University

**Duo Huang**

Yunnan University

**Jinrong Yuan**

Yunnan University

---

**Article**

**Keywords:**

**Posted Date:** July 26th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-3181116/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** There is **NO** Competing Interest.

---

**Version of Record:** A version of this preprint was published at Nature Structural & Molecular Biology on April 22nd, 2024. See the published version at <https://doi.org/10.1038/s41594-024-01272-z>.

# Abstract

A universal characteristic of eukaryotic transcription is that the promoter recruits RNAPII to produce both precursor mRNAs (pre-mRNAs) and short unstable promoter upstream transcripts (PROMPTs) toward the opposite direction. However, how the transcription machinery senses the correct direction to produce pre-mRNAs is largely unknown. Here, through multiple acute auxin-inducible degradation (AID) systems, we show that rapid depletion of a RNAPII-binding protein complex, Integrator, results in robust PROMPTs accumulation throughout the genome. Interestingly, the accumulation of PROMPTs is compensated by the reduction of pre-mRNA transcripts in actively transcribed genes. Consistently, Integrator depletion alters the distribution of polymerase between the sense and antisense direction, which is marked by an increased RNAPII-CTD Tyr1 phosphorylation (Tyr1P) level at PROMPT regions and a refrained Ser2 phosphorylation (Ser2P) level at transcription start sites (TSSs). Mechanistically, the endonuclease activity of Integrator is critical to suppress PROMPTs production in a sequence-independent manner. During this step, the endonuclease activity can be inhibited by the U1 signal on nascent antisense transcripts through the recognition of the U1 snRNA-Integrator which relies on the U1-Integrator axis to govern the direction of gene transcription.

## Introduction

Transcription initiation of eukaryotes is a highly orchestrated process that requires hundreds of protein factors to establish the proper start of gene expression. Active promoters typically harbor two distinct pre-initiation complexes (PICs) assembled with opposite orientations for divergent transcription<sup>1-3</sup>. Compared with pre-mRNA transcripts, PROMPTs, also termed as upstream antisense transcripts (uaRNAs), are short (50-2,000 nucleotides), much less abundant transcripts and are relatively unstable<sup>4-6</sup>. Such phenomenon has been also observed in the transcription of non-coding RNAs, including enhancer RNAs (eRNAs) and super enhancer RNAs (seRNAs)<sup>7-9</sup>.

Although bidirectional transcription is known to localize into a nucleosome-free region with two adjacent PICs, the relationship between two PICs is poorly understood<sup>8,10,11</sup>. Various studies have shown that the U1-PAS (U1-mediated polyadenylation sites) axis at active promoters acts as a *cis*-element to determine the directionality of transcription<sup>12,13</sup>. Specifically, the low occurrence of U1 snRNP sites and high abundance of polyadenylation signal (PAS) sites upstream of sense transcription start sites (TSS) attenuate PROMPT transcription, thus guaranteeing the transcription towards the sense direction<sup>12</sup>. However, thus far, little is known about how this unbalanced bi-directional transcription event is precisely initialized and mechanistically conducted in eukaryotes.

Integrator is a ~1.5 MDa protein complex containing 14-16 subunits in human<sup>14-16</sup>. It was first discovered to be tightly associated with the RNA polymerase II C-terminal domain (RNAPII-CTD) and is required for the maturation of uridylate-rich small nuclear RNAs (UsnRNAs)<sup>14</sup>. Subsequent functional and structural studies revealed that the Integrator complex, together with RNAPII, plays crucial roles in the attenuation of

transcription, pre-mature termination and enhancer RNA biogenesis<sup>17-21</sup>. As a member of the metallo- $\beta$ -lactamase superfamily, it is clear now that INTS11, together with INTS9, forms an endonuclease core of Integrator responsible for the cleavage of RNA during transcription, including enhancer RNAs (eRNAs), telomerase RNA, long non-coding RNAs, viral miRNAs and premature nascent protein coding transcripts<sup>19-23</sup>.

In addition to its endonuclease function, recent structure studies demonstrated that the Integrator complex contains two different enzymatic modules<sup>24,25</sup>. Along with the well-known endonuclease module composed of INTS11, INTS9, and INTS4 subunits, it also contains an interesting phosphatase module including protein phosphatase 2A (PP2A) and 2C subunits associated with the INTS6 and INTS8 subunits<sup>24,25</sup>. These two functional modules are placed on the opposite sides in the complex and the whole Integrator complex mimics a soft hat sitting on the top of RNA polymerase II<sup>25</sup>. The function of the phosphatase module has not been fully uncovered yet, but it has been proposed to associate with the phosphorylation dynamics of the C-terminal domain of RPB1 (RNAPII-CTD) during the pause-release and elongation stages of transcription<sup>26-29</sup>.

Due to the complexity of the Integrator structure, the question about its function during transcription is largely under debate. After its discovery, its role has been proposed into two distinct aspects: facilitating transcription or attenuating gene expression<sup>17-19,26,27,30-32</sup>. Analyses based on the length of genes or specific environmental responses further complicate the determination of its function during different stages of transcription<sup>18,30-36</sup>. However, most of the previous studies on Integrator were based on the use of traditional RNA interference tools, which generally needs 48-60 hours for depletion. Therefore, these observations could mix the direct evidences with nonessential or indirect outputs<sup>37</sup>.

To avoid unwanted secondary effects, we utilized the acute degradation system (AID), in which the depletion occurs in 1~2 hours, to analyze the direct function of Integrator<sup>38,39</sup>. Consistent with previous studies, we detected a massive PROMPT accumulation at most of active human genes after an acute depletion of Integrator in 1 hour<sup>19,40,41</sup>. Meanwhile, through multiple transcription dynamics studies, after Integrator depletion we observed a genome-wide attenuation of pre-mRNA synthesis in the pre-mRNA direction (sense direction), suggesting a counterbalance between the sense and the antisense transcription. To further address the specificity and mechanism, our dominant negative assay of the Integrator endonuclease subunit showed the transcription directionality specifically requires the endonuclease function of Integrator on PROMPTs. Furthermore, by mimicking Integrator cleavage with gapmer antisense oligonucleotides (ASOs) or introduction of U1 sites to the antisense region, we reveal the connection between U1 snRNA and PROMPTs to inhibit Integrator cleavage for the determination of transcription directionality. Together, our results suggest the U1-Integrator axis is critical for the rapid degradation of PROMPTs and facilitates the transcription for actively transcribed genes.

## Results

# Acute depletion of Integrator induces abundant PROMPTs

To accurately assess the function of Integrator, we utilized the mini-Auxin Inducible Degron (mAID) system to rapidly deplete Integrator subunits<sup>38,39,42</sup>. Specifically, the mAID tag was inserted at the C terminus of the endogenous *INTS11* or *INTS9* loci in the HCT116:TIR1 or the F74G mutant cell line to avoid affecting the expression levels of endogenous proteins (Fig. 1a, Extended Data Fig. 1a). After 15-minute treatment of 500  $\mu$ M IAA (indole-3-acetic acid, for INTS11-AID) or one hour with 10  $\mu$ M 5-Ph-IAA (for INTS9-AID2), targeted proteins were nearly undetectable (Fig. 1b). After the depletion of the target protein, we performed chromatin-associated RNA-seq (ChrRNA-seq) to monitor the on-going transcriptome changes<sup>20,43</sup>.

As expected, upon Integrator subunit depletion, abundant transcript signals are accumulated downstream of the annotated 3' end of *RNU11* and *RNU12*, two *UsnRNA* genes which 3' end processing is known to be specifically cleaved by Integrator (Extended Data Fig. 1b, c)<sup>14</sup>. To our surprise, we noticed dramatic accumulation of PROMPTs on 8,377 gene loci for both cell lines (Fig. 1c). Moreover, with INTS11 or INTS9 depletion, PROMPTs with over two-fold accumulations are mostly overlapped (Fig. 1d) and correlated well with each other (Extended Data Fig. 1e). Since PROMPTs have no annotation on the human genome (GRCh38/hg38), we developed an algorithm (PROMPT-Finder, methods for details) to define the range of PROMPTs in the human genome based on ChrRNA-seq results. Briefly, a PROMPT region is defined as a genomic section with significant enrichment of antisense transcripts (> 1kb) initiated immediately upstream of annotated TSSs (< 500bp). Plus, a given PROMPT region should not overlap with any active annotated genes in the same direction (Details in the methods section). Within the identified 8,377 gene loci in INTS11-AID cells, there are 7,070 gene loci harboring PROMPTs and showing significant accumulations of PROMPTs (> 2 fold, FDR < 0.05) upon INTS11 depletion (Fig. 1e-f). Furthermore, under INTS9 depletion, PROMPTs also accumulated at these 7,070 gene loci (Extended Data Fig. 1d, f). The overall PROMPTs accumulation in both INTS11 or INTS9 depletions correlated well with each other, suggesting the acute disruption of the Integrator endonuclease module induces PROMPT accumulation in the antisense direction. The difference of PROMPT numbers and levels called by PROMPT-Finder between INTS11-AID and INTS9-AID cells are likely due to the incomplete depletion of INTS9<sup>34,44</sup>.

Among all 7,070 gene loci that accumulates PROMPTs in the antisense direction, we noticed a broad range of dysregulation of their pre-mRNA on the sense strand (Fig. 1f). We next split all 7,070 gene loci into ten quantiles depending on their pre-mRNA transcript level in control samples (Fig. 1g, pre-mRNA). Among them, the top 90% of gene loci shows various levels of pre-mRNA reduction after the acute depletion of INTS11. Meanwhile, for the bottom 10% genes representing the lowest gene expression levels, both pre-mRNAs and PROMPTs levels are dramatically increased (Fig. 1g). This increase in transcription activity after Integrator depletion has been also reported in a recent publication<sup>29</sup>. To accurately evaluate sense and antisense transcription for actively transcribed genes, we selected the top 30% of highly expressed genes for further analysis (2,121 gene loci).

To confirm that the accumulation of PROMPTs is indeed through the depletion of INTS11, we exchanged the media containing IAA with fresh media, which should restore the level of INTS11 protein (Fig. 1h). As expected, after 7 hours of restoration, the 3' end extension of UsnRNAs (*RNU11* and *RNU12*) can be properly processed through the recovery of the cleavage function of Integrator (Extended Data Fig. 1b). Meanwhile, the PROMPTs were diminished upon the restoration of INTS11 (Fig. 1i, PROMPTs marked in red). As specific examples, we choose the *MYC* gene because it is one of the key transcription factors in cancer and *BMP4* gene for its importance in signaling transduction and embryogenesis<sup>45,46</sup>. Consistently, the massive accumulation of PROMPTs at the *MYC* and *BMP4* loci can be observed after 1 hour of the IAA treatment. In total, all 2,121 PROMPT-harboring gene loci show significant fluctuation of PROMPTs during the process of knockdown and restoration of INTS11 (Fig. 1j-k). Interestingly, for actively transcribed genes, after INTS11 depletion, the increased amount of PROMPTs are correlated with a decreased amount of pre-mRNA. Yet, the total amount of transcripts at each locus remains largely the same (Fig. 1k and 1l). Similar exchange of the levels of pre-mRNAs and PROMPTs can also be observed under INTS9 depletion conditions (Fig. 1k, Extended Data Fig. 2d-g). To avoid the off-target effects from the IAA treatment, we performed the rescue experiments with ectopic expression of INTS11. The gene expression profiles are confirmed the PROMPTs induction is dependent on INTS11 depletion (Extended Data Fig. 2a-c). Together, our results suggest that the acute depletion of Integrator leads to the strong accumulation of PROMPTs for transcriptionally active genes.

