Human promoter directionality is determined by transcriptional initiation and the opposing activities of INTS11 and CDK9

Joshua D Eaton[1,2,3], Jessica Board[1,3], Lee Davidson[1], Chris Estell[1] and Steven West[1*]

[1]The Living Systems Institute, University of Exeter, Stocker Rd, Exeter, EX4 4QD, UK

[2]Current address: MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Ave, Trumpington, Cambridge CB2 0QH, UK

[3]Equal contribution

*Corresponding author: s.west@exeter.ac.uk; +44(0)1392727458

ABSTRACT

RNA polymerase II (RNAPII) transcription initiates bidirectionally at many human protein-coding genes. Sense transcription usually dominates and leads to messenger RNA production, whereas antisense transcription rapidly terminates. The basis for this directionality is not fully understood. Here, we show that sense transcriptional initiation is more efficient than in the antisense direction, which provides initial directionality. After transcription begins, the opposing functions of Integrator (INTS11) and cyclin-dependent kinase 9 (CDK9) maintain directionality. INTS11 terminates antisense transcription, whereas sense transcription is protected from INTS11-dependent attenuation by CDK9. INTS11 terminates transcription in both directions upon CDK9 inhibition, and the engineered recruitment of CDK9 prevents attenuation by INTS11. Therefore, transcriptional initiation and the opposing activities of CDK9 and INTS11 explain the attenuation of antisense transcription, the more extensive nature of sense transcription, and promoter directionality.

61 **INTRODUCTION**

62 In humans, most protein-coding genes initiate transcription bidirectionally. Usually,
63 only sense transcription is extended and leads to mRNA synthesis. Antisense transcription
64 terminates within ~1 kilobase (kb), and the short non-coding (nc)RNA is degraded[1]. Similar
65 asymmetry frequently occurs in unicellular eukaryotes (e.g., budding yeast) and at some plant
66 promoters[2,3]. Thus, promoter directionality is a broadly observed phenomenon. Interestingly,
67 bidirectionality is the ground state of promoters and directionality is acquired over evolutionary
68 time[4]. This is proposed to be through a combination of DNA sequences and proteins that
69 favour the directional initiation and elongation of transcription.

70 The present explanation for the directionality of mammalian RNAPII promoters
71 involves the arrangement of U1 snRNA binding sites and polyadenylation signals (PASs). U1
72 snRNA promotes RNAPII elongation through protein-coding genes by binding to RNA and
73 preventing early termination[5-7]. It does this by silencing PASs and antagonizing other
74 attenuation mechanisms, which include the PP1 regulator, PNUTS, and the restrictor
75 complex[8,9]. In contrast, U1 binding splice sites are rarer in short antisense transcripts, which
76 are often rich in PAS sequences that are proposed to promote transcriptional termination[10].
77 This model predicts that polyadenylation factors control a large fraction of antisense
78 transcriptional termination.

79 The multi-subunit Integrator complex also regulates promoter-proximal transcription[11-
80 15]. The Integrator complex comprises the backbone, arm, phosphatase, and endonuclease
81 modules[11]. Its endoribonuclease is INTS11, and its phosphatase activity is mediated by INTS6
82 and protein phosphatase 2A (PP2A). INTS11 endonuclease broadly affects promoter-
83 proximal transcriptional attenuation whereas INTS phosphatase is proposed to regulate the
84 escape of RNAPII into elongation[16,17]. INTS6/PP2A phosphatase functionally antagonises
85 CDK9, which is vital for RNAPII promoter escape and productive elongation[18]. By this model,
86 CDK9 activity promotes elongation across protein-coding genes and INTS6/PP2A opposes it.

87 We tested the prediction that PAS factors control transcription directionality by
88 terminating antisense transcription, but this is not usually the case. Instead, we find that
89 promoter directionality is conferred by preferential initiation in the sense direction and is
90 thereafter maintained by INTS11 and CDK9. The termination of most antisense transcription
91 is constitutively INTS11-dependent, whereas sense transcription is hypersensitive to INTS11
92 only when CDK9 activity is simultaneously inhibited. We hypothesise that CDK9 activity
93 desensitises sense transcription to INTS11 and that reduced CDK9 activity in the antisense
94 direction exposes RNAPII to INTS11-dependent termination.

95

RESULTS AND DISCUSSION

The present explanation of mammalian promoter directionality invokes early PAS-dependent termination of antisense transcription[10]. This is partly based on the direct detection of polyadenylated antisense transcripts, which provides evidence that some transcriptional termination is PAS-dependent. However, there are non-polyadenylated antisense RNAs that might be attenuated in other ways[19]. To test the relevance of PAS-dependent termination for antisense transcriptional termination and promoter directionality we tagged *RBBP6* with the dTAG degron. Because RBBP6 is required to activate the PAS cleavage machinery[20,21], its depletion should inhibit any PAS-dependent transcriptional termination. Three homozygous *dTAG-RBBP6* clones were isolated (Supplemental Figure 1A) and tagged RBBP6 was efficiently depleted after exposure to the dTAGv-1 degrader (Figure 1A). To test the overall contribution of RBBP6 to transcription, we used POINT (Polymerase Intact Nascent Transcript)-seq[22], which maps nascent transcription by sequencing full-length RNA extracted from immunoprecipitated RNAPII.

Figure 1B shows POINT-seq coverage over *NEDD1* including the upstream antisense transcript. RBBP6 loss causes a termination defect beyond the PAS shown by the extended POINT-seq signal beyond the *NEDD1* gene. However, RBBP6 does not affect the termination of upstream antisense RNA which is not extended when RBBP6 is depleted. Metaplot and heatmap analyses across 1316 genes confirm this as a global trend (Figures 1C and D). Although many antisense transcripts contain multiple AAUAAA sequences, most still terminate RBBP6-independently (Supplementary Figures 1B and C). Similarly, we recently showed that the PAS-dependent 5'→3' exonucleolytic torpedo terminator, XRN2, does not affect antisense transcriptional termination[9]. Therefore, although a fraction of antisense transcription is polyadenylated, most of it can terminate using PAS-independent mechanisms.

Our data argue that PAS-independent termination mechanisms are prevalent for antisense transcription. A major PAS-independent termination pathway is driven by the Integrator complex, which was first identified as the 3' end processing complex for snRNAs[23]. Several reports show that Integrator terminates transcription from most promoters, including those that initiate antisense transcription, so might affect promoter directionality[11,16,17]. To analyse this, we tagged its endonucleolytic subunit (INTS11) with a dTAG degron[24], which enables rapid depletion of its catalytic activity (Figure 2A). We then performed POINT-seq on *INTS11-dTAG* cells depleted or not of INTS11 to assay the global impact of INTS11 on RNAPII transcription. The established effect of INTS11 at snRNAs was detected in our POINT-seq data and demonstrates the efficacy of this approach (Figure 2B).

132

133 As exemplified by *PGK1* (Figure 2C), INTS11 does not affect the termination of protein-
134 coding transcription but its loss causes a strong upregulation of antisense transcription. Meta-
135 analysis shows the generality of antisense transcriptional attenuation via INTS11 (Figure 2D)
136 and heatmap analysis demonstrates the dominant impact of INTS11 on antisense versus
137 sense transcription at most promoters (Figure 2E). Transcription over protein-coding gene
138 bodies is generally more modestly affected by INTS11 loss, but there is evidence of mildly
139 increased transcription over the 5' end of some genes as was recently reported (Supplemental
140 Figure 2A)[12-17]. Overall, INTS11 frequently attenuates antisense transcription whereas a
141 smaller fraction of sense transcription is affected.

142

143 Our analyses show that antisense RNAs are more frequently terminated via INTS11
144 than sense transcripts, which ultimately terminate via the PAS-dependent mechanism. The
145 sensitivity of antisense transcripts to INTS11 might be due to the nature of antisense
146 transcription or some other promoter feature. To interrogate this further, we analysed two
147 other promoter classes: those where protein-coding transcripts are initiated in both directions
148 and those that initiate the bidirectional transcription of unstable enhancer (e)RNAs. In the
149 former case, both transcripts are extended and stable (i.e., like most sense transcripts); in the
150 latter case, both are short and unstable like most antisense transcripts. When both directions
151 are protein-coding, INTS11 depletion causes modest reductions in transcription
152 (Supplemental Figure 2B). Conversely, eRNA transcription was bidirectionally upregulated
153 upon INTS11 elimination like the antisense transcripts in (Supplemental Figure 2C). We
154 conclude that short ncRNAs are more strongly affected by INTS11 than protein-coding
155 transcripts. At directional promoters this results in the attenuation of antisense transcription.

156 The increased antisense POINT-seq signal following INTS11 loss is consistent with
157 defects in transcriptional termination. However, it could result from increased transcriptional
158 initiation. We were also interested in whether preferential sense initiation could contribute to
159 observed directionality as was proposed in budding yeast[4]. To precisely resolve directional
160 aspects of initiation and assay any impact of INTS11, we devised a variant of POINT-seq
161 called short (s)POINT. Briefly, sPOINT follows the POINT-seq protocol but library preparation
162 employs the selective amplification of 5' capped RNAs <150nts (Figure 3A, Supplemental
163 Figure 3A, Experimental Procedures). The sPOINT signal over *PGK1* exemplifies this and
164 demonstrates a very restricted signal close to the TSS (Figure 3B). Figure 3C compares the
165 meta profile of POINT- and sPOINT-seq. While POINT-seq reads span entire transcribed
166 sequences the sPOINT signal is restricted to the promoter-proximal region. Thus, the well-
167 known promoter-proximal RNAPII pause is efficiently assayed by sPOINT-seq. Our metaplot

168 in Figure 3C shows that most capped RNAPII-associated RNA <150nts are promoter-proximal
169 and thus relate to paused polymerases.

