The chromatin remodeler Ino80 mediates alternative RNAPII pausing site determination

Youngseo Cheon[1], Sungwook Han[1], Taemook Kim[1], Daeyoup Lee[1,2,*]

[1]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea.

[2]Lead contact

*To whom correspondence should be addressed:

Tel: +82-42-350-2623

Fax: +82-42-350-2610

E-mail: daeyoup@kaist.ac.kr

Abstract

Promoter-proximal pausing of RNA polymerase II (RNAPII) is a critical step in early transcription elongation for the precise regulation of gene expression. Here, we provide evidence of promoter-proximal pausing-like distributions of RNAPII in *S. cerevisiae*. We found that genes bearing an alternative pausing site utilize Ino80p to properly localize RNAPII pausing at the first pausing site and to suppress the accumulation of RNAPII at the second pausing site, which is tightly associated with the +1 nucleosome. This alternative pausing site determination was dependent on the remodeling activity of Ino80p to modulate the +1 nucleosome position and might be controlled synergistically with Spt4p. Furthermore, we observed similar Ino80-dependent RNAPII pausing in mouse embryonic stem cells (mESCs). Based on our collective results, we hypothesize that the chromatin remodeler Ino80 plays a highly conserved role in regulating early RNAPII elongation to establish intact pausing.

Introduction

Emerging evidence indicates that promoter-proximal pausing is a decisive step in transcription that supports the precise control of gene expression in metazoans (*1*). The establishment and release of paused RNAPII are strictly regulated by several factors during the early transcription elongation stage. Early biochemical studies using a purified system provided key mechanistic insights into pausing, showing that it is governed by two critical factors: DRB-sensitivity-inducing factor (DSIF; the heterodimeric Spt4/Spt5 complex) (*2*) and negative elongation factor (NELF) (*3*). Advances in genomic technology have enabled researchers to track the position of elongation complexes genome-wide by several methods, such as by capturing actively elongating RNAPII (*4-6*) or selecting RNAPII-associated RNAs (*7, 8*). The use of deep sequencing methods revealed that the loss of NELF reduced but did not completely abolish promoter-proximal pausing (*9, 10*), suggested that NELF acted to stabilize pausing rather than

initiate it. Recent studies revealed that RNAPII pausing is found in species that lack NELF homologs, such as *C. elegans* (*11*) and *S. pombe* (*12*). Even *E. coli* RNA polymerases have been shown to pause at the start of the lambda gene (*13*). Further, the capture of nascent transcripts by native elongating transcript sequencing (NET-seq) in *S. cerevisiae* revealed that well-expressed genes exhibit a modest accumulation of read density downstream of the transcription start site (TSS) (*7*), which caused the non-uniform distribution of transcription elongation across genes. Overall, these previous studies have suggested that a conserved regulatory mechanism is involved in the early transcription elongation of yeast.

The nucleosome poses a strong barrier for RNAPII passage at various stages of transcription, and cells benefit from employing highly conserved chromatin remodelers to overcome these physical barriers (*14*). Several studies have shown that pausing occurs in close proximity to nucleosomes (*7, 12, 15, 16*). This suggested that nucleosomes can physically block the elongation of RNAPII, and that this collision causes RNAPII to pause. A genome-wide study targeting mouse Chd1 revealed that an ATPase inactive form of Chd1 results in a particular increase of RNAPII within the promoter regions (*17*), implying that chromatin remodeling could affect the promoter escape and subsequent pause-release of RNAPII. However, it remains unclear whether chromatin remodelers regulate nucleosome architecture to tune promoter-proximal pausing in the early elongation stage.

The chromatin remodeler, Ino80, has been shown to play a key role in the regulation of RNAPII at transcribed genes through its remodeling activity (*18*). The most well-known function of Ino80 is to exchange the highly conserved histone variant H2A.Z for H2A (*19, 20*). However, two recent studies disputed this function (*21, 22*). Several studies have suggested that Ino80p has the intrinsic capability of nucleosome spacing, and that it helps organize the intact nucleosome architecture around the promoter to regulate transcription (*23, 24*). Ino80 is largely enriched at the TSS of most genes in yeast and mammals (*20, 25*) and it has suggested that the recruitment of

Ies6, the component of Ino80 complex, to the 5' end of genes caused removal of histone H3-containing nucleosomes for gene expression in *S. pombe*. (*26*). Ino80 has also been shown to physically interact with the elongating RNAPII (*27, 28*). Nevertheless, the detailed mechanisms through which Ino80 regulates transcription elongation are currently unknown.