## Pre-mRNAs and PROMPTs are counterbalanced in highly expressed genes and induced genes

During the depletion and restoration of INTS11, we noticed that at the *MYC* and *BMP4* loci, pre-mRNA transcripts were attenuated along with the increased level of PROMPTs, suggesting that pre-mRNA and PROMPT transcription might counterbalance each other (Fig. 1l). For most active promoters in eukaryotes, the basal transcription machinery is believed to be inherently directional and the divergent promoter region typically harbors two distinct promoters with inverted orientations<sup>1,3,47</sup>. The assembly of two separated transcription pre-initiation complexes (PICs) is assumed to be correlated with each other<sup>1,48</sup>. Our previous results prompted us to examine if Integrator plays a key role to monitor transcription direction for the active gene loci.

Since ChrRNA-seq performed above did not have a spike-in control, to carefully examine the role of integrator in the dynamics of nascent RNA synthesis, we performed 4sU-seq (with spike-in control) to quantify the newly transcribed RNAs (containing 4-thiouridine) and transient transcriptome sequencing (TT-seq) for the transient transcriptome<sup>49</sup> (Extended Data Fig. 4a-b). In both cases, we observed dramatic inductions of PROMPTs transcription in *MYC*, *BMP4*, *XPO1* and *EP300* loci, while transcription in the sense direction is largely reduced (Fig. 2a, Extended Data Fig. 4d, h and i). With the detailed analysis for the same 2,121 highly transcribed genes, our data suggest that Integrator depletion interrupts the distribution of transcription at the promoter-proximal stage (Fig. 2b-d, TT-seq; Extended Data Fig. 4e-f, 4sU-seq). Moreover, after comparing PROMPTs and gene body signals from our TT-seq results, we noticed that the total amount of newly synthesized transcripts remains largely the same in all of these

loci (Fig. 2e). Similar result was also obtained by using the published Integrator depletion data (Fig. 2f; Extended Data Fig. 4c<sup>29</sup>).

Except for the coding genes, we analyzed our TT-seq results for the enhancer RNAs. According to the RNAPII peaks at the intergenic regions, we identified 3,259 regions as the enhancer loci (details in the methods section). Similar with previous published results, enhancer transcripts are accumulated after Integrator depletion<sup>20</sup>. Contrasting with protein coding genes, nearly all the enhancer loci do not show the strand preferences (Fig. 2g-h). Instead, primary transcripts are accumulated equally in both transcription directions (Fig. 2g, h bottom panel; Extended Data Fig. 3d-e). Similar with the lowly expressed genes in chrRNA-seq results (Fig. 1g), transcription levels at most enhancer regions are quite low (Extended Data Fig. 3d-e). This observation may suggest that transcription speed or capacity could contribute to the selection of transcription direction for eukaryotic gene expression.

To gain further insight into how the transcriptional apparatus is engaged for the actively transcribed genes, we applied epidermal growth factor (EGF) or interferon beta (IFN- $\beta$ ) induced systems to our Integrator AID system. By applying these systems, we can evaluate the function of Integrator on the initiation status of immediate-early genes (IEGs) upon stimulations<sup>1</sup> (Extended Data Fig. 5a, g). As expected, shortly after adding EGF or IFN- $\beta$ , 1,026 and 84 IEGs, which harbor detectable PROMPTs under the INTS11 depletion condition, were significantly induced, respectively (Extended Data Fig. 5d, h). In contrast, upon stimulation after INTS11 depletion, the inductions of IEGs were dramatically attenuated, along with significant accumulation of PROMPTs (*KLF2* and *NR4A1* for EGF induction, Fig. 2i; *DDX58* and *IFIT1* for IFN- $\beta$  response; Extended Data Fig. 5i). These alternating responses were further confirmed with RT-qPCR for *NR4A1* and *IFIT1* pre-mRNA transcripts, as well as for their PROMPTs (Extended Data Fig. 5f and j). From the analyses of 1,026 EGF induced IEGs, the counterbalance effect between PROMPTs and pre-mRNAs could also be observed (Fig. 2j and Extended Data Fig. 5e). Meanwhile, before and after the depletion, the total amount of highly induced nascent transcripts shows no change (Fig. 2k and Extended Data Fig. 5k-l). Together, our results suggest that the depletion of Integrator could lead to rapid changes of transcription distribution between the PROMPTs and pre-mRNAs at the actively transcribed genes.

## **Integrator affects RNAPII status and CTD phosphorylation dynamics**

Given that Integrator is tightly associated with RNAPII-CTD throughout the whole transcription processing<sup>19,26,30,31</sup>, we performed RNAPII chromatin immunoprecipitation sequencing (ChIP-seq) to measure exactly how RNAPII is actively engaged at promoter region to conduct transcription initiation. Upon depletion, the accumulations of RNAPII were observed at upstream regions of TSSs in the *MYC*, *BMP4*, *EP300* and *EZH2* genes, along with decreased levels of RNAPII at the TSS sites (Fig. 3a and Extended Data Fig. 6c). For the same 2,121 highly expressed genes, our heatmap analysis shows clear accumulation of RNAPII occupancy at the upstream of TSS regions and reduction at the TSS sites upon depletion (Fig. 3b). Meanwhile, similar re-distribution of RNAPII occupancies could also be detected from the IEGs with EGF induction after INTS11 depletion (Extended Data Fig. 6f-h).

It is well-known that the phosphorylation dynamics at heptad repeats of the RNAPII-CTD orchestrate different stages of transcription<sup>50</sup>. Previous reports suggest that Tyr1P might associates with the antisense transcription and functions at the transcription initiation stage<sup>51,52</sup>. Moreover, a recent structural study of Integrator indicates that Tyr1P is adjacent to the RNAPII-Integrator interface<sup>25</sup>. Interestingly, our Tyr1P ChIP-seq shows a dramatical enrichment of CTD-Tyr1 phosphorylation at PROMPT regions and an obvious reduction at TSS sites at the *MYC* and *BMP4* loci (Fig. 3a, middle panel). Genome-wide analyses also reveal that the changes of Tyr1P engagement after depletion is similar to that of total RNAPII at PROMPT regions and TSSs (Fig. 3d, left panel), indicating Tyr1P dynamics might be tightly associated with Integrator to regulate PROMPTs during the promoter-proximal stage of transcription (Fig. 3c, left panel).

We next tested several other phosphorylation forms of RNAPII-CTD, within which we observed a noticeable accumulation of Ser2P at TSS sites of *MYC* and *BMP4* loci (Fig. 3a, bottom panel). Interestingly, the amounts of Ser2P showed a remarkable increase at TSS sites of transcriptionally active genes and east only slightly enriched for enhancer TSSs after depletion (Fig. 3c and 3d, right panel; Extended Data Fig. 7f, right panel). Consistent with previous observations<sup>29-31</sup>, our results suggest that the functional disruption of the endonuclease module of Integrator might facilitate the promoter-proximal pausing of RNAPII (Fig. 3h). Together with the total amount of polymerase remaining roughly the same before and after Integrator depletion (Fig. 3i), our results indicate that Integrator controls divergent transcription likely through refraining RNAPII Tyr1 phosphorylation at the initiation stage of transcription, as well as transcription pause-release via RNAPII Ser2 phosphorylation.

The similar changes of Tyr1P and Ser2P were also observed at enhancers and super enhancers after Integrator depletion (Extended Data Fig. 6e, Extended Data Fig. 7f). Meanwhile, within the 2,121 highly expressed gene loci or 1,026 EGF induced gene loci, the combination of RNAPII for both sense and antisense still remains at a similar level (Fig. 3g and Extended Data Fig. 6h, gray bars). Together with the western blot of RNAPII and TT-seq for newly synthesized transcripts, our results suggest that Integrator depletion might not affects the RNAPII loading on the genome, but rather it interrupts the distribution of bi-directional transcription for actively transcribed genes.

## The endonuclease activity of Integrator is critical for bidirectionality

Integrator is known to possess the endonuclease activity for small nuclear RNAs, viral miRNAs and enhancer RNAs<sup>14,20,23</sup>. Given that INTS11 functions as an endonuclease and regulates genome-wide non-productive transcription<sup>19</sup>, we speculated that Integrator might directly cleave PROMPTs to destabilize the antisense transcripts. Previous studies have shown that a single point mutation (E203Q) in the catalytic domain of INTS11 impairs the processing of UsnRNAs and enhancer RNAs<sup>14,20</sup>. We thus performed a dominant negative assay by expressing INTS11 or its catalytic-dead E203Q mutant in INTS11-depleted cells, so as to investigate whether the endonuclease function of Integrator is necessary for the regulation of PROMPT transcription (Fig. 4a). Indeed, the ectopic expression of INTS11 could



substantially reduce the 3'-end accumulation of the UsnRNA genes (*RNU11* and *RNU12*) and PROMPT production of *MYC* and *CCND1* genes after depleting the endogenous protein (Fig. 4b, Extended Data Fig. 8a). Notably, the accumulation of PROMPTs caused by INTS11 depletion cannot be suppressed by overexpression of the E203Q mutant, suggesting that the endonuclease activity of Integrator is required for PROMPT processing on chromatin (Fig. 4b).

To confirm the endonuclease activity to PROMPTs is critical for transcription directionality, we designed chemically modified gapmer antisense oligonucleotides (ASOs), which can hybridize with nascent transcripts on chromatin in a sequence-specific manner and cleave the nascent RNA via an RNase H1 mechanism<sup>53-55</sup>. We assumed that the PROMPT cleavage by gapmer ASO could mimic the Integrator endonuclease function without affecting the other functions of the Integrator complex (Fig. 4c). Upon INTS11 depletion, PROMPTs accumulation and the pre-mRNA reduction at *MYC*, *RBM14* and *SRRT* gene loci were measured by RT-qPCR (Fig. 4d and Extended Data Fig. 8b, first two columns for each panel). Meanwhile, after the treatment with sequence-specific gapmer ASOs, the reduction of PROMPTs and increased level of pre-mRNAs were monitored accordingly (Fig. 4d and Extended Data Fig. 8b, ASO+). These results indicate that targeted ASO cleavage of PROMPTs indeed mimics Integrator cleavage for the transcriptional balance between sense and antisense strands. More importantly, because ASO approaches can only mimic the endonuclease function of Integrator for the antisense strand without affecting transcription at the sense strand or interrupting other functions of Integrator, it clearly suggests that the PROMPT cleavage is critical for the preference of transcriptional direction for sense over antisense strands in eukaryotic cells.

In humans, INTS11 is also termed CPSF73L, due to its highly homologous protein sequence with CPSF73, which is known to cleave the 3' end of pre-mRNAs and some PROMPTs<sup>19,56,57</sup>. To distinguish the endonuclease function of INTS11 and CPSF73, we created a CPSF73-AID2 cell line, which showed a dramatic decrease of CPSF73 level after 1 hour of the 5-Ph-IAA treatment (Extended Data Fig. 8c). Upon CPSF73 depletion, we observed significant extension at the 3' end of *MYC*, *ACTB* and *EGFR* pre-mRNAs (Extended Data Fig. 8e-f). However, for all 3,847 genes, which have at least two-fold accumulation of 3'-end pre-mRNA (Extended Data Fig. 8g), we observed no significant enrichment of PROMPTs, albeit these genes have considerable amounts of PROMPTs in INTS11-depleted cells (Fig. 4e-g, Extended Data Fig. 8h). These data suggest that Integrator and CPSF73 might have distinct preference to process RNAs during transcription<sup>19</sup>.

Taken together, our results suggest there is a balance between sense and antisense strand transcription for actively transcribed gene loci. Under normal conditions, antisense strand transcription is likely suppressed through rapid cleavage of PROMPTs by Integrator and degraded through an exosome dependent RNA degradation mechanism<sup>5,58</sup>. Without the cleavage function of Integrator on the antisense strand, transcription could be activated on the antisense strand, which leads to PROMPTs accumulation and pre-mRNA reduction for the actively transcribed gene loci.