170 We then performed sPOINT-seq in *INTS11-dTAG* cells treated or not with dTAGv-1
171 and plotted the coverage of full-length reads within 1kb of TSSs (Figure 3D). Because
172 sPOINT-seq maps the 5' and 3' ends of these reads, it also detects TSSs at single-nucleotide
173 resolution and these 5' ends are plotted in the lower meta profile in Figure 3D. INTS11 loss
174 results in lower sPOINT-seq signal arguing against increased transcriptional initiation in its
175 absence. Together with the increased POINT-seq – especially for antisense transcription –
176 this indicates more RNAPII promoter escape upon INTS11 loss. Interestingly, this experiment
177 reveals sense transcriptional initiation to be more efficient and focused compared to the
178 antisense direction. This is clear from the TSS-aligned metaplot in Figure 3D and the
179 dispersed nature of antisense transcription shown in Figure 3E. Quantitation of the TSS-
180 derived sense vs. antisense reads demonstrates the higher signal in the sense direction in
181 untreated cells (Supplemental Figure 3B). Figure 3F (*PGK1*) and supplemental figure 3C
182 (*ACTB*) exemplify these features on individual genes. INTS11 loss does not dramatically alter
183 TSS position or focus. Although a lower resolution technique, RNAPII chromatin-
184 immunoprecipitation and sequencing (ChIP-seq) confirmed these promoter characteristics
185 and the mild impact of INTS11 (Supplemental Figures 3D and E). Overall, these data show
186 that directionality is partly determined by preferential initiation in the sense direction.

187 After initiation, directionality is maintained because it is well-established that sense
188 transcription goes further than antisense transcription. If INTS11 regulates this, an opposing
189 force is required to prevent it from attenuating sense transcription. INTS11 affects
190 transcription very early, occupies promoters, and becomes less active as RNAPII moves into
191 elongation[16-18,25,26]. Therefore, if INTS11 is counteracted to allow sense transcription, any
192 responsible mechanism needs to act early. One of the first transcriptional checkpoints
193 involves the phosphorylation of RNAPII and other factors by CDK9, which releases promoter-
194 proximally paused RNAPII into elongation[27]. During this process, the Integrator-associated
195 phosphatase, PP2A, antagonizes CDK9 and presumably regulates the sensitivity of RNAPII
196 to INTS11[18]. Because Integrator phosphatase has little effect on antisense transcription[16], we
197 hypothesised that INTS11 sensitivity and CDK9 activity are inversely correlated to maintain
198 directionality after initiation.

199 This hypothesis predicts that sense transcription will be attenuated via INTS11 when
200 CDK9 is inactive. To test this genome-wide, we depleted INTS11 from *INTS11-dTAG* cells in
201 the presence or absence of a specific CDK9 inhibitor (NVP-2[28]) and performed POINT-seq.
202 As exemplified by *CDCA7*, the depletion of INTS11 alone caused an antisense transcriptional

203 termination defect with a milder impact on the protein-coding sense direction (Figure 4A). As
204 expected, NVP-2 treatment reduced transcription over the protein-coding gene body.
205 Antisense transcription also displays some CDK9 sensitivity. Importantly, in NVP-2-treated
206 cells, INTS11 loss increased transcription in both directions - effectively compromising the
207 maintenance of directionality. Further examples are displayed in Supplemental Figures 4A
208 and B. This is a genome-wide trend as shown in the metaplots in Figures 4B and C. Heatmap
209 analysis of the effect of INTS11 loss after CDK9 inhibition shows bidirectional upregulation of
210 transcription at most assayed genes (Figure 4D). This experiment shows a clearer effect of
211 INTS11 loss over the 5' end of sense transcripts versus our experiment in Figure 2 (although
212 this is again more modest than the antisense effect). We speculate this to be due to the
213 slightly extended period of INTS11 depletion employed alongside NVP-2 treatment (2.5 hr
214 here vs. 1.5 hr for Figure 2).

215 Following CDK9 inhibition and INTS11 loss, the largest recovery of protein-coding
216 transcription is often over the 5' end of genes. This presumably reflects a continued
217 requirement of CDK9 for elongation, even when INTS11 is absent. Furthermore, Integrator
218 phosphatase remains RNAPII-associated after acute INTS11 loss[17], which would antagonise
219 elongation especially if CDK9 is lacking. Importantly, the POINT-seq signal remains higher in
220 the sense vs. antisense direction even when CDK9 and INTS11 are both compromised. This
221 is consistent with our sPOINT-seq conclusion that levels of sense transcriptional initiation are
222 higher compared to the antisense direction. Finally, we confirmed the CDK9-mediated
223 suppression of INTS11 on four selected protein-coding genes using an alternative inhibitor, 5,
224 6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) (Supplemental Figure 4C). In sum, most
225 RNAPII promoters are bidirectionally affected by INTS11 loss when CDK9 is inhibited.
226

227 If CDK9 counteracts INTS11, its recruitment should also prevent transcriptional
228 attenuation by Integrator. To assay this, we employed a plasmid where transcription is driven
229 by the human immunodeficiency virus (HIV) promoter. Transcription from the HIV promoter
230 results in the synthesis of the trans-activating response (TAR) element and promoter-proximal
231 RNAPII pausing. Pause release requires the trans-activator of transcription (TAT), which
232 promotes RNAPII elongation by recruiting CDK9[29]. INTS11 suppresses transcription from the
233 HIV promoter when TAT is not present[30]. To test the effect of CDK9 on this process, *INTS11-*
234 *dTAG* cells were transfected with the HIV reporter with or without TAT before treatment or not
235 with dTAGv-1. Transcription from the reporter was analysed by qRT-PCR (Figure 4E). HIV
236 transcription is induced by INTS11 loss in the absence of TAT. Transcription was greatly
237 stimulated by TAT (~200 fold) but was no longer enhanced by INTS11 loss (there was slightly

238 less reporter RNA recovered). Therefore, CDK9 recruitment by TAT alleviates INTS11-
239 dependent attenuation of transcription.

240 How CDK9 opposes INTS11 is unresolved but, of possible relevance, INTS11 and
241 SPT5 are adjacent in the RNAPII: Integrator structure[31]. As SPT5 is a substrate of CDK9[32-34],
242 its phosphorylation might evict INTS11 or prevent its association with the complex. We show
243 that sense transcriptional initiation dominates over the antisense direction. This could reflect
244 the positioning of DNA elements that attract initiation factors to advantage sense
245 transcription[4]. Consistently, mammalian initiation components show peak occupancy over
246 annotated sense TSSs[35]. Thereafter, the maintenance of directionality by CDK9 and INTS11
247 is consistent with their well-established roles in promoting and terminating transcription.
248 Published ChIP-seq indicates their respective enrichment in sense and antisense directions,
249 which supports our model (Supplemental Figure 4D[18]). Nevertheless, some antisense
250 transcription is NVP-2 sensitive, revealing CDK9 activity on a fraction of RNAPII (Figure 4).
251 As CDK9 opposes INTS11, this fraction of antisense transcription may be terminated via other
252 mechanisms like restrictor and the PAS-dependent machinery[9,10,36-38]. Overall, transcriptional
253 initiation and the opposing activities of CDK9 and INTS11 establish and maintain promoter
254 directionality (see model in Figure 4F).

255

256 **ACKNOWLEDGEMENTS**

257 We thank Hiroshi Kimura for the total RNAPII antibody used for POINT-seq. Our
258 research was funded by a Wellcome Trust Investigator Award to SW (223106/Z/21/Z). This
259 project used the University of Exeter Sequencing Service, and their equipment was funded by
260 the Wellcome Trust (Multi-User Equipment Grant award number 218247/Z/19/Z).

261

262 **AUTHOR CONTRIBUTIONS**

263 Conceptualization J.D.E, S.W; data curation J.D.E, J.B, L.D, C.E; formal analysis
264 J.D.E., J.B., L.D; C.E and S.W; methodology, J.D.E. and S.W.; investigation, J.D.E., J.B., L.D;
265 C.E and S.W.; supervision and funding acquisition, S.W.; writing and editing, J.D.E., J.B., L.D;
266 C.E and S.W.

267

268

269 **EXPERIMENTAL PROCEDURES**

270 **Sequencing data**

271 Deposited at Gene Expression Omnibus under accession: GSE243266.

272 **Cloning**

273 HIV reporter constructs were made by removing the CMV promoter, the entire beta-
274 globin sequence, and its PAS from a pcDNA5 FRT/TO plasmid containing the WT β-globin
275 (βWT) gene (described in[39]) and inserting an HIV promoter and downstream TAR element
276 derived from βΔ5-7 (described in[40]). *INTS11* targeting constructs were modified from those
277 we previously described to generate *INTS11-SMASh* cells (described in[41]). The SMASh tag
278 was removed and replaced with 2xHA dTAG derived from Addgene plasmid 91792[24]. Guide
279 RNA expressing Cas9 plasmids to modify *INTS11* or *RBBP6* were made by inserting annealed
280 oligonucleotides, containing the targeting sequence, into px330 digested with BbsI.