Here, we reveal the non-uniform distribution of elongating RNAPII in *S. cerevisiae* and show that it resembles promoter-proximal pausing in metazoans. Using the auxin-inducible degron (AID) system (*29, 30*), we found that Ino80p plays a key function in determining the position of RNAPII pausing in budding yeast. The genes whose pausing sites are regulated by Ino80p exhibited the use of an alternative pausing site rather than a focused pausing peak. The chromatin remodeling activity of Ino80p to properly localize the +1 nucleosome is essential to suppress RNAPII pausing at the second pausing site, thereby facilitating the utilization of the first pausing site. Further, we observed similar Ino80-dependent RNAPII pausing in mESCs, suggesting that Ino80 plays a highly conserved role in the regulation of RNAPII pausing.

Results

PRO-seq reveals genome-wide promoter-proximal pausing-like distributions in *S. cerevisiae*

To investigate the genome-wide distribution of elongation-competent RNAPII in *S. cerevisiae*, we first used Precision Run-On sequencing (PRO-seq) and Precision Run-On 5' cap sequencing (PRO-cap) with 2-biotin run-on (biotion-11-CTP and UTP) (*6, 12*) (table S1). To more precisely define the transcription initiation sites, we chose the single base pair with the most PRO-cap reads within 250 bases upstream and downstream of the annotated TSS. We herein refer to the newly defined observed TSS as a "TSS" unless otherwise noted. Unexpectedly, PRO-seq revealed that transcription elongation was non-uniformly distributed: It was concentrated near TSSs, with a pattern resembling that associated with promoter-proximal pausing in metazoans (Fig. 1A; See black line of the profile). This was surprising because a previous study using PRO-

seq in *S. cerevisiae* had captured a relatively uniform distribution of RNAPII across genes (*12*). To examine the significance of these apparent pausing-like features, we next classified genes as being paused or not paused, as described in the previous study (*12*). We identified 2,599 (45.5%) high-confidence paused and 2,099 (37.7%) not-paused genes among 5,315 filtered protein-coding genes (Fig. 1, A and B). The prevalence of pausing in *S. cerevisiae* was thus higher than that observed in *S. pombe* (28%) (*12*) and human (41%) (*4*) but lower than that observed in *D. melanogaster* (63%) (*9*). However, these differences in the relative number of paused genes could be a consequence of using different methods or gene sets to define RNAPII pausing (*31*).

We next sought to identify the general features of the paused genes in *S. cerevisiae*. To investigate whether the obtained PRO-seq reads were related to transcriptional activity or nucleosome density, we generated heatmaps of our data sets obtained from PRO-seq, PRO-cap, and existing data sets obtained using MNase-seq (*29*) and ChIP-seq against TBP and phosphor-Ser5 of RNAPII C-terminal domain (pSer5) (*32*). The RNAPII intensity within the promoter-proximal regions (TSS to TSS+250bp) generally correlated with the transcriptional activity (Fig. 1C); in this, our results were consistent with those of a previous study (*4*). In addition, the heatmaps showed that the higher the PRO-seq intensity, the lower the nucleosome occupancy and the wider the nucleosome free region (NFR) (Fig. 1C); this, too, was in accordance with earlier reports (*33*). Further, we compared our PRO-seq data with previously reported data sets obtained using Rpb3 NET-seq (*7*) and ChIP-seq (*34*). Consistent with our gene classification results, the NET-seq and ChIP-seq data displayed much higher enrichment at the TSS of paused genes compared to not-paused genes (Fig. 1D). Overall, we concluded that slowed RNAPII in *S. cerevisiae* is a conserved aspect of early transcription elongation that resembles promoter-proximal pausing in metazoans.