# Integrator coordinates with upstream U1 signal for transcription directionality

Given the well-recognized U1-PAS theory<sup>12,13</sup>, we next asked if Integrator could act as the trans-factor to connect U1 snRNP and its U1 sites (the cis-element) on the production of nascent transcript for the unbalanced sense/antisense transcription. We thus carried out a genome-wide search for the 1st U1 snRNP sites near TSSs with the *de novo* motif analysis as described before<sup>12</sup>. We grouped genes according the numbers of predicted strong or medium U1 sites in the 1kb window upstream of TSS sites (Fig. 5a and Extended Data Fig. 9a-b). Notably, under normal condition, the amount of PROMPT transcripts is generally higher if the PROMPT regions harbor two or more U1 binding sites in our chrRNA-seq data sets (Fig. 5b, left panel). After INTS11 depletion, the induction of PROMPTs is much less for these genes with multiple U1 sites (Fig. 5b, right panel), suggesting that U1 sites may counteract with Integrator to facilitate PROMPT production.

Apart from the numbers of U1 sites, the distance between the U1 site and TSS could also affect the productivity of PROMPTs. After the categorization of the 1st U1 sites in a 2 kb window upstream of TSS sites (Fig. 5c), our data suggests that the further the 1st U1 site away from the TSS in the PROMPT region, the less PROMPTs can be generated under native conditions (Fig. 5d, left panel). Meanwhile, upon INTS11 depletion, PROMPTs were produced more and less controlled by the appearance of U1 sites (Fig. 5d, right panel). Combining with the number of U1 sites and the position of the 1st U1 site in the PROMPT regions, we speculate that the U1 site might inhibit Integrator cleavage for the production of transcription on the antisense strand. Indeed, comparing with the sense direction that harbors large amount of U1 sites, there are much fewer U1 sites distributed in the antisense direction (Extended Data Fig. 9a).

A recent publication shows that U1 snRNP enhances transcription in the pre-mRNA direction<sup>59</sup>, suggesting U1 sites could also contribute to PROMPT transcription. Based on the number and distance of U1 sites on the antisense strand (in a 2 kb window upstream of TSS), we calculated a U1 score for every identified PROMPT, which should reflect the contribution of U1 sites to PROMPT production (Fig. 5e, details in methods). We found that the higher a U1 score on a specific PROMPT region, the higher level of antisense transcription and more PROMPTs were accumulated in our chrRNA-seq data sets (Fig. 5f, left panel). Interestingly, upon INTS11 depletion, there are less changes for the PROMPTs possessing a high U1 score, suggesting they are less regulated by Integrator cleavage activity (Fig. 5f, right panel). This observation is not only reflected by our chrRNA-seq data, but also exhibited in our TT-seq results (Fig. 5g).

To carefully check if our analysis is correct, we selected two gene loci with high U1 scores and two loci with low U1 score (Fig. 5e). For *MYC* gene loci, lacking or having fewer U1 sites in the antisense direction, our multiple sequencing results suggest that Integrator could rapidly cleave PROMPTs and guide RNAPII machinery to the pre-mRNA direction (Fig. 5h left panel). Upon Integrator depletion, there was induced transcription on the antisense strand and attenuated transcription on the sense strand (+ IAA lanes). The

balance of transcription at these gene loci is strongly regulated by Integrator cleavage. In contrast, for *CCNY* gene loci, with high U1 scores in the antisense direction, Integrator cleavage is inhibited by U1 sites on the antisense strand (Fig. 5h right panel). Consistently, after Integrator depletion, the bi-directional transcription pattern remains largely the same.

## U1 snRNA together with Integrator mediate transcription dynamics

To further understand the connection between Integrator and U1 snRNA during transcription, similar with previous analysis, we compared all the predicted U1 sites for the 2,121 actively transcribed gene loci with 3,259 enhancer regions (Fig. 6a). These results showed an obvious bias toward pre-mRNA side for the active genes (red line) and no difference for the enhancer regions (blue line). Comparing with our TT-seq results, there is a shift of transcription direction for active genes after Integrator depletion (Fig. 6b, red line). However, such a shift is absent at enhancer regions, which exhibit elevated nascent RNA synthesis at both sides (Fig. 6b, blue line). Based on our analyses, enhancer RNA transcription is directly regulated by Integrator cleavage and it fits with the previously published results<sup>20</sup>.

Moreover, after we split all 7,070 transcribed genes by their reads number from our TT-seq results, the transcription levels of eRNAs or PROMPTs are quite low (Fig. 6c, top panel). Meanwhile, upon Integrator depletion, the regulation of low expressed transcripts is quite different with highly transcribed genes (Fig. 6c, bottom panel, red bars vs. blue bars). Thus, Integrator-dependent regulation might act in a dose dependent manner (Fig. 6c, bottom panel).

To directly link the U1 snRNA with Integrator, we next performed RNA-IP by pulling down the ectopic FLAG-INTS1, and using FLAG-CPSF73 as the negative control (Fig. 6d). Interestingly, we observed strong and specific interaction of *RNU1* (*U1 snRNA*) with INTS1, in contrast to CPSF73 which has little interaction with *RNU1* (Fig. 6e). Moreover, this interaction is specific for U1 snRNA with INTS1, but not for U2 snRNA (Fig. 6e), suggesting this U1-Integrator interaction is not within the processing steps of UsnRNA maturation<sup>14</sup>.

To further measure the importance of U1-Integrator interaction, especially for transcription directionality, we inserted either a DNA fragment containing 3 x U1 site sequences or a natural sequence with multiple U1 sites (from the PROMPT region of the *Cttn* gene locus) in the PROMPT region of the *FUS* gene (Fig. 6f, Extended Data Fig. 9d-e). After Integrator depletion, significant accumulations of PROMPTs can be detected at the PROMPT region for both cases (Fig. 6g, panel + IAA in PROMPT section). Meanwhile, the reduction of *FUS* pre-mRNA could simultaneously be observed (Fig. 6g, panel + IAA in pre-mRNA section). Interestingly, after the insertion of 3 x U1 sites in both cases, significant amounts of PROMPTs had accumulated with a reduction of nascent transcript in the sense direction (Fig. 6g, panel + U1). Moreover, there was no significant inhibition of cleavage response for both PROMPT and pre-mRNA directions after the Integrator depletion (Fig. 6g, panel + U1 + IAA). Taken together, these results suggest that during transcription, U1 signals on PROMPTs could function together with U1-Integrator to inhibit Integrator cleavage and impact on bi-directional transcription.

## Discussion

Bi-directional transcription has been well-documented at the promoters of mammalian protein-coding genes, but little is known about the regulatory bases that guide transcription to the sense direction and refrain the antisense transcription. By applying the rapid depletion systems for multiple subunits of Integrator, we show a clear interplay of transcription for both strands at active gene promoters through the endonuclease function of the Integrator complex. This transcriptional interplay mainly depends on the local transcription activity and U1 snRNA binding sites around the TSS region. Similar with previous studies, we find Integrator cleaves enhancer RNAs and other lowly expressed genes<sup>19,20,40</sup>. As most of the enhancer sites lack of U1 sites, the Integrator cleavage cannot be constrained by U1 snRNA. For the actively transcribed gene loci, U1 snRNA binding sites surrounding TSSs exhibit an uneven distribution, which is far more abundant at sense than antisense direction (Fig. 6h). In this case, PROMPTs could be efficiently cleaved by Integrator and rapidly degraded through the RNA exosome mechanism<sup>5,19,58</sup>. Moreover, because Integrator cleavage can be inhibited by U1 sites on the sense strand, the entire transcription machinery on the gene locus favors the transcription to the pre-mRNA side. During this step, the U1 snRNA reverse-complementarily binding with nascent transcripts to inhibit Integrator cleavage is the critical step to perform the directional selection locally. Taken together, we propose a U1-Integrator model to govern the transcription bi-directionality for actively transcribed genes (Fig. 6h).

Untill now, the function of Integrator for the early steps of transcription was still under debate<sup>15,26</sup>. The cleavage activity of Integrator for non-polyadenylated RNAs has been well documented to have functions on RNA maturation, biogenesis and transcription activation<sup>14,20,31</sup>. However, a genetic repressor screen has suggested that the complex may have the function to abrogate transcription<sup>17-19</sup>. Meanwhile, an association of Integrator with non-productive transcription was reported within ~ 3kb downstream of promoter-proximal regions in a sequence-independent manner<sup>19</sup>. As Integrator has multiple different functional modules and participates at different stages of transcription, rapid and clean tools are needed to dissect its function<sup>26</sup>. Along this line, a recent paper also suggests that the elongation velocity at protein-coding genes shows broadly decreased productive elongation after acute INTS11 depletion<sup>60</sup>. Moreover, by expression of a truncated form of INTS8 in an acute INTS8 depletion system, a distinct activation of transcription has been reported for the function of the phosphatase module, suggesting different enzymatic modules in the Integrator complex may present different functions during transcription<sup>29</sup>. Consistent with this notion, our results are focused on the endonuclease function of Integrator complex. It seems that the rapid degradation of PROMPTs is initiated through Integrator cleavage. Moreover, through multiple structure and mechanism studies, it suggests that Integrator may serve as a scaffold complex, working together with other transcription factors or RNAs, for the context-based access control on chromatin. The detailed picture of how the Integrator complex involved into different steps of transcription still needed to be addressed.

In eukaryotes, it is known that the abundance of U1 snRNP is much higher than that of the other snRNPs. A “Telescripting” mechanism has been proposed to suppress premature termination in the pre-mRNA

strand from Dreyfuss's laboratory<sup>61-63</sup>. Consistent with the telescripting mechanism, our data do support that the presence of U1 snRNAs on chromatin and their pairing with pre-mRNAs are critical for the controlling of transcription. Moreover, our results further suggest that the pairing between U1 snRNAs with nascent transcripts could inhibit Integrator cleavage for premature termination. Importantly, our data pointed out that this inhibition is crucial for the natural selection of transcription direction. From another point, somatic mutations in the 5' splice site binding region of U1 snRNA has been identified in different cancer types<sup>64,65</sup>. As one of the most abundant noncoding RNAs in eukaryotic cells, it has been shown to regulate long noncoding RNA (lncRNA) retention in the nucleus, even though the mechanism is not clear<sup>24</sup>. Similar with blocking the 5' end of U1 snRNA, our results of INTS11 depletion show a reduction of transcription of active genes, suggesting a counterbalance model of endonuclease activity of Integrator with U1 snRNA or U1 snRNP during transcription. As Integrator and U1 snRNA were discovered to be biochemically associated with the RNAPII complex, our results suggest that the U1-Integrator axis may serve as a decisive factor for the rapid removal of transcription from the anti-sense strand, which determines the proper direction for pre-mRNA transcription in eukaryotic cells.

## Declarations

### Acknowledgements

We thank the Lai and Dang lab members for constructive discussions and help for genome-wide sequencing works. We also acknowledge Dahua Chen lab at Institute of Biomedical Research in Yunnan University for their help on high-throughput sequencing.

### Funding:

This work was supported by grants from the Yunnan Province Science and Technology Department (202201BF070001-015 to F.L.), the National Natural Science Foundation of China (32070626 to F.L., 32171262 and 31871254 to Y.D.), the Open Research Program of State Key Laboratory for Conservation and Utilization of Bio-Resource in Yunnan (2021KF002 to F.L.) and the startup funds from Yunnan University for F.L. and Y. D.

### Competing interests:

The authors declare that they have no competing interests.

### Author Contributions

F. Lai and Y. Dang conceived and supervised the overall project. J.Y. performed most of the experiments under the supervision of G.R. J.L. performed bioinformatics analyses and figure preparation with J.Y. L.M and X.G. generated cell line with 3 X U1 sequences. W.S. generated CPSF73-AID2 cell lines and performed the ChrRNA-seq. L.M. performed TT-seq. S.L. generated *HCT116:TIR1(F74A)* TET-ON cell line. B.P. performed the cut-tag experiments for RNAPII and helped for INTS11 overexpression experiments. B.P.,

Z.L. and W.S. generated INTS9 cell line and performed the INTS9 ChrRNA-seq. S.C., X.G., B.W., A.D., D.H. and J.Y. prepared all the samples for the high-throughput sequencing. F.L., Y.D., J.Y. and J.L. wrote the manuscript with critical feedback from all co-authors.