281 **Cell culture and cell lines**

282 HCT116 cells were maintained in DMEM supplemented with penicillin/streptomycin at
283 37°C, 5% CO_2. *dTAG-RBBP6* cells were generated using the "CHoP in" protocol[42]. Briefly, a
284 24-well dish was transfected with 250ng of px330[43] containing the *RBBP6*-targeting guide and
285 250ng of PCR product containing the dTAG degron preceded by a blasticidin or puromycin
286 selection marker (derived from Addgene plasmid 91792[24]). Jetprime (Polyplus) was used for
287 transfection. Three days later, cells were passaged into media containing 10µg/ml
288 Blasticidin/1µg/ml Puromycin and colonies were PCR screened ~10 days later. *INTS11-dTAG*
289 cells were generated by homology-directed repair. A 6-well dish of cells was transfected with
290 1µg px330 containing the *INTS11*-targeting guide (described in[41]) and 1µg each of the repair
291 templates. Three days later, cells were passaged into media containing 30µg/ml Hygromycin
292 and 800µg/ml Neomycin (G418). ~10 days later, colonies were picked and screened by PCR.
293 For protein depletion 1µM dTAGv-1 (Tocris) was used for 1-14hrs (see figure legends for
294 timings used in each experiment). NVP-2 was used at 250nM for 2.5 hrs. DRB was used at
295 100µM for 2.5 hr.

296 **POINT-seq**

297 For POINT-seq, we followed the protocol provided in[22]. The only modification was that
298 we started with a confluent 10cm dish of cells and performed the immunoprecipitation with
299 6µg of anti-RNAPII. ~2% cell volume of Drosophila S2 cells was included as a spike in control.
300 Libraries were prepared using the NEBNext Ultra™ II Directional RNA Library Prep Kit for
301 Illumina (New England Biolabs). sPOINT was performed in the same manner except
302 immunoprecipitated RNA was treated with Terminator™ 5-Phosphate-Dependent

303 Exonuclease (lucigen) to remove any uncapped transcripts and libraries were prepared with
304 the SMARTer® smRNA-Seq Kit for Illumina (Takara Bio).

305 **ChIP-seq**

306 For each experiment, 1x10cm dish of cells was used. Protein: DNA crosslinks were
307 formed by adding Formaldehyde (1% v/v) to culture media for 10 mins then quenching with
308 125mM glycine. Cells were rinsed 2x with PBS, scraped off the dish, and pelleted in 10ml
309 PBS at 500xg for 5 mins. We then employed the simple ChIP enzymatic kit (Cell Signalling
310 Technologies) to fragment chromatin and purify RNAPII-bound DNA. We followed the kit
311 protocol except for conjugating 5µg of anti-total RNAPII to sheep anti-mouse dynabeads (Life
312 Technologies). Sequencing libraries were generated using the NEBNext® Ultra™ II DNA
313 Library Prep Kit for Illumina.

314 **Total RNA isolation and qRT-PCR**

315 A 24-well dish of cells was transfected with 100ng HIV reporter plasmid and, where
316 co-transfected, 50ng TAT plasmid[44] using Jetprime (Polyplus) following the manufacturer's
317 protocol. Media was refreshed 5 hours post-transfection and dTAGv-1 was added where
318 appropriate. The next day, total RNA was isolated using Trizol (Thermo Fisher) following the
319 manufacturer's protocol. RNA was DNase treated then 1µg was reverse transcribed using
320 Protoscript II reverse transcriptase (New England Biolabs). qPCR was performed using LUNA
321 SYBR green reagent (New England Biolabs) on a Qiagen Rotorgene instrument. Quantitative
322 analysis used the ΔCT method.

323 **Antibodies**

324 INTS11 (Abbexa; abx234340); total RNAPII (gift from Hiroshi Kimura), INTS8
325 (RRID:AB_2683403); RBBP6 (RRID:AB_2621169); SPT5 (RRID:AB_2196394).

326 **qRT-PCR primers and gRNA target sites**

327 See supplemental table 1.

328

329 **BIOINFORMATICS**

330 **sPOINT**

331 For sPOINT TSS metaplots showing full-length and 5' derived coverage, gene lists
332 were determined by selecting principal protein-coding transcript isoforms from gencode v42
333 human annotation – specifically, those containing both "appris_principal_1" and

334 “Ensembl_canonical” labels (15301 in total). Of these, the top 20% expressed (based on -
335 dTAG) were used to generate meta profiles (3060 TSSs in total).

336 **Antisense PAS analysis**

337 Bed files from Figure 1C were edited to map the antisense transcript from each gene
338 by fitting each gene's start and end point to 3kb upstream of their TSS. Sequences underlying
339 these regions were obtained via BEDTools getFASTA (hg38) and consensus PAS (AATAAA)
340 enumerated in R. PAS frequency per transcript was plotted using ggplot2. Heatmaps for
341 transcripts containing > 1 or no PAS motifs were plotted with DeepTools.

342 **POINT-Seq alignment and visualisation**

343 Adapters were removed from raw reads using Trim Galore! and mapped to GRCh38
344 using HISAT2 using default parameters. Additionally, reads were also mapped to Dm6 to
345 quantify unique Drosophila reads in each sample for spike-in. Reads with a MAPQ score ≤30
346 were removed from alignment files using SAMtools[45]. Biological replicates were normalised
347 and merged using SAMtools merge before, spike-in scaled bigwig files using DeepTools were
348 visualised using IGV[46].

349 **ChIP-seq alignment**

350 Adapters were removed from raw reads using Trim Galore! and mapped to GRCh38
351 using HISAT2 with default parameters. Reads were also mapped to Dm6 to identify spike-in
352 signal. Reads with MAPQ score of ≤30 were removed with SAMtools[45]. Replicates were
353 normalised and merged using SAMtools merge.

354 **POINT-seq metaplots and heat maps**

355 A metagene list of 1316 genes was generated containing genes with no overlapping
356 regions within 10kb of any other expressed transcription unit. Split strand metagene plots were
357 produced using Drosophila spike-in normalised sense and antisense (scaled to -1) bigwig
358 coverage files separately with further graphical processing performed in R. For heat maps,
359 computematrix (DeepTools) was used to generate score files from the normalised bigwig files
360 using the 10kb non-overlapping gene list. A log2 ratio (depletion/control) was applied to
361 identify changes in reads. Plots were redrawn in R; parameters used for each heat map are
362 detailed in figure legends.

363 **ChIP-seq datasets**

364 Data used to produce CDK9 and INTS11 metaplot (GSE163804), GSM numbers:
365 INTS11 (GSM4987768) and CDK9 (GSM4987764).

366 **ChIP-seq metaplot**

367 ChIP-seq metaplots were created from bigwig coverage data files using DeepTools
368 with further processing performed within R. For Supplementary Figure 3E, plots used the same
369 gene list described above for sPOINT promoter analyses. For Supplementary Figure 4D, a
370 metagene list was generated from non-overlapping 918 antisense (promoter upstream
371 transcripts) regions determined by de novo transcript assembly of previously published
372 nuclear RNA-seq from cells depleted of the catalytic exosome subunit, DIS3[47].

373

374 **REFERENCES**

375 1. Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S.,
376 Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion
377 reveals transcription upstream of active human promoters. Science *322*, 1851-1854.
378 1164096 [pii]

379 10.1126/science.1164096.

380 2. Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J.,
381 Regnault, B., Devaux, F., Namane, A., Seraphin, B., et al. (2005). Cryptic pol II
382 transcripts are degraded by a nuclear quality control pathway involving a new poly(A)
383 polymerase. Cell *121*, 725-737. 10.1016/j.cell.2005.04.030.

384 3. Thieffry, A., Vigh, M.L., Bornholdt, J., Ivanov, M., Brodersen, P., and Sandelin, A.
385 (2020). Characterization of Arabidopsis thaliana Promoter Bidirectionality and
386 Antisense RNAs by Inactivation of Nuclear RNA Decay Pathways. Plant Cell *32*,
387 1845-1867. 10.1105/tpc.19.00815.

388 4. Jin, Y., Eser, U., Struhl, K., and Churchman, L.S. (2017). The Ground State and
389 Evolution of Promoter Region Directionality. Cell *170*, 889-898 e810.
390 10.1016/j.cell.2017.07.006.

391 5. Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G.
392 (2010). U1 snRNP protects pre-mRNAs from premature cleavage and
393 polyadenylation. Nature *468*, 664-668. nature09479 [pii]

394 10.1038/nature09479.

395 6. Mimoso, C.A., and Adelman, K. (2023). U1 snRNP increases RNA Pol II elongation
396 rate to enable synthesis of long genes. Molecular cell. 10.1016/j.molcel.2023.03.002.

397 7. Chiu, A.C., Suzuki, H.I., Wu, X., Mahat, D.B., Kriz, A.J., and Sharp, P.A. (2018).
398 Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature
399 Polyadenylation Suppressed by U1 snRNP. Molecular cell *69*, 648-663 e647.
400 10.1016/j.molcel.2018.01.006.

401 8. So, B.R., Di, C., Cai, Z., Venters, C.C., Guo, J., Oh, J.M., Arai, C., and Dreyfuss, G.
402 (2019). A Complex of U1 snRNP with Cleavage and Polyadenylation Factors

Controls Telescripting, Regulating mRNA Transcription in Human Cells. Molecular cell *76*, 590-599 e594. 10.1016/j.molcel.2019.08.007.