Pausing-like feature in budding yeast is broader and more distal than that in metazoans

Despite this expectation, we observed a striking difference in the PRO-seq distributions of *S. cerevisiae* compared to metazoans. The peak of the PRO-seq enrichment within the promoter-proximal regions was located ~100bp downstream of the TSS (Fig. 1A, Left). This localization was much farther downstream than the peak of the Global Run-On sequencing (GRO-seq) or the PRO-seq enrichment in metazoans, which showed the read peaks at ~50bp downstream of the TSS in the sense strand (*4, 5*). Given that PRO-seq can track transcription elongation at almost single base-pair resolution, we defined the single-nucleotide of the maximum PRO-seq read within the promoter-proximal regions as the pausing site, as described in a recent study (*16*). The cumulative curve demonstrated that the 25th and 75th percentiles of the distance from the TSS to pausing site were 74 and 154 bp, respectively (Fig. 1E; The red dotted lines). To determine the association between pausing and the +1 nucleosome, we generated an average profile of existing MNase-seq data used in Fig. 1C around the TSS (Fig. 1F, Left; The 25th and 75th percentiles of the distance from the TSS to the pausing site are represented as the two red dotted lines) and our PRO-seq data around the +1 dyad, which was determined by the same MNase-seq data (Fig. 1F, Right). Interestingly, the majority of pausing sites were found to be located downstream of the +1 nucleosome. This is consistent with the previous report using NET-seq, which showed the peak of mean pause density downstream of the +1 nucleosome dyad (*7*). In contrast, RNAPII is generally paused upstream of the +1 dyad (*5, 15, 16*) in metazoans. This difference suggests that RNAPII undergoes a longer elongation period before pausing in budding yeast compared to metazoans and it could likely be attributed to differences in promoter structure. The +1 nucleosome typically includes a TSS for most yeast genes, whereas the +1 nucleosome of metazoans is located downstream of the TSS (*35, 36*).

Loss of Ino80p causes variation in fitness and the PRO-seq pattern

Given the apparent role of Ino80p in transcription elongation (*18*), we used PRO-seq to investigate the role of Ino80p in nascent transcription at nearly single-nucleotide resolution. We

first set out to map transcription elongation in wild-type and *ino80*Δ cells. Interestingly, *ino80*Δ cells exhibited a 5'-direction skew of the promoter-proximal peak in replicates 2 (*ino80*Δ_2) and 3 (*ino80*Δ_3) relative to that in wild-type cells (fig. S1A). However, these results were not comparable, as replicate 1 (*ino80*Δ_1) did not show a skewed PRO-seq distribution (fig. S1A). Consistent with our PRO-seq results, the cells of same batch used to generate *ino80*Δ_1 exhibited better growth than *ino80*Δ_2 and *ino80*Δ_3 (fig. S1B). These variations in the PRO-seq distribution and cellular fitness of *ino80*Δ cells seem to be attributed to the appearance of revertant due to the severe growth defect of *ino80*Δ. Thus, we employed an auxin-inducible degradation system (*29, 30*), with the goal of generating highly reproducible data and exclude the indirect effect of Ino80p in transcription elongation. Briefly, Ino80-AID strains were grown to mid-log phase in yeast peptone dextrose (YPD) containing ethanol (Ctrl). Ethanol was washed from the media and the cells were incubated with auxin (0.5mM) for 3 hrs (KD). Auxin was washed from the media and cells were incubated in auxin-free medium for an additional 3 hrs (Rescue) (fig. S1C). Western blot analysis confirmed the conditional depletion and recovery of Ino80p in an AID-tag-dependent manner after the 3 hrs incubations with or without auxin (fig. S1D).