## **DATA and Materials availability**

All sequencing data generated as a part of this study have been deposited to NCBI GEO under the accession number GEO: GSE207268 and GSE207269 (NCBI tracking system #23081142).

All pipelines and customized codes are available on GitHub ([https://github.com/Berialim/PROMPTs\\_analysis\\_visualization](https://github.com/Berialim/PROMPTs_analysis_visualization)). Further information and requests for resources and reagents generated in this study are available upon request.

## **Methods**

### **Cell culture and generation of AID Cell lines**

HCT116 cell lines were cultured in DMEM (Thermo Fisher Scientific), supplemented with 10% (v/v) FBS, 2 mM L-glutamine at 37°C. To generate HCT116-AID cell lines, the HCT116-OsTIR1 cells were transfected into a 10 cm petri dish with 4.8 µg of guide RNA plasmid (based on pX330-U6-Chimeric\_BB-CBh-hSpCas9 (addgene #42230)) and 3.6 µg of donor plasmids (pMK289 (addgene #72827) and pMK290 (addgene #72828)) by using the Calcium Phosphate Cell Transfection Kit (Beyotime, C0508). 8 hours after transfection, the media were changed with fresh media and cells were selected with 500 µg/mL neomycin (Biofrox, 1150GR005) and 100 µg/mL hygromycin (Sino Biological Inc, 50708-mcch). After 10 ~ 15 days of selection, individual clones were isolated and screened by genomic DNA PCR with corresponding primer sets (details in Supplementary Table 1). The correct clones were further confirmed by the western blot with corresponding antibodies. Before the experiments, 500 mM Indole-3-Acetic Acid Solution (Phytotech, I364) was dissolved in DMSO as the stock solution.

To build the inducible AID system, HEK293T cells were transfected with pSW-2XFlag-TIR1 (F74A) BLA-TET-ON plasmid together with two helper plasmids (psPAX2 and pMD2.G) using the standard transfection protocol of Lipofectamine 2000 (Invitrogen, 11668019). HCT116 cells were infected twice and screened with 6 µg/ml Blastidicin for 14 days. The HCT116-OsTIR1(F74A) clones were confirmed with the western blot against OsTIR1 (Anti-OsTIR1 pAb, MBL PD048). The development of HCT116-AID2 cell lines were similar, except with 10 µM of 5-Ph-IAA (MCE, HY-134653) after 24 hours of the 1µg/ml Doxycycline induction.

### **Chromatin RNA sequencing (ChrRNA-seq) and data processing**

ChrRNA-seq was performed as previously described<sup>53</sup>. Briefly, 5 ~ 10 millions of cells were suspended in cold cytoplasmic lysis buffer (0.15% NP-40, 10mM Tris pH7.5, 150mM NaCl) and incubated on ice for 10 min. The cell lysate was carefully layered onto a cold sucrose buffer (10mM Tris pH7.5, 150mM NaCl,

24% sucrose W/V) and the cytoplasm fraction (supernatant) was removed after centrifugation. The nuclear pellet was gently resuspended in cold glycerol buffer (20mM Tris pH 7.9, 75mM NaCl, 0.5mM EDTA, 50% glycerol, 0.85 mM DTT) and lysed with nuclear lysis buffer (20mM HEPES pH7.6, 7.5mM MgCl<sub>2</sub>, 0.2mM EDTA, 0.3m NaCl, 1M urea, 1% NP-40, 1mM DTT). Following a quick centrifugation, the chromatin was isolated from the pellet fraction.

The chromatin-associated RNA was isolated with the standard Trizol protocol (Invitrogen, cat. no. 15596018). Genomic DNA was removed following the protocol of DNase I treatment (Thermo, EN0521). Chromatin RNA-seq (ribo-depleted) libraries were produced with a VAHTS Universal V8 RNA-seq Library Prep kit for MGI (Vazyme, NRM605-02). ChrRNA-seq libraries were quantified with Bio-Fragment Analyzer (Bioptic, c100001) and sequenced on MGI 2000 platform (MGI-SEQ, BGI).

Raw reads of chromatin RNA-seq were processed with Trim Galore (v0.6.6)

([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) to remove adaptors and low-quality reads with the parameter “-q 30”. Ribosomal RNA contaminations were removed using Bowtie (v1.3.0)<sup>66</sup> with “-un”. Remaining reads were aligned to the human genome (UCSC hg38) with STAR (v2.7.9a)<sup>67</sup>, which were converted to BAM file and sorted with SAMtools (v1.7)<sup>68</sup>. The quantification of gene and PROMPTs counts matrix was carried out with the featureCounts tool from Subread (v2.0.1)<sup>69</sup>. Differential expression analysis and normalization of gene and PROMPTs counts matrices were performed using the DESeq2 R package (v1.34.0)<sup>70</sup>. Biological replicates BAM files were merged by SAMtools (v1.7)<sup>68</sup>. Bigwig file were generated and normalized using merged BAM files by bamCoverage of deepTools (v3.5.1)<sup>71</sup> with parameter “-normalizeUsing RPKM” (RPKM = Reads Per Kilobase per Million mapped reads), and visualized by Integrative Genomics Viewer (IGV) (v2.9.4)<sup>72</sup>.

### **Chromatin immunoprecipitation sequencing (ChIP-seq) and data processing**

About  $1 \times 10^7$  cells were cross-linked with 1% formaldehyde for 10 min and quenched in 125 mM of glycine for 5 min at room temperature. After washing twice with ice cold PBS, the cells were re-suspended in cold ChIP buffer (150 mM NaCl, 1% Triton X-100, 0.7% SDS, 500 mM dithiothreitol, 10 mM Tris-HCl and 5 mM EDTA with fresh protease inhibitors) on ice for 20 min. Chromatin shearing was performed using Covaris ME220 ultrasonic generator. After clearance of the sonicated chromatin by centrifugation at 14,000 rpm for 10 min, chromatin fragments were immuno-precipitated overnight at 4°C with 2–4 µg of appropriate antibodies and 30 µL of Dyna Protein A or G beads (Invitrogen, 11204D or 11202D). The next day, beads were washed twice with cold Mixed Micelle Buffer (150 mM NaCl, 1% Triton X-100, 0.2% SDS, 20 mM Tris-HCl, 5 mM EDTA, and 65% sucrose), twice with cold Buffer 500 (500 mM NaCl, 1% Triton X-100, 0.1% Na deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, and 1 mM EDTA), twice with cold LiCl/detergent Buffer (250 mM LiCl, 0.5% Na deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, and 1 mM EDTA) and one wash with 1 x cold TE buffer. The immunoprecipitated chromatin fragments were eluted with 1 x TE buffer containing 1% SDS and incubated overnight at 65°C to reverse crosslinks. The chromatin fragments were treated with 0.5 mg/mL proteinase K for 3 h. DNA was purified by phenol/chloroform and

precipitated with isopropanol with Glyco-Blue (Invitrogen, AM9516). The DNA libraries were constructed with the VAHTS Universal DNA Library Prep Kit for MGI (Vazyme, NDM607-02) and sequenced on MGI 2000 instrument (MGI-SEQ).

Raw reads were first processed as described above and mapped to the human genome (UCSC hg38) using Bowtie2 (v2.3.4.1)<sup>66</sup> with default parameters. Duplicate reads were removed by Picard (v2.25.7) (<https://broadinstitute.github.io/picard/>) and Bigwig files were generated with deepTools<sup>71</sup> and normalized with parameter “--normalizedUsing RPGC” and “--effectiveGenomeSize 2913022398”. Peak calling was performed with MACS2 (v2.2.7.1)<sup>73</sup> with default parameters.

### **4sU-seq and TT-seq and data processing**

Detailed procedures of 4sU-seq and TT-seq were followed with previous published papers<sup>74,75</sup>. For 4sU-seq, cells were treated with 4-thiouridine (4sU) at a final concentration of 2 mM for 15min, and total RNA was extracted by Trizol (Invitrogen, cat. no. 15596018). The mixture of 80 µg of total RNA and 2 µg 4sU-labeled RNA of *Drosophila* S2 cells (as spike-in) was biotinylated in 4sU-seq biotinylation mix at room temperature for 2 h. Subsequently, RNA was purified by chloroform extraction twice, and precipitated with isopropanol. After denaturation for 10min at 65°C, RNA was incubated with 50 µl pre-washed streptavidin magnetic C1 beads (Invitrogen, cat. no. 65002), with gentle rotation for 15 min at room temperature. Beads were washed five times with beads wash buffer (1 M NaCl, 5 mM Tris-HCl pH = 7.5, 0.5 mM EDTA, 0.05% Tween 20). The 4sU-labeled RNA was eluted with 100 mM DTT, and purified by kit (ZYMO RESEARCH, Cat R1016). RNA library was constructed with VAHTS Universal V8 RNA-seq Library Prep kit for MGI (Vazyme, NRM605-02). The quality of library was monitored by Bio-Fragment Analyzer (Bioptic, c100001) and library was sequenced with MGI 2000 instrument (MGI-SEQ, BGI). For TT-seq, cells were treated with 2 mM 4sU for 7 min and quenched with direction RNA extraction by Trizol reagent (Invitrogen, cat. no. 15596018). The mixture of 80 µg total RNA and 2 µg 4sU-labeled RNA of *Drosophila* S2 cells (as spike-in) was fragmented using 20 µl 1 M NaOH for 23 min on ice, and neutralized with 80 µl 1 M Tris-HCl (pH = 6.8). RNA was biotinylated in TT-seq biotinylation mix for 30 min at room temperature.

Raw reads were processed as described above and mapped to the human genome and *drosophila* genome (UCSC dm6) using STAR<sup>67</sup> with parameter “--outFilterMultimapNmax 1” to remove multi-mapped reads. Low mapping quality (MAPQ lower than 30) and duplicate reads were further removed from BAM files by SAMtools<sup>68</sup>. The number of spike-in dm6 reads counted by SAMtools<sup>68</sup> was used to calculate the normalization factor  $\alpha = 1e6/dm6\_count$ . Bigwig files were generated and normalized with merged BAM files by deepTools<sup>71</sup> with scaling factors of spike-in. Gene expression quantification was performed with featureCounts<sup>69</sup>. Reads counts were normalized by both scaling factors of spike-in and gene length.

### **Lentiviral transduction**

Lenti expression plasmids INTS11 or INTS11 (E203Q) were transfected with two helper plasmids (psPAX2 and pMD2.G) into HEK293T cells by Lipofectamine 2000 (Invitrogen,11668019). The fresh



culture media were replaced and the viral supernatants were collected twice after 24 hours and 48 hours of transfection. The HCT116-AID cells were infected with virus for 70 hours and harvested 60 hours after the IAA treatment. The efficiencies of protein expression were measured by western blots with appropriate antibodies and quantitative RT-PCR for the products of transcription. All the antibodies and PCR primer sequences are listed in the supplementary table 1.

### **Antisense oligonucleotide transfection**

The HCT116-AID cell line was cultured in 5% CO<sub>2</sub> at 37°C. When cell density reached to 80–90%, 100nM gapmer ASO was transfected into cells by calcium transfection method. After 6–8 hours of transfection, fresh medium was replaced and IAA was provided at the same time. After 18 hours of transfection, Trizol (Invitrogen, cat. no. 15596018) was used to extract total RNA for RT-qPCR. ASOs used in this study to cleave PROMPTs of MYC, SRRT and RBM14 are 20 nucleotides in a standard sandwich structure (10 unmodified deoxynucleotides flanking by 5 MOE-modified ribonucleotides with phosphorothioate backbone)<sup>76</sup>. ASOs were solubilized in water (DNase-/RNase-free)

The sequence of the MYC ASO is 5'- **TACTGCTACGGAGGAGCAGC**-3'

The sequence of the RBM14 ASO is 5'-**AATTAATGGCACGAGGGCTT**-3'

The sequence of the SRRT ASO is 5'- **TGTGCCTGGCCCTAAATATT**-3'

The bold letters represent MOE-modified bases.