9. Estell, C., Davidson, L., Eaton, J.D., Kimura, H., Gold, V.A.M., and West, S. (2023). A restrictor complex of ZC3H4, WDR82, and ARS2 integrates with PNUTS to control unproductive transcription. Molecular cell *83*, 2222-2239 e2225. 10.1016/j.molcel.2023.05.029.

10. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360-363. 10.1038/nature12349.

11. Wagner, E.J., Tong, L., and Adelman, K. (2023). Integrator is a global promoter-proximal termination complex. Molecular cell *83*, 416-427. 10.1016/j.molcel.2022.11.012.

12. Lykke-Andersen, S., Zumer, K., Molska, E.S., Rouviere, 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. Molecular cell *81*, 514-529 e516. 10.1016/j.molcel.2020.12.014.

13. 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 reports *32*, 107917. 10.1016/j.celrep.2020.107917.

14. 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 & development *33*, 1525-1538. 10.1101/gad.330167.119.

15. Elrod, N.D., Henriques, 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. Molecular cell *76*, 738-752 e737. 10.1016/j.molcel.2019.10.034.

16. 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. Molecular cell *83*, 1588-1604 e1585. 10.1016/j.molcel.2023.03.022.

17. 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. Molecular cell *82*, 4232-4245 e4211. 10.1016/j.molcel.2022.10.004.

18. 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-

442 Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in
443 cancer. Cell *184*, 3143-3162 e3132. 10.1016/j.cell.2021.04.022.

444 19. Gockert, M., Schmid, M., Jakobsen, L., Jens, M., Andersen, J.S., and Jensen, T.H.
445 (2022). Rapid factor depletion highlights intricacies of nucleoplasmic RNA
446 degradation. Nucleic acids research *50*, 1583-1600. 10.1093/nar/gkac001.

447 20. Schmidt, M., Kluge, F., Sandmeir, F., Kuhn, U., Schafer, P., Tuting, C., Ihling, C.,
448 Conti, E., and Wahle, E. (2022). Reconstitution of 3' end processing of mammalian
449 pre-mRNA reveals a central role of RBBP6. Genes & development *36*, 195-209.
450 10.1101/gad.349217.121.

451 21. Boreikaite, V., Elliott, T.S., Chin, J.W., and Passmore, L.A. (2022). RBBP6 activates
452 the pre-mRNA 3' end processing machinery in humans. Genes & development *36*,
453 210-224. 10.1101/gad.349223.121.

454 22. Sousa-Luis, R., Dujardin, G., Zukher, I., Kimura, H., Weldon, C., Carmo-Fonseca, M.,
455 Proudfoot, N.J., and Nojima, T. (2021). POINT technology illuminates the processing
456 of polymerase-associated intact nascent transcripts. Molecular cell *81*, 1935-1950
457 e1936. 10.1016/j.molcel.2021.02.034.

458 23. Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R.
459 (2005). Integrator, a multiprotein mediator of small nuclear RNA processing,
460 associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265-276.
461 10.1016/j.cell.2005.08.019.

462 24. Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett,
463 A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for
464 immediate and target-specific protein degradation. Nature chemical biology *14*, 431-
465 441. 10.1038/s41589-018-0021-8.

466 25. Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., Chen, X., Li, Y., Liu, W., Li, J., et
467 al. (2020). Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II
468 phosphatase. Science *370*. 10.1126/science.abb5872.

469 26. Ramamurthy, L., Ingledue, T.C., Pilch, D.R., Kay, B.K., and Marzluff, W.F. (1996).
470 Increasing the distance between the snRNA promoter and the 3' box decreases the
471 efficiency of snRNA 3'-end formation. Nucleic acids research *24*, 4525-4534.
472 10.1093/nar/24.22.4525.

473 27. Fujinaga, K., Huang, F., and Peterlin, B.M. (2023). P-TEFb: The master regulator of
474 transcription elongation. Molecular cell *83*, 393-403. 10.1016/j.molcel.2022.12.006.

475 28. Olson, C.M., Jiang, B., Erb, M.A., Liang, Y., Doctor, Z.M., Zhang, Z., Zhang, T.,
476 Kwiatkowski, N., Boukhali, M., Green, J.L., et al. (2018). Pharmacological
477 perturbation of CDK9 using selective CDK9 inhibition or degradation. Nature
478 chemical biology *14*, 163-170. 10.1038/nchembio.2538.

479 29. Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). A novel
480 CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its

481 high-affinity, loop-specific binding to TAR RNA. Cell *92*, 451-462. 10.1016/s0092-
482 8674(00)80939-3.

483 30. Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R.,
484 Sobhian, B., Severac, D., Rialle, S., et al. (2014). Integrator complex regulates
485 NELF-mediated RNA polymerase II pause/release and processivity at coding genes.
486 Nat Commun *5*, 5531. 10.1038/ncomms6531.

487 31. Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H., and Cramer,
488 P. (2021). Structural basis of Integrator-mediated transcription regulation. Science
489 *374*, 883-887. 10.1126/science.abk0154.

490 32. Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H.
491 (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for
492 processive transcription elongation. Molecular cell *21*, 227-237. S1097-
493 2765(05)01812-5 [pii]

494 10.1016/j.molcel.2005.11.024.

495 33. Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan,
496 K., and Bentley, D.L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5
497 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism.
498 Molecular cell *76*, 896-908 e894. 10.1016/j.molcel.2019.09.031.

499 34. Parua, P.K., Booth, G.T., Sanso, M., Benjamin, B., Tanny, J.C., Lis, J.T., and Fisher,
500 R.P. (2018). A Cdk9-PP1 switch regulates the elongation-termination transition of
501 RNA polymerase II. Nature *558*, 460-464. 10.1038/s41586-018-0214-z.

502 35. Petrenko, N., and Struhl, K. (2021). Comparison of transcriptional initiation by RNA
503 polymerase II across eukaryotic species. Elife *10*. 10.7554/eLife.67964.

504 36. Estell, C., Davidson, L., Steketee, P.C., Monier, A., and West, S. (2021). ZC3H4
505 restricts non-coding transcription in human cells. Elife *10*. 10.7554/eLife.67305.

506 37. Rouviere, J.O., Salerno-Kochan, A., Lykke-Andersen, S., Garland, W., Dou, Y.,
507 Rathore, O., Molska, E.S., Wu, G., Schmid, M., Bugai, A., et al. (2023). ARS2
508 instructs early transcription termination-coupled RNA decay by recruiting ZC3H4 to
509 nascent transcripts. Molecular cell *83*, 2240-2257 e2246.
510 10.1016/j.molcel.2023.05.028.

511 38. Austenaa, L.M.I., Piccolo, V., Russo, M., Prosperini, E., Polletti, S., Polizzese, D.,
512 Ghisletti, S., Barozzi, I., Diaferia, G.R., and Natoli, G. (2021). A first exon termination
513 checkpoint preferentially suppresses extragenic transcription. Nat Struct Mol Biol *28*,
514 337-346. 10.1038/s41594-021-00572-y.

515 39. Muniz, L., Davidson, L., and West, S. (2015). Poly(A) Polymerase and the Nuclear
516 Poly(A) Binding Protein, PABPN1, Coordinate the Splicing and Degradation of a
517 Subset of Human Pre-mRNAs. Molecular and cellular biology *35*, 2218-2230.
518 10.1128/MCB.00123-15.

519 40. Dye, M.J., and Proudfoot, N.J. (2001). Multiple transcript cleavage precedes
520 polymerase release in termination by RNA polymerase II. Cell *105*, 669-681. S0092-
521 8674(01)00372-5 [pii].

522 41. Davidson, L., Francis, L., Eaton, J.D., and West, S. (2020). Integrator-Dependent and
523 Allosteric/Intrinsic Mechanisms Ensure Efficient Termination of snRNA Transcription.
524 Cell reports *33*, 108319. 10.1016/j.celrep.2020.108319.

525 42. Manna, P.T., Davis, L.J., and Robinson, M.S. (2019). Fast and cloning-free
526 CRISPR/Cas9-mediated genomic editing in mammalian cells. Traffic *20*, 974-982.
527 10.1111/tra.12696.

528 43. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X.,
529 Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using
530 CRISPR/Cas systems. Science *339*, 819-823. science.1231143 [pii]

531 10.1126/science.1231143.

532 44. Adams, S.E., Johnson, I.D., Braddock, M., Kingsman, A.J., Kingsman, S.M., and
533 Edwards, R.M. (1988). Synthesis of a gene for the HIV transactivator protein TAT by
534 a novel single stranded approach involving in vivo gap repair. Nucleic acids research
535 *16*, 4287-4298.

536 45. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
537 Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The
538 Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
539 10.1093/bioinformatics/btp352.

540 46. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools:
541 a flexible platform for exploring deep-sequencing data. Nucleic acids research *42*,
542 W187-191. 10.1093/nar/gku365.

543 47. Davidson, L., Francis, L., Cordiner, R.A., Eaton, J.D., Estell, C., Macias, S., Caceres,
544 J.F., and West, S. (2019). Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the
545 Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and
546 Transcriptional Control. Cell reports *26*, 2779-2791 e2775.
547 10.1016/j.celrep.2019.02.012.

548

549 **FIGURE LEGENDS**

550 **FIGURE 1: RBBP6 loss disrupts PAS-dependent termination of sense transcription.**

551 **A.** Western blot demonstrating the depletion of dTAG-RBBP6 over a time course of
552 dTAGv-1 addition. SPT5 serves as a loading control.