Ino80p is critical to the proper positioning of RNAPII pausing at the genes bearing an alternative pausing site

We carried out PRO-seq experiments to determine whether Ino80p knockdown (Ino80p-KD) caused a similar skewed pattern of the promoter-proximal peak in *ino80*Δ cells. The average profile of the median PRO-seq intensity for paused genes (N = 2,599) displayed a general skew of the promoter-proximal signal in the 5’ direction upon Ino80p-KD (fig. S2A). The pausing sites observed after 0 hr and 3 hrs of auxin treatment in Ino80-AID cells, which were assigned as described above, were designated “the 1st pausing site” and “the 2nd pausing site”, respectively. Cumulative curves and boxplots of the distance between the TSS and the pausing site revealed

that the pausing site was significantly shifted in the 5' direction upon Ino80p-KD (fig. S2B). To analyze this RNAPII transition in detail throughout the genome, we generated a heatmap of the PRO-seq $\log_2$ fold change around the 1st pausing site. Interestingly, 12.6% of the paused genes (N = 2,599) were shifted toward 3' upon Ino80p-KD; however, consistent with our cumulative curve and boxplot data, the 5' shift was stronger and more frequent (28.5%) (fig. S2C). We chose transcripts that showed shifts of more than 30bp in their pausing site upon Ino80p-KD (to enable us to observe a clear change in position) and investigated whether the pausing site tended to be restored upon Ino80p rescue. Indeed, a boxplot analysis indicated that the changes of pausing site were significantly recovered under auxin withdrawal for both shifted gene sets (fig. S2D). In contrast, Ino80p rescue did not affect pausing sites in genes that were not shifted upon Ino80p-KD. These results suggest that Ino80p plays a previously unrecognized function in proper localization of RNAPII pausing sites in the early elongation stage for a subset of genes.

To select genes directly regulated by Ino80p, we chose transcripts that were shifted more than 30bp upstream in knockdown cells and restored in rescued cells (N = 221) and those that failed to show any shift under knockdown or subsequent restoration (N = 1,211). We referred to the former gene set as "shift-to-5' genes" and the latter as "no-shift genes". To examine whether the assigned pausing site precisely reflects the peak of pausing, we generated the average profile of the median PRO-seq signal around the TSS for shift-to-5' genes. We observed that the PRO-seq peak exhibited striking upstream movement upon Ino80p-KD (Fig. 2A; The arrows and below dotted lines represent the median of the pausing sites). As expected, this shifted pausing was almost completely restored to the 1st pausing site when Ino80p was rescued (Fig. 2A). Thus, our results suggest that this pausing site determination is totally Ino80p expression-dependent.

To investigate the general features of Ino80p-dependent genes, we compared the PRO-seq pattern at the 1st pausing site under untreated conditions for no-shift genes and shift-to-5' genes. Surprisingly, we observed only a small accumulation of PRO-seq signal upstream of the peak at

the 1st pausing site for shift-to-5' genes, whereas no-shift genes displayed only a single sharp and distinct peak (Fig. 2B). The 25th and 75th percentiles of the 2nd pausing site relative to the 1st pausing site for shift-to-5' genes (represented by the red dotted lines in the Fig. 2B) indicated that Ino80p-KD displaces RNAPII solely at locations where pausing can also occur but not a main pausing site. These results implied that Ino80p-dependent genes bear an alternative pausing site and utilize Ino80p to facilitate RNAPII pausing at the main pausing site. Further, we found that pausing at the 1st pausing site was decreased upon Ino80p-KD and recovered upon Ino80p rescue (Fig. 2C, Upper) for shift-to-5' genes, whereas pausing at the 2nd pausing site showed an opposite tendency (Fig. 2C, Bottom). Thus, we propose that Ino80p is critical for the passage of RNAPII from the 2nd pausing site to the 1st pausing site for the establishment of intact pausing. We also noted that the elongating RNAPII at the 2nd pausing site in untreated Ino80-AID cells was above the basal level (Fig. 2C, Bottom; Ctrl sample), which is consistent with the hypothesis that RNAPII could also pause at the 2nd pausing site in the physiological state, but it does not represent a major pausing site. Supporting this idea, the sequence preferences at the 1st and the 2nd pausing site generated by WebLogo (*37*) exhibited a marked similarity (fig. S2E), suggesting the involvement of *cis*-acting nucleic acid sequences in RNAPII pausing. Overall, these results implied that Ino80p might play a pivotal role in determining proper pausing sites on genes bearing alternative pausing sites.