### **RNA immunoprecipitation (RIP)**

About 1 x 10<sup>6</sup> cells were transfected with pCMV2-INTS1-FLAG and pCMV2-CPSF73-FLAG for 36 h and lysed with lysis buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.1% Triton, 1mM EDTA). The RIP experiment did not include any crosslinking steps. After centrifuging at 14000 rpm for 10 min, the supernatant was incubated with 50µL Anti-FLAG® M2 Magnetic Beads (Sigma-Aldrich) for 2 h at 4°C. The beads were washed three times with ice-cold lysis buffer and twice with ice-cold PBS. Immunoprecipitated RNA was extracted with Trizol (Invitrogen, cat. no. 15596018) reagent and used for qPCR assay.

### **Identification of active promoters and enhancers**

Active promoters were defined within 500bp regions immediately upstream of TSS (transcription start site), which are overlapped with peaks called from RNAP II ChIP-seq. To identify active enhancers, we first selected genomic regions that contain RNAP II peaks and are at least ± 10kb away from any annotated gene. Next, we use ROSE (v0.1)<sup>77,78</sup> to identify active enhancers and super-enhancers from those regions. Most BED files are processed using BEDtools<sup>79</sup>.

### **PROMPT-Finder**

First, we set background area as intergenic regions which are 20kb and 10kb from the upstream or downstream of annotated genes (UCSC hg38). To generate the empirical distribution of ChrRNA-seq background signals, we randomly selected 10,000 windows (200bp) from background areas of each chromosome and calculated the chrRNA-seq density of each window, resulting in an empirical distribution function for each chromosome. Next, we used a sliding window (200bp in length, 10bp steps) to scan across the genome. ChrRNA-seq signals of each sliding window were evaluated with corresponding empirical distribution function (e.g., chromosome I). The probability of each window was further adjusted by false discovery rate (FDR). Windows with FDR > 0.05 were removed. Remaining windows in upstream antisense region of active promoters (50kb upstream and 2kb downstream of TSS) were merged if the gap between windows is less than 400bp. For a given PROMPT region, the portion overlapping with downstream active genes of the same direction were truncated. We eliminated the PROMPT region if length < 1kb. We next estimated the differential expression of these PROMPT regions between treatment and control by featureCounts (v2.0.1)<sup>69</sup> and DESeq2 (v1.34.0)<sup>70</sup>. PROMPT regions were defined as FDR > 0.05 and FC > 2.

### **Transcript activity and RNAP II loading balance analysis**

Reads in identified promoter regions and genes were quantified using featureCounts<sup>69</sup>. ChrRNA-seq reads counts were normalized with library size and region (gene or PROMPTs) length. TT-seq reads counts were normalized by scaling factors and region length. As chrRNA-seq and TT-seq are strand-specific, the total read counts were calculated as the sum of reads in PROMPT and gene regions. The total read counts in ChIP-seq were calculated from the end of the PROMPT region to the end of the gene. Then, the + IAA and CTRL were compared to identify the RNAP II loading and transcription activity changes in PROMPT, gene(pre-mRNA), and total.

### **U1 site prediction**

Prediction of U1 snRNA recognition sites was performed as described<sup>80</sup>. The 5' splice site motif was calculated in the known intron 5' site (3nt in exon and 6nt in an intron) of the human genome (UCSC hg38). The motif of the 5' splice site was used by FIMO<sup>81</sup> to search for significant matches (P < 0.01). Matches were then scored by the maximum entropy model<sup>82</sup>. All annotated 5' splice sites were also calculated with maximum entropy score to classify the predicted sites. Sites with scores larger than the median of annotated 5' splice sites were classified as strong. Sites with scores lower than the median but higher than the first quartile were classified as medium.

### **Classification of genes by predicted U1 site and U1 score calculation**

Active genes were removed if their upstream 5kb region overlapped with the putative promoter of annotated genes. For the remaining genes, we only took into account predicted U1 sites that were located upstream 2kb in antisense direction and calculated the distance between 1st U1 and TSS. For estimation of U1 distance in the antisense direction, genes were classified as "0-0.5kb", "0.5-1kb" and "1-2kb" by the

1st U1 distance. For estimation of U1 abundance in the antisense direction, we only counted the number of U1 sites within 1kb upstream of TSSs. Genes were classified as “0,” “1,” and “2+” by the U1 number in the upstream antisense regions. For the estimation of PROMPT level, we took into account reads in 5kb regions in the upstream antisense direction with featureCounts<sup>69</sup>. Reads counts were normalized by both feature-length and scaling factors (TT-seq and 4sU-seq were spike-in) generated by deepTools<sup>71</sup>. U1 Scores were calculated as the following equation:

$$U1Score = \sum_{i=1}^n \left(1 - \frac{U_i - T}{2000}\right)$$

## Visualization of data through Heatmaps, average line plot, boxplots, and violin plots

The heatmap plots of Bigwig files were generated using computeMatrix and plotHeatmap from deepTools<sup>71</sup>, while average line plots were created with computeMatrix and plotProfile, and log2 scale plots were produced using computeMatrix and customized Python scripts. For chrRNA-seq, boxplots were constructed through featureCounts<sup>69</sup> and DESeq2<sup>70</sup> for reads count and normalization, and R scripts were utilized for plotting. Meanwhile, boxplots for ChIP-seq were generated by featureCounts<sup>69</sup> and normalized by library size count via SAMtools<sup>68</sup>. Python scripts are used for data visualization. Violin plots were created using featureCounts<sup>69</sup> for reads count, normalized with scaling factors (or library size for ChIP-seq and chrRNA-seq), and plotted via customized Python scripts.

### Statistical analysis

We used Mann-Whitney test throughout for high-throughput sequencing analyses.

## References

1. Scruggs, B.S. et al. Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol Cell* **58**, 1101-12 (2015).
2. Adelman, K. & Lis, J.T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720-31 (2012).
3. Rhee, H.S. & Pugh, B.F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295-301 (2012).
4. Seila, A.C. et al. Divergent transcription from active promoters. *Science* **322**, 1849-51 (2008).
5. Preker, P. et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851-4 (2008).

6. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-8 (2008).
7. Core, L.J. et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**, 1311-20 (2014).
8. Kim, T.K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-7 (2010).
9. De Santa, F. et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* **8**, e1000384 (2010).
10. Sigova, A.A. et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A* **110**, 2876-81 (2013).
11. Djebali, S. et al. Landscape of transcription in human cells. *Nature* **489**, 101-8 (2012).
12. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. & Sharp, P.A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360-3 (2013).
13. Chiu, A.C. et al. Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. *Mol Cell* **69**, 648-663 e7 (2018).
14. Baillat, D. et al. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265-76 (2005).
15. Welsh, S.A. & Gardini, A. Genomic regulation of transcription and RNA processing by the multitasking Integrator complex. *Nat Rev Mol Cell Biol* **24**, 204-220 (2023).
16. Offley, S.R. et al. A combinatorial approach to uncover an additional Integrator subunit. *Cell Rep* **42**, 112244 (2023).
17. Elrod, N.D. et al. The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. *Mol Cell* **76**, 738-752 e7 (2019).
18. Tatomer, D.C. & Wilusz, J.E. Attenuation of Eukaryotic Protein-Coding Gene Expression via Premature Transcription Termination. *Cold Spring Harb Symp Quant Biol* **84**, 83-93 (2019).
19. Lykke-Andersen, S. et al. Integrator is a genome-wide attenuator of non-productive transcription. *Mol Cell* **81**, 514-529 e6 (2021).
20. Lai, F., Gardini, A., Zhang, A. & Shiekhattar, R. Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399-403 (2015).
21. Rubtsova, M.P. et al. Integrator is a key component of human telomerase RNA biogenesis. *Sci Rep* **9**, 1701 (2019).
22. Barra, J. et al. Integrator restrains paraspeckles assembly by promoting isoform switching of the lncRNA NEAT1. *Sci Adv* **6**, eaaz9072 (2020).
23. Cazalla, D., Xie, M. & Steitz, J.A. A primate herpesvirus uses the integrator complex to generate viral microRNAs. *Mol Cell* **43**, 982-92 (2011).
24. Zheng, H. et al. Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science* **370**(2020).

25. Fianu, I. et al. Structural basis of Integrator-mediated transcription regulation. *Science* **374**, 883-887 (2021).
26. Stein, C.B. et al. Integrator endonuclease drives promoter-proximal termination at all RNA polymerase II-transcribed loci. *Mol Cell* **82**, 4232-4245 e11 (2022).
27. Huang, K.L. et al. Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. *Mol Cell* **80**, 345-358 e9 (2020).
28. Vervoort, S.J. et al. The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell* **184**, 3143-3162 e32 (2021).
29. Hu, S. et al. INTAC endonuclease and phosphatase modules differentially regulate transcription by RNA polymerase II. *Mol Cell* (2023).
30. Beckedorff, F. et al. The Human Integrator Complex Facilitates Transcriptional Elongation by Endonucleolytic Cleavage of Nascent Transcripts. *Cell Rep* **32**, 107917 (2020).
31. Gardini, A. et al. Integrator regulates transcriptional initiation and pause release following activation. *Mol Cell* **56**, 128-139 (2014).
32. Yue, J. et al. Integrator orchestrates RAS/ERK1/2 signaling transcriptional programs. *Genes Dev* **31**, 1809-1820 (2017).
33. Stadelmayer, B. et al. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat Commun* **5**, 5531 (2014).
34. Skaar, J.R. et al. The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res* **25**, 288-305 (2015).
35. Dasilva, L.F. et al. Integrator enforces the fidelity of transcriptional termination at protein-coding genes. *Sci Adv* **7**, eabe3393 (2021).
36. Shersher, E. et al. NACK and INTEGRATOR act coordinately to activate Notch-mediated transcription in tumorigenesis. *Cell Commun Signal* **19**, 96 (2021).
37. Jaeger, M.G. & Winter, G.E. Fast-acting chemical tools to delineate causality in transcriptional control. *Mol Cell* **81**, 1617-1630 (2021).
38. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M.T. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep* **15**, 210-218 (2016).
39. Yesbolatova, A. et al. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat Commun* **11**, 5701 (2020).
40. Liu, X. et al. The PAF1 complex promotes 3' processing of pervasive transcripts. *Cell Rep* **38**, 110519 (2022).
41. Nojima, T. et al. Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Mol Cell* **72**, 970-984 e7 (2018).
42. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**, 917-22 (2009).

43. Bhatt, D.M. et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279-90 (2012).
44. Wu, Y., Albrecht, T.R., Baillat, D., Wagner, E.J. & Tong, L. Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc Natl Acad Sci U S A* **114**, 4394-4399 (2017).
45. Lourenco, C. et al. MYC protein interactors in gene transcription and cancer. *Nat Rev Cancer* **21**, 579-591 (2021).
46. Lawson, K.A. et al. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* **13**, 424-36 (1999).
47. Duttke, S.H.C. et al. Human promoters are intrinsically directional. *Mol Cell* **57**, 674-684 (2015).
48. Marquardt, S. et al. A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* **157**, 1712-23 (2014).
49. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225-8 (2016).
50. Phatnani, H.P. & Greenleaf, A.L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20**, 2922-36 (2006).
51. Hsin, J.P., Li, W., Hoque, M., Tian, B. & Manley, J.L. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *Elife* **3**, e02112 (2014).
52. Descostes, N. et al. Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife* **3**, e02105 (2014).
53. Lai, F., Damle, S.S., Ling, K.K. & Rigo, F. Directed RNase H Cleavage of Nascent Transcripts Causes Transcription Termination. *Mol Cell* **77**, 1032-1043.e4 (2020).
54. Bennett, C.F. & Swayze, E.E. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu Rev Pharmacol Toxicol* **50**, 259-93 (2010).
55. Roberts, T.C., Langer, R. & Wood, M.J.A. Advances in oligonucleotide drug delivery. *Nat Rev Drug Discov* **19**, 673-694 (2020).
56. Colgan, D.F. & Manley, J.L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11**, 2755-66 (1997).
57. Ryan, K., Calvo, O. & Manley, J.L. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* **10**, 565-73 (2004).
58. Davidson, L. et al. Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. *Cell Rep* **26**, 2779-2791 e5 (2019).
59. Mimoso, C.A. & Adelman, K. U1 snRNP increases RNA Pol II elongation rate to enable synthesis of long genes. *Mol Cell* **83**, 1264-1279 e10 (2023).
60. Wang, H. et al. H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature* **615**, 339-348 (2023).