553 **B.** Genome browser track of *NEDD1* in POINT-seq data from *dTAG-RBBP6* cells treated
554 or not (2hr) with dTAGv-1. RBBP6 depletion induces a termination defect in the
555 protein-coding direction (downstream of the indicated PAS) but not the upstream

556 antisense direction. The Y-axis shows Reads Per Kilobase per Million mapped reads
557 (RPKM).
558 **C.** Metaplot of POINT-seq data from *dTAG-RBBP6* cells treated (RBBP6 dep) or not (Ctrl)
559 (2hr) with dTAGv-1. This shows 1316 protein-coding genes selected as separated
560 from any expressed transcription unit by ≥10kb. Signals above and below the x-axis
561 are sense and antisense reads, respectively. The Y-axis scale is RPKM.
562 TSS=transcription start site; TES=transcription end site (this marks the PAS position).
563 **D.** Heatmap representation of the data in C, which displays signal as a log2 fold change
564 (log2FC) in RBBP6 depleted versus un-depleted conditions.

565 **FIGURE 2: INTS11 loss disrupts the termination of antisense transcription.**

566 **A.** Western blot demonstrating homozygous tagging of *INTS11* with dTAG and the
567 depletion of INTS11-dTAG after 1.5 hr treatment with dTAGv-1. INTS8 serves as a
568 loading control.
569 **B.** Genome browser track showing POINT-seq signal over *RNU5A-1* and *RNU5B-1* in
570 *INTS11-dTAG* cells treated or not (1.5 hr) with dTAGv-1. Note, that INTS11 depletion
571 induces a termination defect in each case. Y-axis shows RPKM following spike-in
572 normalisation.
573 **C.** Genome browser track showing POINT-seq signal over *PGK1* and its upstream
574 antisense region in *INTS11-dTAG* cells treated or not (1.5 hr) with dTAGv-1. Y-axis
575 shows RPKM following spike-in normalisation.
576 **D.** Metaplot of POINT-seq data from *INTS11-dTAG* cells treated or not (1.5 hr) with
577 dTAGv-1. This shows 1316 expressed protein-coding genes that are separated from
578 any expressed transcription unit by ≥10kb. Signals above and below the x-axis are
579 sense and antisense reads, respectively. Y-axis shows RPKM following spike-in
580 normalisation.
581 **E.** Heatmap representation of the data in D, which displays signal as a log2 fold change
582 (log2FC) in INTS11 depleted versus undepleted conditions over a region 3kb upstream
583 and downstream of annotated TSSs.

584 **FIGURE 3: Transcription initiation is more efficient and focused in the sense direction.**

585 **A.** Schematic of sPOINT-seq protocol. The POINT-seq protocol is followed, in which
586 chromatin is isolated and engaged RNAPII is immunoprecipitated. Short transcripts
587 are preferentially amplified during library preparation (see Experimental Procedures
588 for full details).
589 **B.** Comparison of POINT-(top trace) and sPOINT-seq (lower trace) on *PGK1*. Y-axis
590 units are RPKM.

591 **C.** Metaplot comparison of POINT- (top plot) and sPOINT-seq (lower plot) profiles across
592 the 684 highest expressed protein-coding that are separated from expressed
593 transcription units by ≥10kb.. Signals above and below the x-axis are sense and
594 antisense reads, respectively. Y-axis shows RPKM following spike-in normalisation.
595 **D.** Top metaplot shows full read coverage for sPOINT-seq performed in *INTS11-dTAG*
596 cells treated or not (1.5hr) with dTAGv-1. The lower metaplot is the same data but
597 only the 5' end of each read is plotted. Both plots display a region +/-1kb from the TSS
598 of the top 20% (based on sPOINT signal) promoters. Y-axis signals are RPKM
599 following spike-in normalisation.
600 **E.** Metaplot zoom of the antisense TSS signals deriving from the lower plot in D. This
601 makes clear the dispersed sites of initiation. The Y-axis scale is RPKM following spike-
602 in normalisation.
603 **F.** Genome browser track of *PGK1* promoter region in sPOINT-seq. This showcases the
604 focused sense TSS (black arrows) and the dispersed antisense reads (brackets)
605 generalised in the metaplots in D and E. Note the higher y-axis scale (RPKM) for
606 sense vs. antisense.

607 **FIGURE 4: Promoters lose their directionality when CDK9 is inhibited.**

608 **A.** Genome browser track of *CDCA7* in POINT-seq data derived from *INTS11-dTAG* cells
609 either untreated, dTAG treated, NVP-2 treated or dTAG and NVP-2 treated (2.5 hr).
610 Signals above and below the Y-axis are sense and antisense reads, respectively. The
611 Y-axis scale shows RPKM following spike-in normalisation.
612 **B.** Metaplot of POINT-seq analysis of POINT-seq in *INTS11-dTAG* cells depleted or not
613 of INTS11 and treated or not with NVP-2 (2.5 hr). This shows 1316 protein-coding
614 genes selected as separated from any expressed transcription unit by ≥10kb. The
615 regions 3kb upstream and downstream of genes are included. Y-axis units are RPKM
616 following spike-in normalisation.
617 **C.** Metaplot of the same CDK9i + and -dTAG data shown in B but zoomed into the region
618 3kb upstream and downstream of the TSS.
619 **D.** Heatmap representation of the data in C, which displays signal as a log2 fold change
620 (log2FC) in INTS11 depleted versus un-depleted conditions covering a region 3kb
621 upstream and downstream of the TSS.
622 **E.** qRT-PCR analysis of *INTS11-dTAG* cells transfected with the HIV reporter construct
623 with or without TAT then depleted or not of INTS11. Quantitation shows signals
624 relative to those obtained in the presence of INTS11 and the absence of TAT after
625 normalising to MALAT1 RNA. n=3. Error bars show standard deviation. ** denoted

626 p=0.01. Note that INTS11 depletion was performed concurrently with transection (14hr
627 in total).
628 **F.** Model for promoter directionality depicting higher levels of focused transcriptional
629 initiation in the sense direction together with opposing gradients of CDK9 and INTS11
630 activity that peak in sense and antisense directions, respectively.

631

632 **SUPPLEMENTAL FIGURE LEGENDS**

633 **SUPPLEMENTAL FIGURE 1, related to figure 1.**

634 **A.** Western blot showing three separate *dTAG-RBBP6* cell clones. In each case,
635 homozygous tagging is demonstrated by the size shift versus endogenous RBBP6
636 (HCT116 lanes). Tagged RBBP6 is completely depleted by 2hr treatment with dTAGv-
637 1 whereas endogenous RBBP6 is unaffected. SPT5 serves as a loading control.
638 Clone number 2 (red asterisk) was selected for the experiments in Figure 1.
639 **B.** Graph plotting the number of AAUAAA sequences in antisense transcripts derived
640 from a 3kb window upstream of TSSs. Y-axis is the number of transcripts, and the x-
641 axis shows the AAUAAA count per transcript.
642 **C.** Heatmap of control or RBBP6-depleted POINT-seq showing the RBBP6 effect on
643 antisense transcripts without an AAUAAA or those that contain at least one AAUAAA.
644 Most are unaffected by RBBP6 loss. The Y-axis is RPKM.

645 **SUPPLEMENTAL FIGURE 2, related to figure 2.**

646 A. Genome browser track showing POINT-seq signal over the first 7kb of *GLCCI1* in
647 *INTS11-dTAG* cells treated or not (1.5 hr) with dTAGv-1. This is an example of where
648 INTS11 loss affects protein-coding transcription as previously described[17]. Y-axis
649 shows RPKM following spike-in normalisation.
650 **B.** Metaplot of POINT-seq data from *INTS11-dTAG* cells treated or not (1.5hr) with
651 dTAGv-1. These are protein-coding genes arranged head-to-head. Signals above
652 and below the x-axis are sense and antisense reads, respectively. Y-axis shows
653 RPKM following spike-in normalisation.
654 **C.** Metaplot of POINT-seq data from *INTS11-dTAG* cells treated or not (1.5hr) with
655 dTAGv-1. This shows RNAs derived from enhancer clusters, which generally initiate
656 transcription bidirectionally. Signals above and below the x-axis are sense and
657 antisense reads, respectively. Y-axis shows RPKM following spike-in normalisation.