The transition of the pausing site in Ino80p-KD is independent of both TSS usage and H2A.Z^{Htz1}

We next determined whether the observed transition of RNAPII pausing was due to a defect in TSS usage. The precise transcription initiation sites for shift-to-5' genes upon Ino80p-KD were identified by the maximum PRO-cap read within 100bp around the defined TSS in control samples. Histograms of the distance between the TSS in control and knockdown samples indicated that the majority of these genes exhibited no differences in transcription initiation site

(fig. S2F, Left). This suggested that the Ino80p-dependent positioning of RNAPII pausing may result from a defect in transcription elongation rather than a defect in TSS usage. To further investigate whether the PRO-cap intensity was altered upon Ino80p-KD, we measured the $\log_2$ fold change of the PRO-cap signal 100bp around the TSS for shift-to-5' or no-shift genes (fig. S2F, Right). Our boxplot analysis demonstrated that the changes in PRO-cap signal for shift-to-5' genes showed a greater decrease than those for no-shift genes, suggesting that there may be an association between pausing site determination and the abundance of 5' capped RNA.

Given that previous reports proposed a connection between Ino80p and H2A.Z^{Htz1} (*19, 22*), we questioned whether the transition of RNAPII could be caused by the insufficient removal of H2A.Z^{Htz1}. To test this possibility, we carried out PRO-seq experiments in *htz1*Δ cells. The average profile, however, revealed that *htz1*Δ did not result in a skewed pattern of the promoter-proximal peak similar to that seen for the shift -to-5' genes upon Ino80p-KD (fig. S2G, Left). Moreover, the distance of pausing site shift in Ino80p-KD showed no significant correlation with H2A.Z^{Htz1} enrichment in the +1 nucleosome, which is calculated from an existing MNase-ChIP-seq (*38*) (fig. S2G, Right). Thus, we conclude that the function of Ino80p in pausing site positioning is independent of its role in restricting the localization of H2A.Z^{Htz1}, at least for shift-to-5' genes. Consistent with this proposal, previous studies showed that the occupancy of H2A.Z^{Htz1} on chromatin is not altered under Ino80 depletion (*21, 22*).

The transition of RNAPII pausing is closely associated with the +1 nucleosome

We questioned whether the transition of RNAPII upon Ino80p-KD was associated with the nucleosome architecture around the pausing site. To address this, we first analyzed the average profile of an existing MNase-seq data generated using untreated Ino80-AID cells (*29*), which is the same cell background used in our PRO-seq. Surprisingly, the nucleosome distribution relative to the 1st pausing site of shift-to-5' genes displayed a much better phase than that of no-shift genes (Fig. 3A). Also, the +1 nucleosome tended to be located in closer proximity to the 2nd

pausing site than the 1st pausing site (Fig. 3A; The 25th and 75th percentiles of the 2nd pausing site relative to the 1st pausing site are represented as the two dotted lines). This implicated that the ability of Ino80p to suppress RNAPII pausing at the 2nd pausing site in a nucleosome context-dependent manner. We also examined existing MNase-seq data obtained upon Ino80p-KD (*29*). While the nucleosome distribution for no-shift genes showed almost no difference, that for shift-to-5' genes exhibited a moderate disturbance in nucleosome positioning and a decrease in the +1 nucleosome occupancy upon knockdown (Fig. 3B). When we performed the same analysis for MNase-seq data distributed by other laboratories (*32*), which differed slightly in the cell background, incubation temperature, and auxin treatment time, we obtained similar results (fig. S3, A and B).