61. Kaida, D. et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664-8 (2010).
62. Berg, M.G. et al. U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53-64 (2012).
63. So, B.R. et al. A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls Telescripting, Regulating mRNA Transcription in Human Cells. *Mol Cell* **76**, 590-599 e4 (2019).
64. Shuai, S. et al. The U1 spliceosomal RNA is recurrently mutated in multiple cancers. *Nature* **574**, 712-716 (2019).
65. Suzuki, H. et al. Recurrent noncoding U1 snRNA mutations drive cryptic splicing in SHH medulloblastoma. *Nature* **574**, 707-711 (2019).
66. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
68. Li, H. et al. The sequence alignment/map format and SAMtools. *bioinformatics* **25**, 2078-2079 (2009).
69. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general-purpose read summarization program. *Bioinformatics* **30**, 923-930 (2014).
70. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 1-21 (2014).
71. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-W165 (2016).
72. Robinson, J.T. et al. Integrative genomics viewer. *Nature biotechnology* **29**, 24-26 (2011).
73. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, 1-9 (2008).
74. Gregersen, L.H., Mitter, R. & Svejstrup, J.Q. Using TTchem-seq for profiling nascent transcription and measuring transcript elongation. *Nature Protocols* **15**, 604-627 (2020).
75. Mata, J. & Wise, J.A. 4-Thiouridine Labeling to Analyze mRNA Turnover in *Schizosaccharomyces pombe*. *Cold Spring Harb Protoc* **2017**(2017).
76. Roberts, T.C., Langer, R. & Wood, M.J.A. Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery* **19**, 673-694 (2020).
77. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-334 (2013).
78. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013).
79. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
80. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. & Sharp, P.A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360-363 (2013).

81. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018 (2011).
82. Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. in *Proceedings of the seventh annual international conference on Research in computational molecular biology* 322-331 (2003).

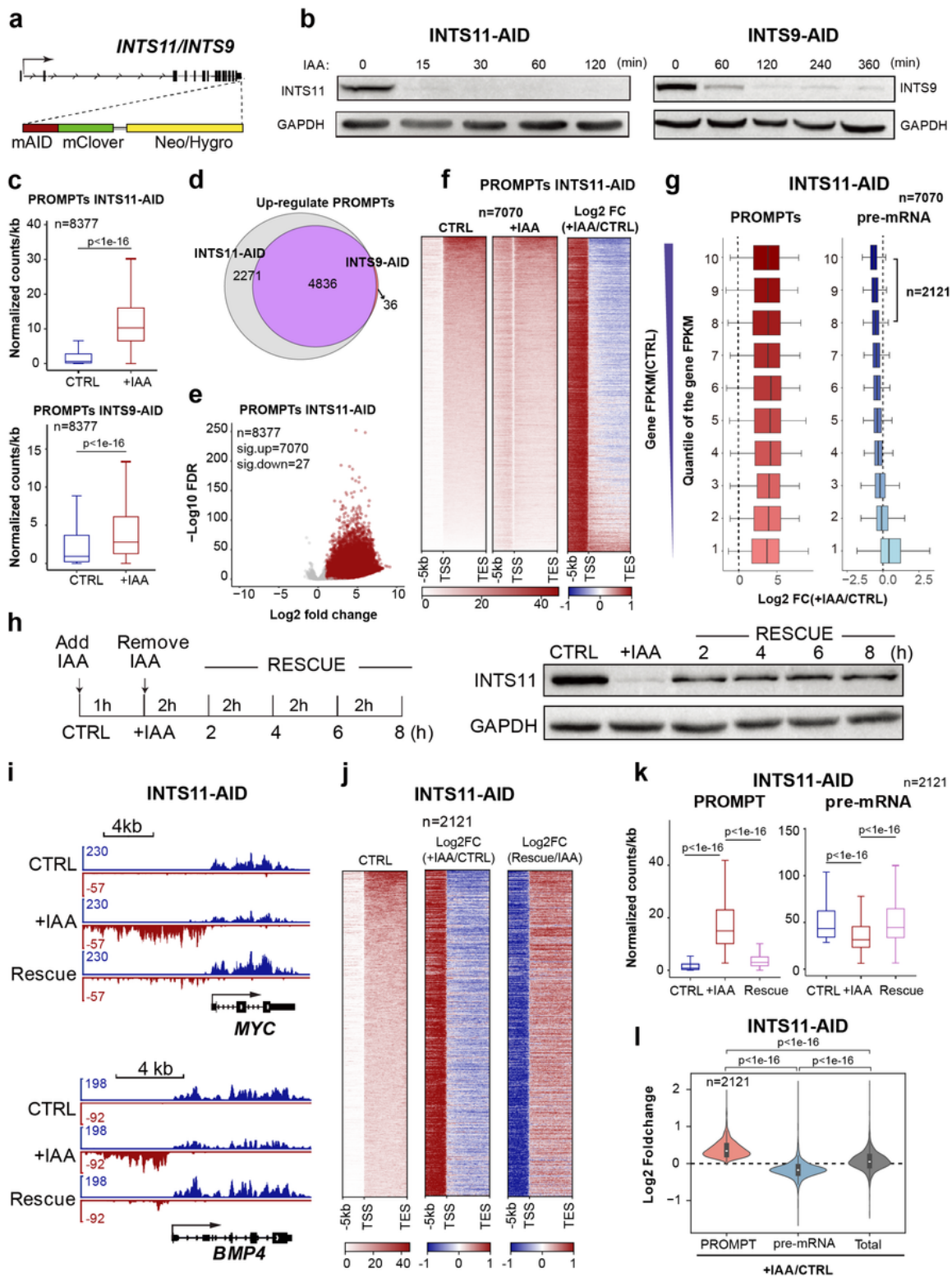
## Supplementary Tables

Supplementary Table 1 is not available with this version

Supplementary Table 1: List of Plasmids, Cell lines, Antibodies and primers in this study.

## Figures



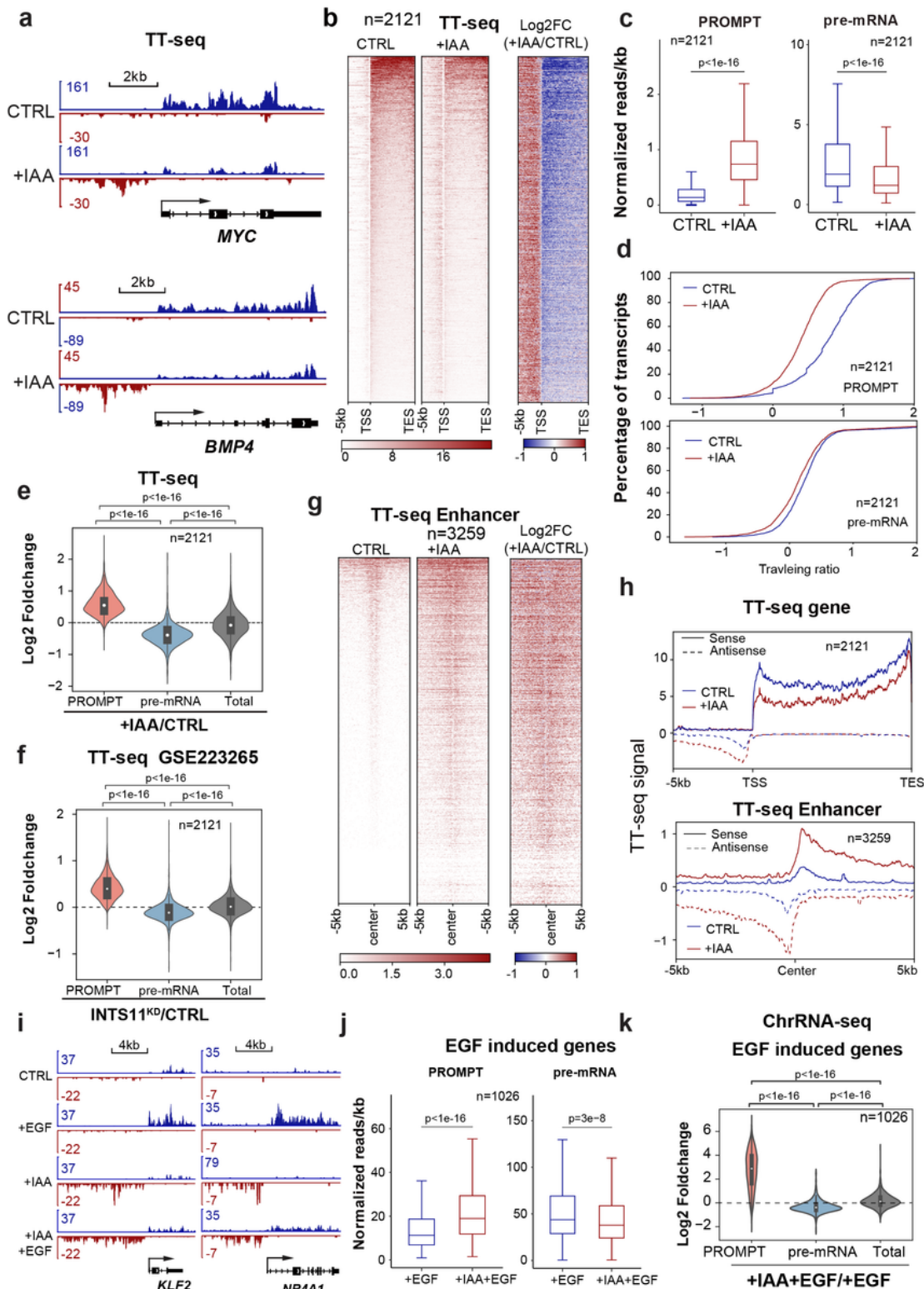


**Figure 1**

**Rapid degradation of Integrator leads massive accumulation of PROMPTs.**

**a**, Schematic diagram showing the generation of INTS11-AID and INTS9-AID2 cell lines. **b**, Western blot representing the INTS11 depletion (left) with the time course after 500  $\mu$ M IAA treatment, and the same of INTS9 with 10  $\mu$ M 5-Ph-IAA treatment (right). **c**, Box plots of induced expression of 8377 PROMPTs after

the INTS11 depletion (top) or INTS9 depletion (bottom).  $p < 1.0E-16$ . **d**, Venn diagram showing the intersection of up-regulate (FC>2) PROMPTs after INTS11 and INTS9 depletion. **e**, Volcano plot showing the log<sub>2</sub> fold change of PROMPTs before and after INTS11 depletion. **f**, Heatmap representing 7,070 genes and their PROMPTs from chrRNA-seq with an overlaying heatmap of log<sub>2</sub> (+IAA / CTRL) at the right panel. Up-regulation shown in red, and down-regulation shown in blue. **g**, 7,070 genes are ranked by gene expression level (FPKM) and divided into 10 quantiles in control samples. Box plots representing the log<sub>2</sub> (+IAA / CTRL) fold change of each quantile in PROMPTs (left) and pre-mRNA (right). The top 3 quantiles with 2121 genes have been selected as the highly expressed genes. **h**, Flow diagram of the time course for the rescue experiments (left). INTS11 western blot shown on the right. **i**, IGV tracks of *MYC* and *BMP4* gene loci based on ChrRNA-seq results in INTS11-AID cells. **j**, Heatmap representing 2121 active genes and their PROMPTs from chrRNA-seq with an overlaid heatmap of log<sub>2</sub> (+IAA / CTRL) or log<sub>2</sub> (Rescue / +IAA) at the right panel. The rescue time is 7 hours. **k**, Boxplots representing the transcript levels from PROMPT (left) and pre-mRNA (right) for 2121 active genes. **l**, Violin plot showing the changes at PROMPT and gene regions before and after INTS11 depletion. The total represents the combined signals of pre-mRNA and PROMPT for 2121 active genes.  $p < 1.0E-16$ . Mann-Whitney test was used for all statistical analyses.