658

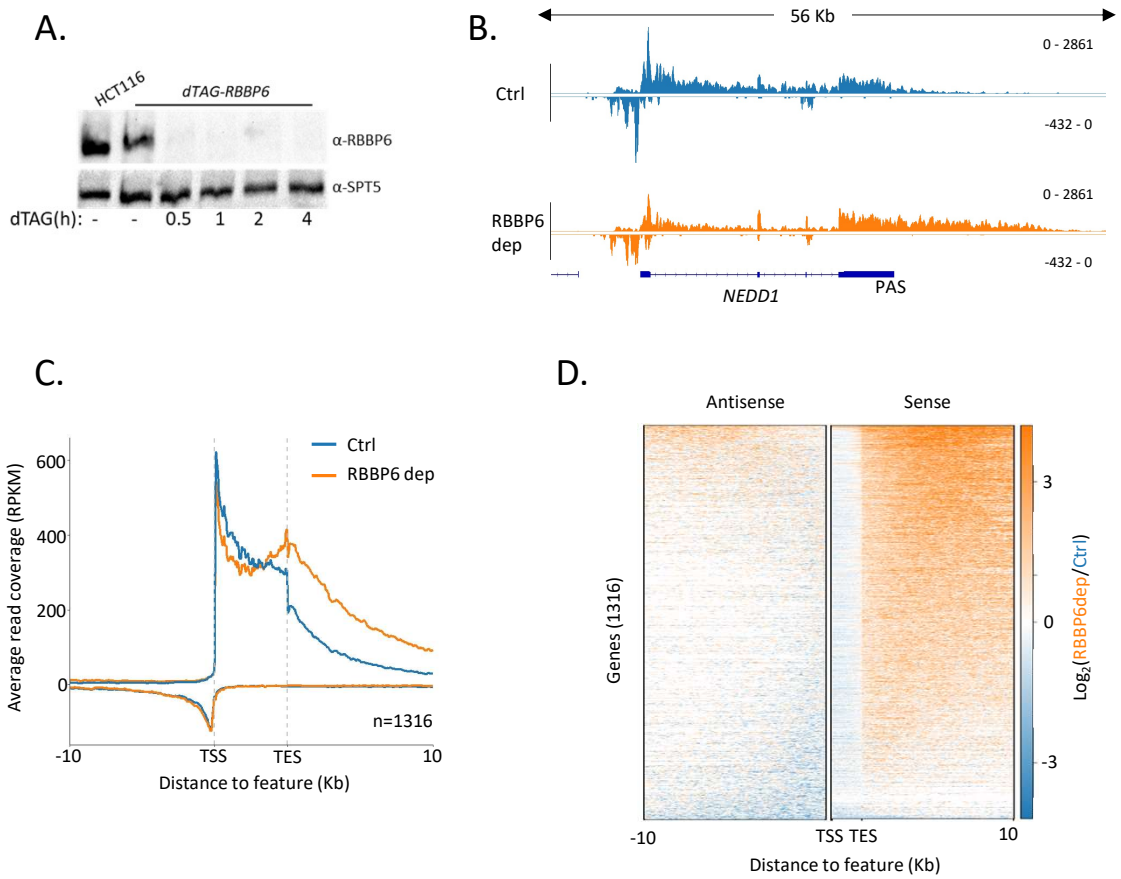
659 **SUPPLEMENTAL FIGURE 3, related to figure 3.**

660 **A.** Plot showing the size distribution of fragments mapped by sPOINT-seq. This
661 demonstrates its selectivity toward transcripts <150nts.
662 **B.** Violin plot of the sum sPOINT 5'end spike-in normalised RPKM signal across a 500bp
663 window from the protein-coding TSSs shown in Figure 3D (n=3060). These samples
664 are untreated with dTAG to quantitate the normal levels of sense vs. antisense
665 initiation/pausing in a window -/+ 500bp from the TSS in respective directions for
666 antisense and sense. This shows more initiation/pausing in sense vs. antisense
667 directions.
668 **C.** Genome browser track of *ACTB* promoter region in sPOINT-seq from *INTS11-dTAG*
669 cells treated or not with dTAGv-1 (1.5 hr). This showcases the focused sense TSS
670 (black arrows) and the dispersed antisense reads (brackets). Note the different y-axis
671 scales (RPKM) for sense vs. antisense.
672 **D.** Genome track of *RNU5A-1* and *RNU5B-1* in RNAPII ChIP-seq performed on *INTS11-*
673 *dTAG* cells treated or not with dTAGv-1 (2hr). INTS11 depletion causes RNAPII build-
674 up beyond both genes indicating a transcription termination defect. The Y-axis scale
675 is RPKM.
676 **E.** Metaplot of RNAPII occupancy of protein-coding TSS regions (the same gene set
677 employed for the sPOINT analysis in Figure 3) derived from RNAPII ChIP-seq
678 performed on *INTS11-dTAG* cells treated or not with dTAGv-1 (2hr). The Y-axis scale
679 is log10 qvalue of peak pileup values normalised to spike in control.

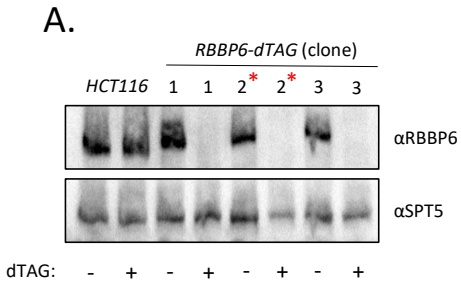
680 **SUPPLEMENTAL FIGURE 4, related to figure 4.**

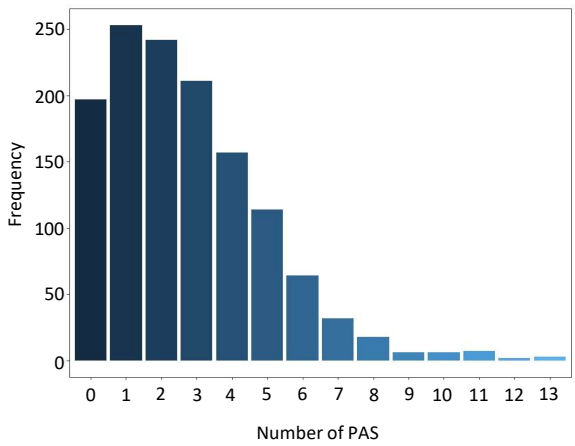
681 **A.** Genome browser track of *TARDBP* in POINT-seq data derived from *INTS11-dTAG*
682 cells either untreated, dTAG treated, NVP-2 treated or dTAG and NVP-2 treated (2.5
683 hr). The Y-axis scale shows RPKM.
684 **B.** As for A, but for *NEDD1*.
685 **C.** qRT-PCR analysis of PGK1, ENY2, PLEKHF2, and NUDCD1 pre-mRNAs in *INTS11-*
686 *dTAG* cells treated or not with dTAGv-1 and at the same time exposed or not to DRB
687 (all 2.5 hr). To enrich nascent transcripts, primers detect intronic RNA. Quantitation
688 shows fold change versus spliced actin relative to samples untreated with dTAGv-1 or
689 DRB. DRB treatment substantially reduces signal, which is restored when INTS11 is
690 co-depleted. Error bars show standard deviation. n=3, * denotes $p \leq 0.05$.
691 **D.** Analysis of published[18] INTS11 and CDK9 ChIP-seq data showing their respective
692 occupancy within 1kb of the TSS. Based on this data, they are slightly enriched
693 upstream (INTS11) and downstream (CDK9) of the TSS. The Y-axis units are RPKM.

FIGURE 1


A.
HCT116
dTAG-RBBP6
α-RBBP6
α-SPT5
dTAG(h): - - 0.5 1 2 4
B.
56 Kb
Ctrl
0 - 2861
-432 - 0
RBBP6 dep
0 - 2861
-432 - 0
NEDD1
PAS
C.
Ctrl
RBBP6 dep
Average read coverage (RPKM)
600
400
200
0
n=1316
-10
TSS
TES
10
Distance to feature (Kb)
D.
Antisense
Sense
Genes (1316)
3
0
-3
Log2(RBBP6dep/Ctrl)
-10
TSS TES
10
Distance to feature (Kb)

SUPPLEMENTAL FIGURE 1, related to figure 1





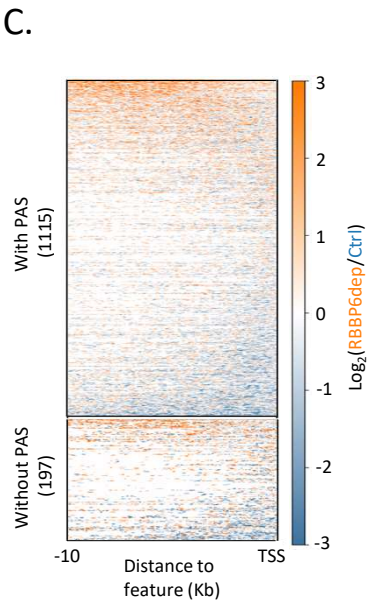
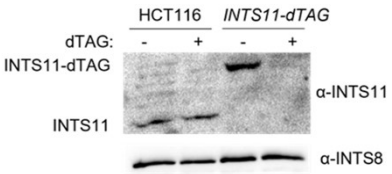
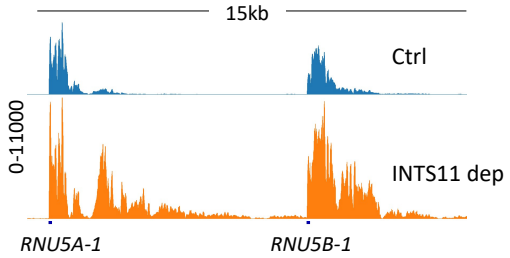


FIGURE 2

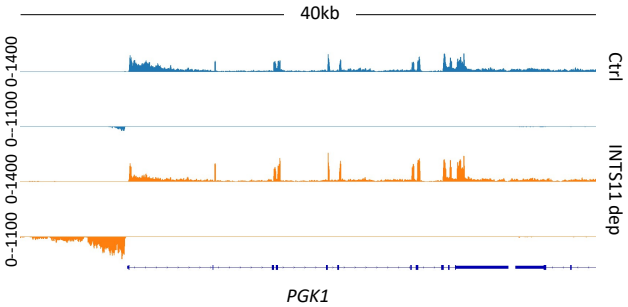
A.