To continue addressing the association of the 2nd pausing site with the +1 nucleosome, we next evaluated the PRO-seq distribution around the +1 dyad (defined in Fig. 1F). To discard false-positive nucleosome positions, we excluded nucleosomes that did not overlap the H3K4me3 ChIP-seq enrichment calculated from the existing data (*34*), as previously reported (*16*). We generated a composite profile centered on the +1 dyad and a boxplot of the distance to pausing site from the +1 dyad. No-shift genes were divided by whether they exhibited a pausing site outside (N = 463) or inside (N = 401) of the +1 nucleosome, to clearly distinguish the changes in PRO-seq distribution. Neither group of no-shift genes displayed any distinct change in pausing site relative to the +1 dyad upon Ino80p-KD (fig. S3C). In striking contrast, we observed that the transition from the 1st to the 2nd pausing site occurred through the +1 nucleosome for shift-to-5' genes (N = 160) (Fig. 3C). It seemed that the large fraction of elongating RNAPII could not pass the +1 nucleosome upon Ino80p-KD. Rescue of Ino80p expression caused almost perfect restoration of the pausing site to the downstream of the +1 nucleosome, further supporting the direct function of Ino80p. Based on these findings, we propose that the regulation of the

nucleosome positioning around the 1st pausing site by Ino80p is critical to suppress RNAPII accumulation at the 2nd pausing site.

Ino80 remodeling activity might be critical in pausing site determination

To distinguish whether the regulation of nucleosome positioning around the pausing site depends on direct remodeling activity of Ino80p or other *trans*-activating factors associated with Ino80p, we carried out PRO-seq experiments in *arp5*Δ cells, which lack a component that is essential for the chromatin remodeling activity of Ino80p complex in *S. cerevisiae* (*39-41*). When we analyzed the defined paused genes as described for Ino80p-KD cells, we observed that a similar shift of pausing site toward the inside of the +1 nucleosome occurred solely for pausing site shifted toward 5' genes (> 30bp) in *arp5*Δ (N = 264) (Fig. 3D and fig. S3D). Corroborating this, a Venn diagram analysis revealed that there was a significant overlap (*P* value = 1.52×10^{-12}) between shift-to-5' genes upon Ino80p-KD and genes showing an upstream shift of RNAPII pausing in *arp5*Δ (fig. S3E). Thus, we concluded that the chromatin remodeling activity of the Ino80p complex to the +1 nucleosome plays an important role in well-positioning of RNAPII pausing.

The conserved pausing factor, Spt4p, plays a similar role in regulating alternative pausing site

Since a previous study suggested that the conserved pausing factor, Spt4p, is required to facilitate productive transcription elongation in *S. cerevisiae* (*12*), we postulated Spt4p is also critical for transitiong the elongating RNAPII from the 2nd to the 1st pausing site. To examine this, we conducted PRO-seq in *spt4*Δ cells. Using the same set of analyses performed with the PRO-seq data in Ino80p-KD, we selected pausing site shifted toward 5' direction (> 30bp) gene sets in *spt4*Δ (N = 784). As expected, we observed a similar upstream skewed pattern of the PRO-seq peak in *spt4*Δ relative to wild-type within the promoter-proximal regions (Fig. 4, A and B). Further, PRO-seq signal in *spt4*Δ showed a prominent increase immediately downstream of the

TSS, which was consistent with the previous report (*12*). There was a significant overlap (*P* value = 5.40×10^{-8}) between shift-to-5' genes upon Ino80p-KD and Spt4p-regulated genes (Fig. 4C).We also observed strong negative genetic interactions between *INO80* and both *SPT4* and *PAF1* (Fig. 4D), the latter of which has been shown to physically interact with DSIF and RNAPII and facilitate RNAPII pausing (*42-44*). Recent work showed that Spt4, Spt5, and the additional elongation factor Elf1 could cooperatively prevent RNAPII pausing at SHL(-6), SHL(-5) and SHL(-2) *in vitro* (*45*), suggesting that Spt4 plays a role in overcoming the nucleosome barrier. Based on these results, we propose that Ino80p and Spt4p synergistically regulate the ability of elongating RNAPII to travel along the +1 nucleosome and establish intact pausing in *S. cerevisiae*.