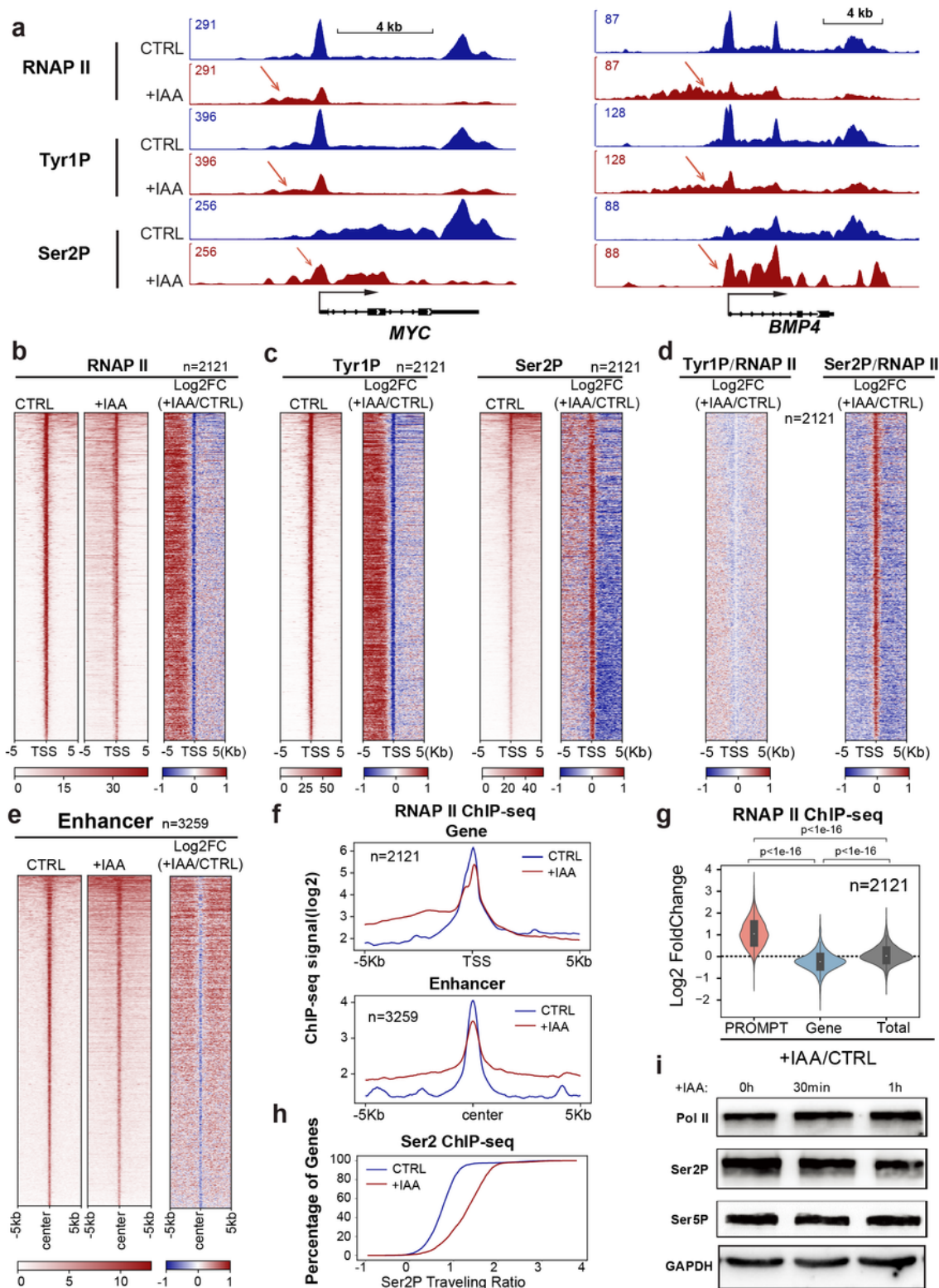


**Figure 2**

**Integrator depletion alters transcription dynamics in the sense and antisense direction.**

**a**, IGV tracks of MYC and BMP4 gene loci based on TT-seq results in INTS11-AID cells. **b**, Heatmap representing 2121 active genes and their PROMPTs from TT-seq (left). Overlying heatmap of log<sub>2</sub> (+IAA / CTRL) (right). (Up-regulation showing in red, down-regulation showing in blue). **c**, Boxplots representing

the transcript levels from PROMPT (left) and pre-mRNA (right) for 2121 active genes  $P < 1.0E-16$ . **d**, TT-seq traveling ratio representing the extension of PROMPT and the ineffective elongation of pre-mRNA. **e**, Violin plot showing the transcription dynamics at PROMPT and gene regions before and after INTS11 depletion. The total represents the combined signals of pre-mRNA and PROMPT for 2121 active genes. **f**, Violin plot showing the transcription dynamics at PROMPT and gene regions before and after INTS11 depletion. The total represents the combined signals of pre-mRNA and PROMPT for 2121 active genes. Data from GSE223265. **g**, Heatmap representing 3259 enhancers from TT-seq(left). Overlaying heatmap of  $\log_2$  (+IAA / CTRL) (right). **h**, Metagene analysis representing novel transcripts at PROMPT regions and reduced transcription at gene bodies for 2121 active genes. **i**,IGV tracks of KLF2 and NR4A1 gene loci before and after EGF induction. **j**,Quantitative boxplots representing the transcript levels of PROMPT (left) and pre-mRNA (right) of 1026 EGF-induced genes. Mann-Whitney test was used for all statistical analyses. **k**,Violin plot showing the responses of 1026 EGF-induced.



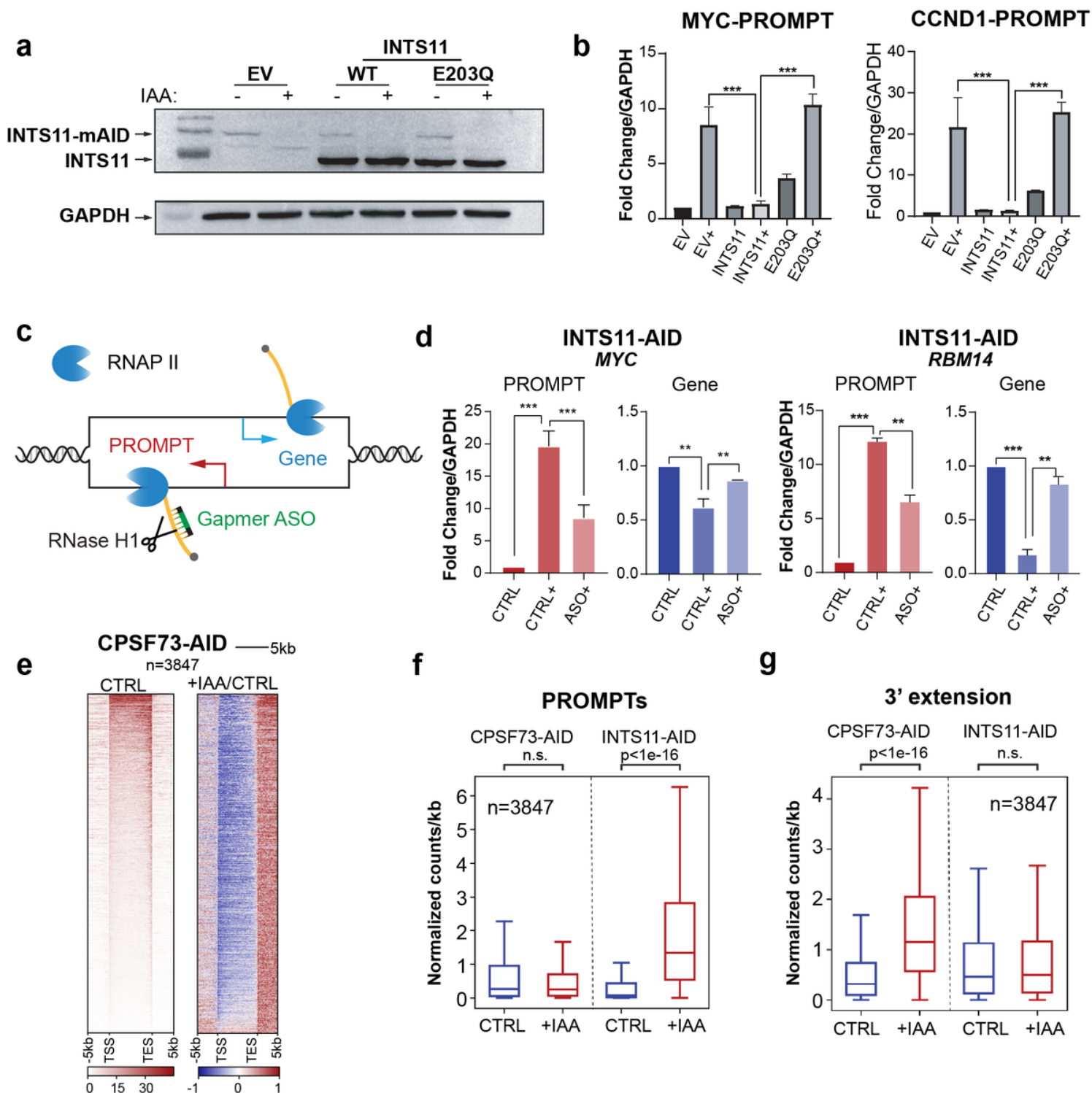
**Figure 3**

**Integrator depletion alters the RNAPII dynamics.**

**a**, The IGV tracks of total RNAPII ChIP-seq, Tyr1P ChIP-seq, and Ser2P ChIP-seq at *MYC* (Left) or *BMP4* (Right) loci before and after INTS11 depletion (Control shown in blue and IAA treatment shown in red). **b**, Heatmap of total RNAPII occupancy in 5 kb around TSS sites of 2121 genes. (Right panel, overlay log2

(+IAA / CTRL)). **c**, The same heatmaps of Tyr1P and Ser2P occupancies. **d**, Heatmap of Tyr1P / total RNAPII (left) and Ser2P / total RNAPII (right) for the change of  $\log_2$  (+IAA / CTRL) around TSS sites. **e**, Heatmap of RNAPII ChIP-seq (left) and  $\log_2$  (+IAA / CTRL) (right) around 3259 enhancers. **f**, Metagene analysis of the occupancies of RNAPII ChIP-seq signal near TSS sites (top) and enhancers (bottom). **g**, Comparison of RNAPII changes at PROMPT or gene regions before and after INTS11 depletion. The total represents the combined signals of pre-mRNA and PROMPT for 2121 active genes. Mann-whitney test was used for statistical analyses. **h**, Ser2P traveling ratio at the initiation stage of transcription. **i**, Western blot of total RNAP II and different CTD phosphorylation states after the INTS11 depletion in the nucleus.