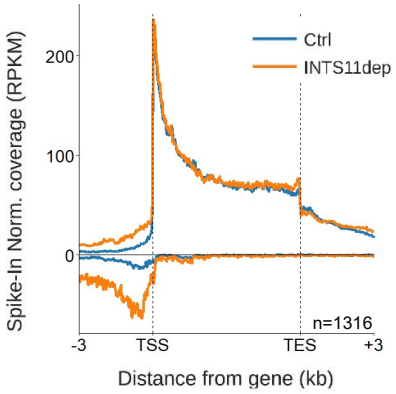
B.



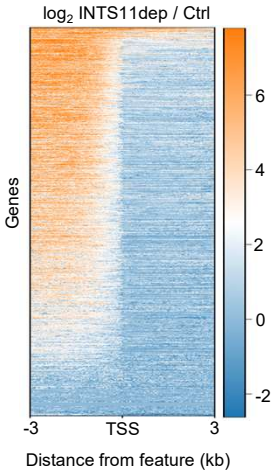
C.



D.



E.



SUPPLEMENTAL FIGURE 2, related to figure 2

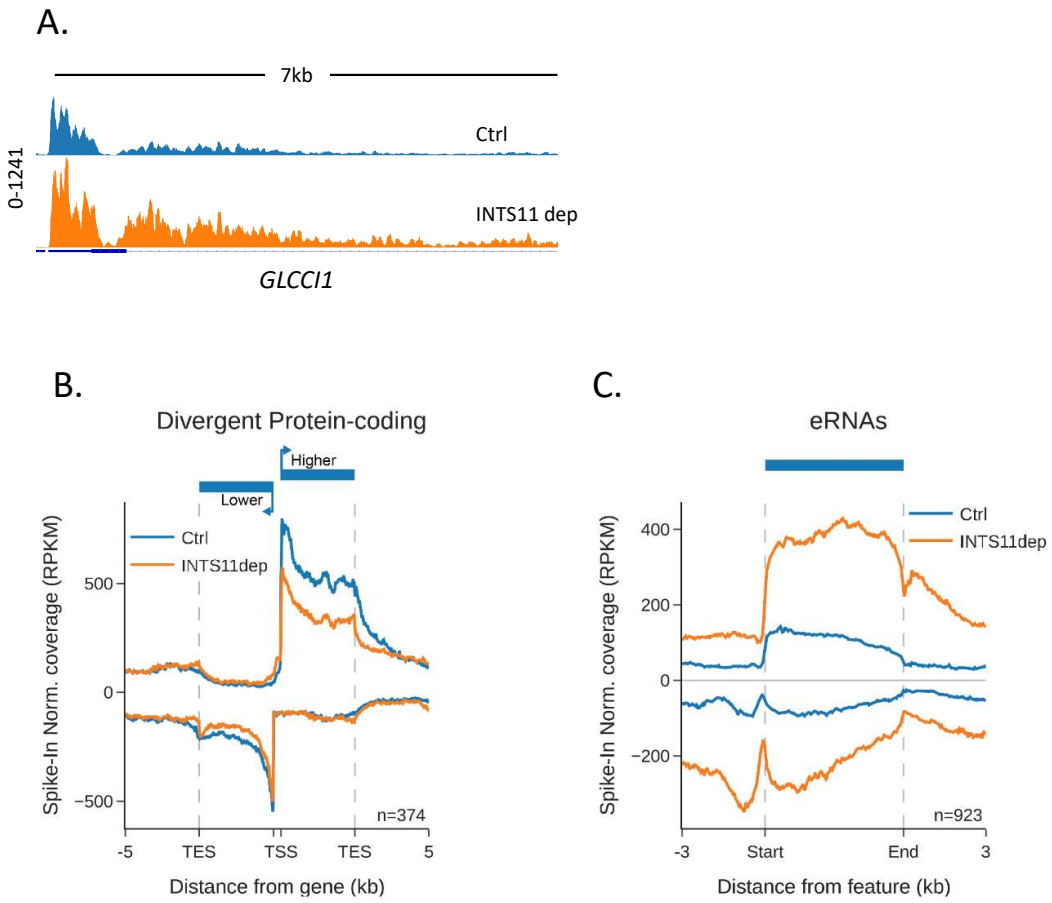
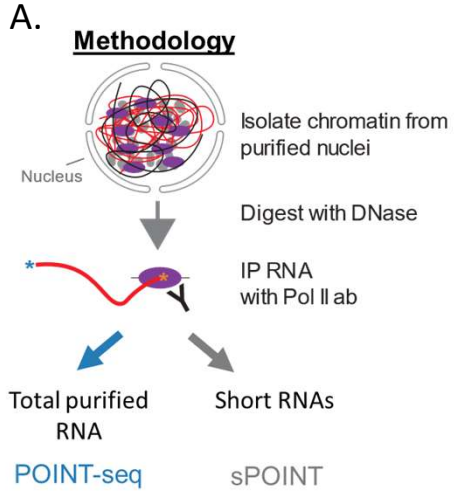
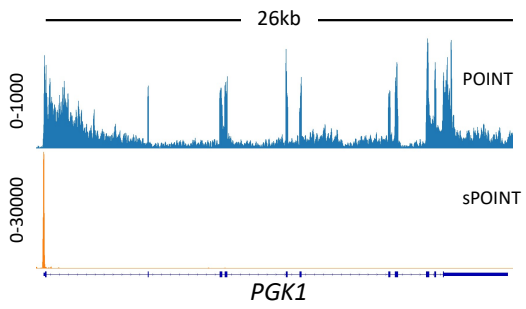


FIGURE 3

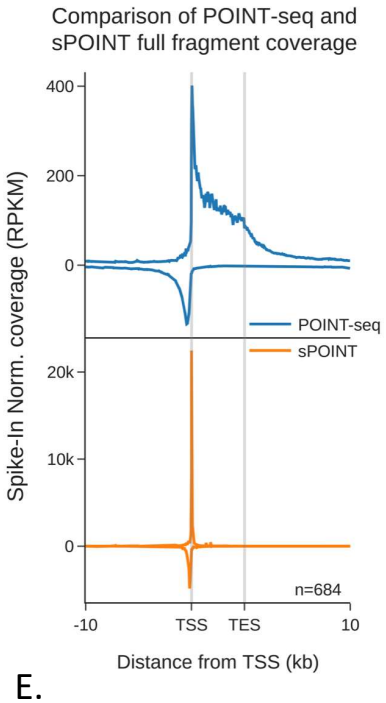
A.



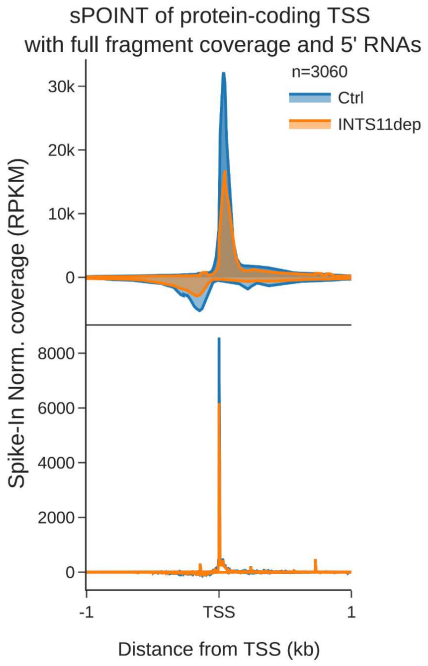
B.



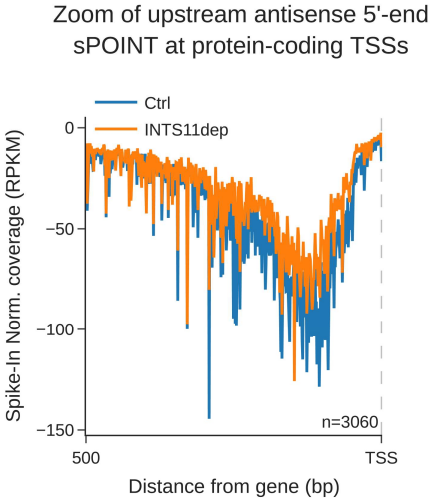
C.



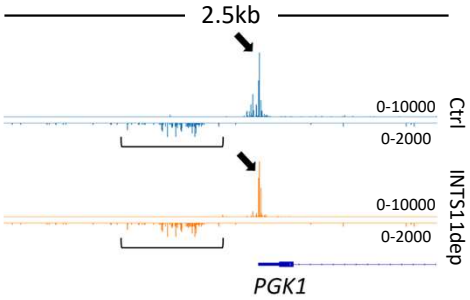
D.



E.

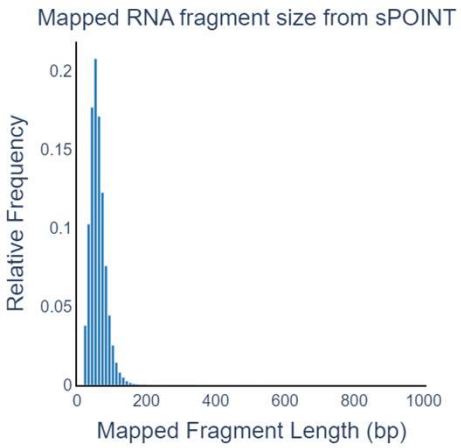


F.