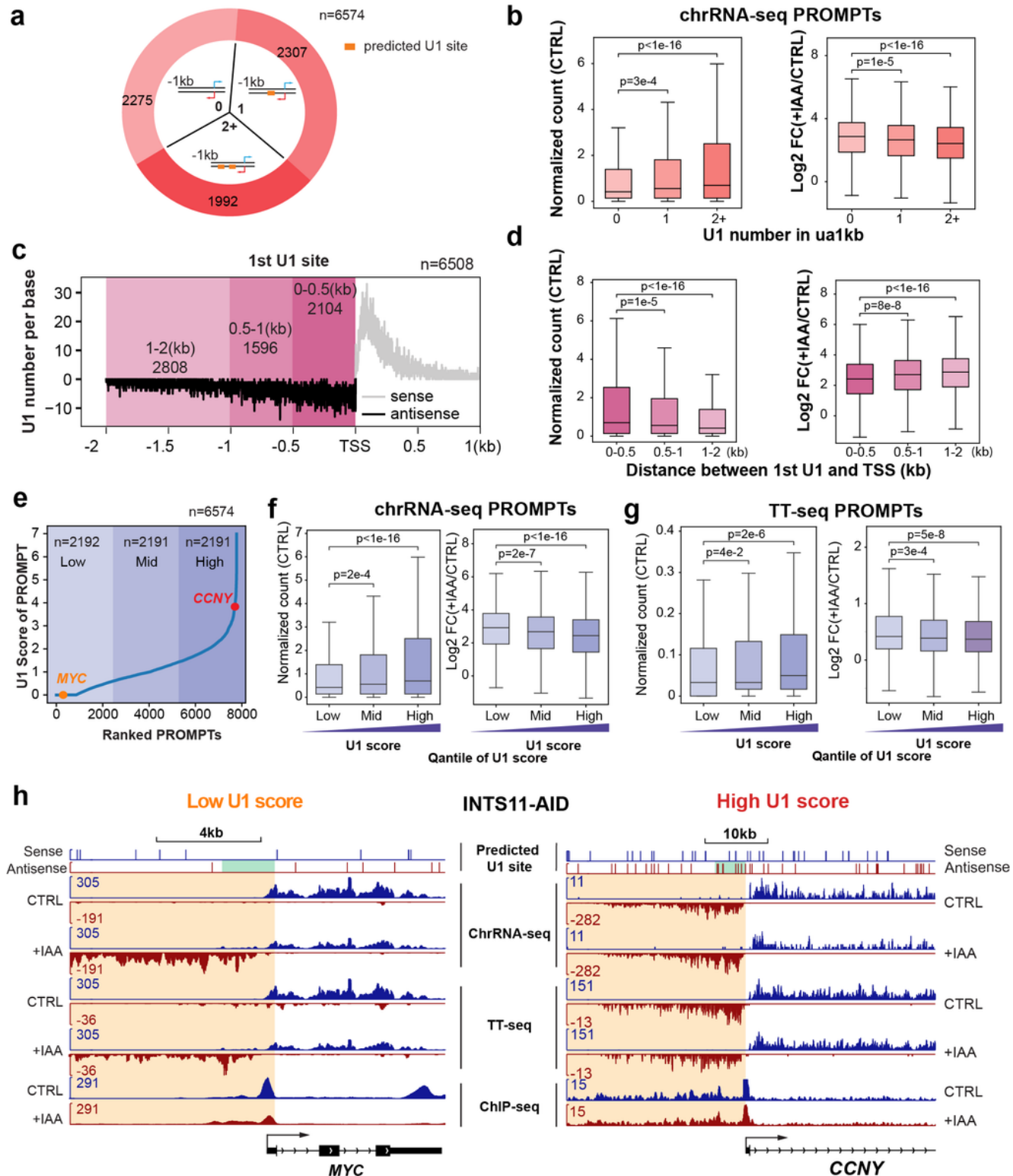


**Figure 4**

**Endonuclease function of Integrator is specifically required for transcription directionality**

**a**, Ectopic expression of INTS11 or the E203Q mutant in the INTS11-AID cells. EV means empty vector. "+" stands for the IAA treatments. **b**, Quantitative RT-PCR of PROMPTs accumulation at *CCND1* or *MYC* loci through the ectopic expression of INTS11 or the E203Q mutant. **c**, Schematic diagram showing targeted degradation of PROMPT by gapmer ASO through RNase H1 mechanism. **d**, Quantitative RT-PCR of

PROMPT and mRNA changes upon the gapmer ASO treatments targeting PROMPTs of *MYC* (left) or *RBM14* (right). **e**, Heatmap of the 3' accumulation after CPSF73 depletion. **f, g**, Boxplot showing PROMPTs **f** and gene 3' extension **g** for CPSF73 or INTS11 depletion for 3847 gene loci.  $p < 1.0E-16$ .  $p$  value was calculated with the Mann-Whitney test. n.s. not significant. For **b** and **d**, unpaired two-tailed student's  $t$ -test was used.  $p < 0.05$ . (\*\*) $p < 0.01$ . (\*\*\*) $p < 0.001$ . Error bars represent s.e.m. (n = 3 biologically independent experiments).

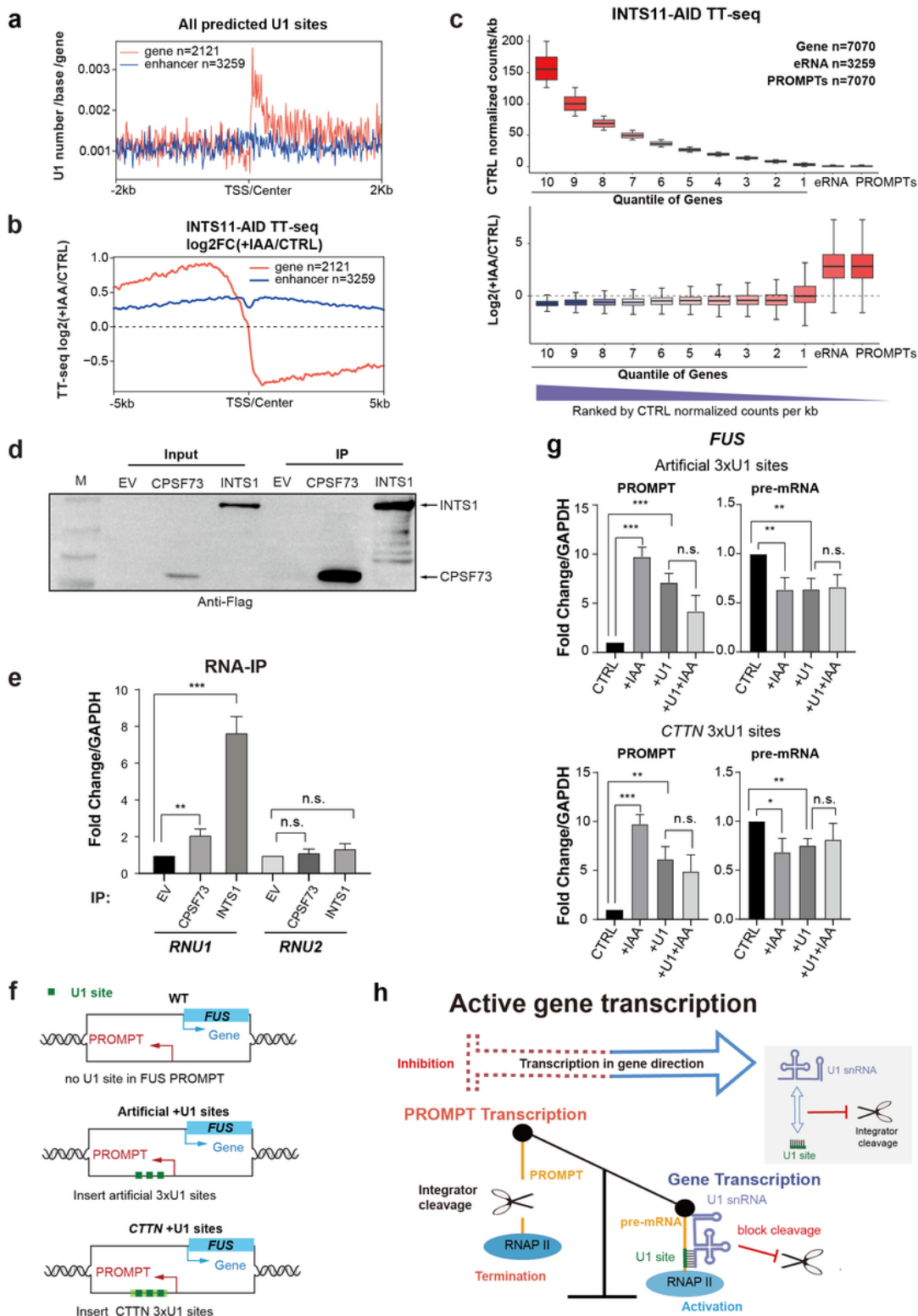




## Figure 5

### Dynamics of transcription directionality depends on the distribution of U1 sites and Integrator cleavage.

**a**, Diagram of predicted U1 sites in 1 kb window upstream of TSSs. **b**, Quantitative boxplots of PROMPT distribution with the number of U1 binding sites defined in **a**. (Left, reads count from CTRL sample; Right, Fold change after INTS11 depletion). **c**, Diagram of 1<sup>st</sup> U1 site in 2 kb window upstream of TSSs. **d**, Analysis of PROMPT distribution by 1<sup>st</sup> U1 sites at PROMPT region. The detailed definition of the x-axis is in Extended Data Fig. 9b. (Left, reads count from CTRL sample; Right, Fold change after INTS11 depletion). **e**, The rank of U1 score at PROMPT region (details in the Methods). **f**, Quantitative boxplot of PROMPTs categorized by the U1 score defined in **e**. Left, reads count from CTRL sample. Right, PROMPT fold change after INTS11 depletion. **g**, Same classification with TT-seq results. Mann-whitney test was used for statistical analyses. **h**, IGV examples of *MYC* (low U1 score) and *CCNY* (high U1 score) loci from chrRNA-seq, TT-seq and RNAPII ChIP-seq results. The predicted strand-specific U1 sites are marked on the top.



**Figure 6**

**Integrator together with U1 snRNA orchestrates transcription directionality.**

**a**, Metagene analysis of all predicted U1 sites centered with TSS sites of genes or enhancers. **b**, Metagene analysis of TT-seq log<sub>2</sub> fold change (+IAA/CTRL) for gene TSS sites or center of enhancers, only downstream of TSS show a decrease in transcription. **c**, 7070 genes are ranked by TT-seq results in CTRL

samples and divided into 10 quantiles with the normalized counts for 3259 eRNAs and 7070 PROMPTs (top). Box plots representing the  $\log_2$  (+IAA / CTRL) fold change of each quantile shown above. **d**, Ectopic expression of Flag-tagged CPSF73 or INTS1 to detect their associations with U1 snRNA. **e**, Quantitative RT-PCR of *RNU1* or *RNU2* transcripts associated with ectopically expressed CPSF73 or INTS1. **f**, Schematic diagrams of inserting artificial 3 U1 sequence (middle) or 3 x U1 sequence from CTTN gene (bottom) in a U1-free PROMPT region of FUS gene. **g**, Quantitative RT-PCR indicating the changes of PROMPT and pre-mRNA at the FUS gene locus inserted with 3XU1 sequence (left) or with CTTN endogenous 3 x U1 sequence (right). “+u1” for inserting the U1 sequence. *p*-value was calculated with unpaired two-tailed student's t-test. *n.s.* not significant. (\*)  $p < 0.05$ . (\*\*)  $p < 0.01$ . (\*\*\*)  $p < 0.001$ . Error bars represent s.e.m. (n = 3 biologically independent experiments). **h**, Model of U1-Integrator axis to control transcription directionality for enhancer RNA transcription (left) and active gene transcription (right).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Extendeddatafigures.docx](#)