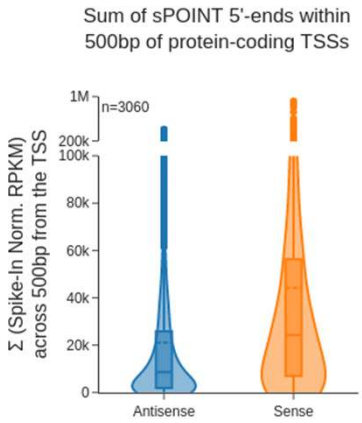


SUPPLEMENTAL FIGURE 3, related to figure 3

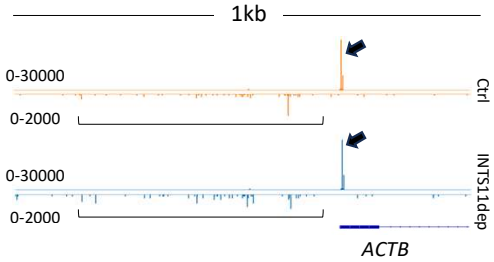
A.



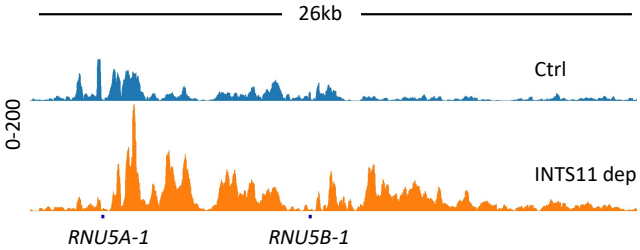
B.



C.



D.



E.

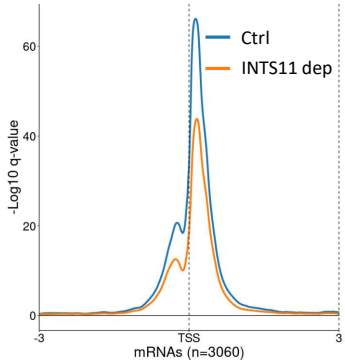
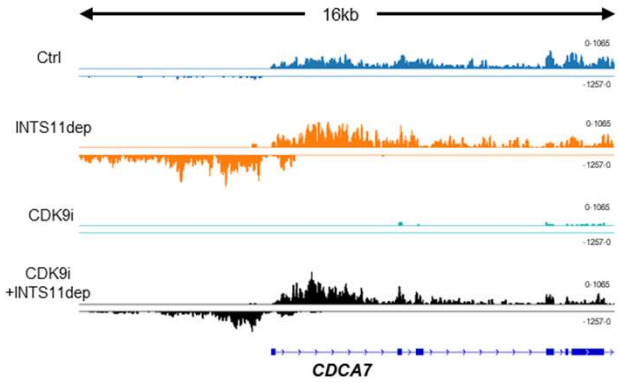
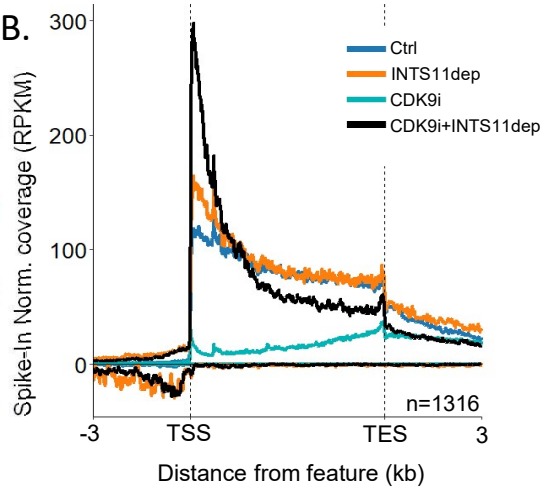


FIGURE 4

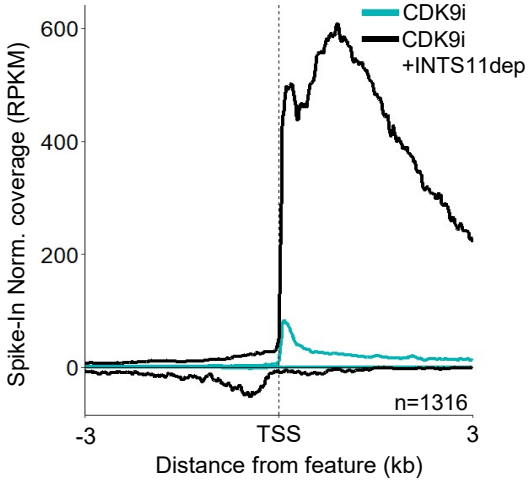
A.



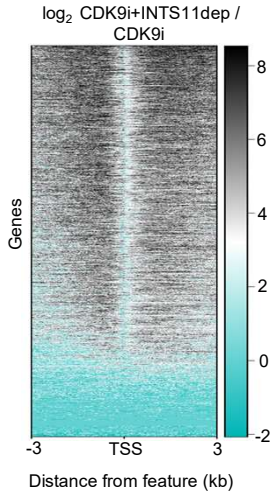
B.



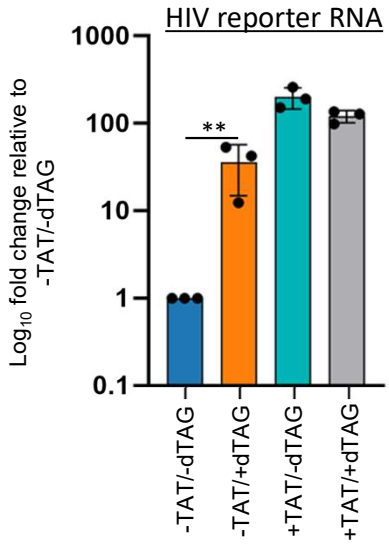
C.



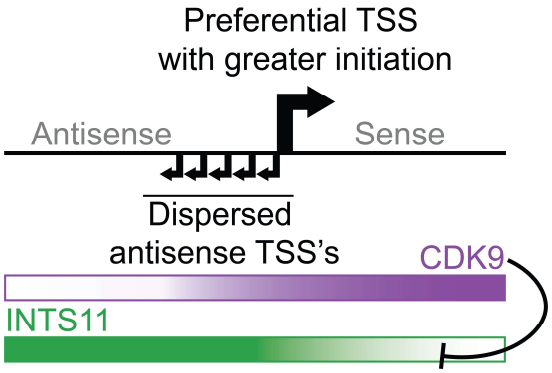
D.



E.

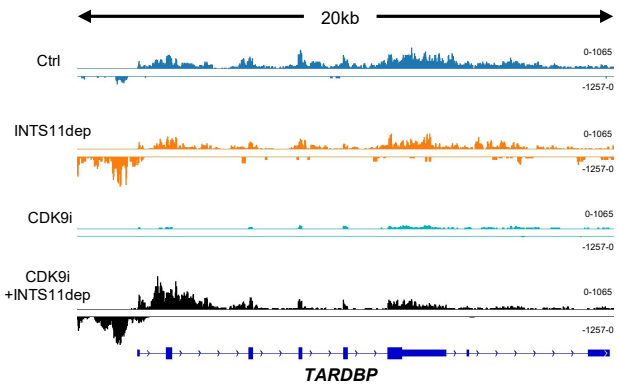


F.

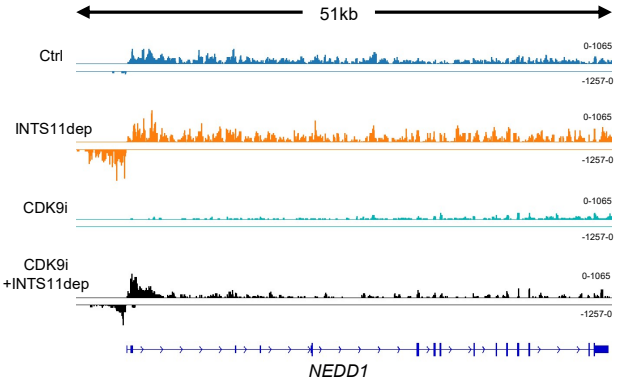


SUPPLEMENTAL FIGURE 4, related to figure 4

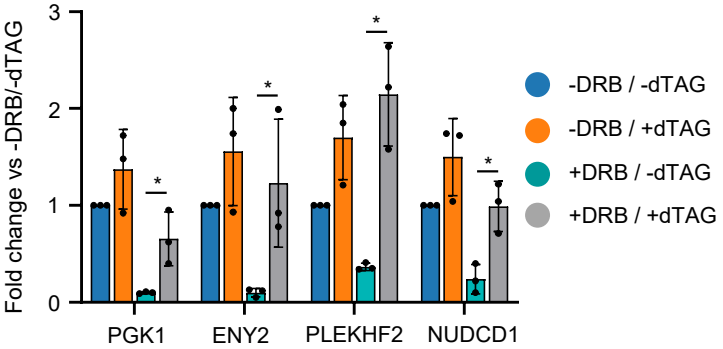
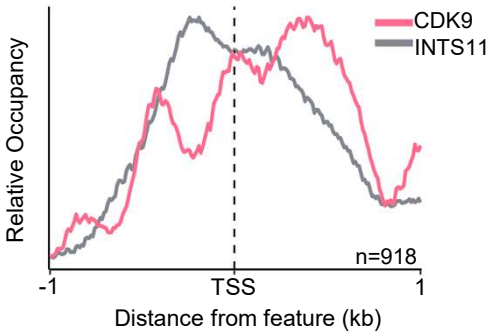
A.



B.



C.



D.