The role of Ino80p in RNAPII pausing is conserved in mESCs

Since the Ino80 complex is a highly conserved chromatin remodeler from yeast to humans (*18*), we investigated whether Ino80 loss also caused a defect in elongating RNAPII positioning before pause-release in mESCs. Toward this end, we carried out PRO-seq experiments in mESCs treated with either *siEgfp* or *siIno80* (fig. S4A). To more precisely identify the peak of promoter-proximal pausing, we tiled 1kb around the annotated TSSs (TSS ± 1kb) in a 50-bp window with a 5-bp shift, as previously reported (*4*). We selected the window showing the maximum PRO-seq reads, and designated the 5' end of the selected window as the "PRO-seq peak". The average profile of median PRO-seq intensity in mESCs treated with *siEgfp* revealed an almost 2-fold higher PRO-seq intensity at the PRO-seq peak than at the annotated TSS (fig. S4B). Because the PRO-seq signal was highly confined around the PRO-seq peak regions, we determined the regions from 100bp upstream to 200bp downstream of the PRO-seq peak as the promoter-proximal regions for mESCs. Based on the PRO-seq coverage of these regions, we classified genes as being high-confidence paused (N = 27,406; 78.5%) or high-confidence not paused (N = 5,427; 16.3%) among the total protein-coding genes (N = 31,173). To analyze the RNAPII

pausing shift under Ino80-KD, we defined the 1st and the 2nd pausing site for each gene in the same manner as in *S. cerevisiae*. To distinguish the pattern of RNAPII transition for each gene, we generated a heatmap of the PRO-seq $\log_2$ fold change around the pausing site (fig. S4C). However, the results obtained from these analyses differed from those observed in *S. cerevisiae* (fig. S2C and fig. S4C). Although the 2nd pausing site was shifted upstream relative to the 1st pausing site for a small fraction of genes (N = 1,082; 8.9%) upon Ino80-KD, a much larger fraction of genes (N = 2,635; 23.2%) displayed a downstream shift in mESCs (fig. S4C).

Except for the direction of shift, however, Ino80-KD in mESCs yielded a pausing site-determination defect similar to that observed in *S. cerevisiae*. First, the average profile of the median PRO-seq signal around the PRO-seq peak regions for genes whose pausing site shifted more than 30bp in the 3' direction upon Ino80-KD ("shift-to-3' genes"; N = 2,324) revealed that Ino80-KD induced prominent downstream movement of RNAPII pausing (Fig. 5A; The arrows and below dotted lines represent the median of the pausing sites). Second, these shift-to-3' genes exhibited the use of an alternative pausing site rather than a focused pausing peak in the physiological conditions (Fig. 5B). Ino80-KD displaces RNAPII solely at locations where pausing can also occur (Fig. 5B; The 25th and 75th percentiles of the 2nd pausing site relative to the 1st pausing site for shift-to-3' genes are represented as the two red dotted lines). Third, whereas pausing at the 1st pausing site was decreased by Ino80-KD, pausing at the 2nd pausing site was significantly increased under the same condition (Fig. 5C). The sequence preferences at each pausing site were also similar (fig. S4D). Furthermore, *de novo* motif finding analysis using HOMER (*46*) indicated that the YY1 motif was significantly enriched in the promoter regions of shift-to-3' genes relative to no-shift genes (fig. S4E). YY1 is physically associated with the mammalian Ino80 complex (*47*), providing additional support for the engagement of Ino80 with these genes. We did not observe a similar motif in *S. cerevisiae*, perhaps reflecting its lack of an identified homolog for YY1 (*48*). Based on our findings, we conclude that Ino80 plays a

conserved role in determining where RNAPII should mainly pause at the genes bearing an alternative pausing site in the early transcription elongation stage in mESCs.

We next deciphered whether the 2nd pausing site was also associated with the +1 nucleosome in mESCs. We analyzed an existing MNase-seq data set obtained from untreated mESCs (*49*). We observed the better positioned +1 nucleosome for shift-to-3' genes (N = 2,324) relative to no-shift genes (N = 16,903) (Fig. 5D), and found the proximity of the 2nd pausing site to the +1 nucleosome (Fig. 5D; The dotted lines represent the 25th and 75th percentiles of the 2nd pausing site relative to the 1st pausing site). This suggested that the nucleosome context around the 1st pausing site are important for mESCs to suppress the 2nd pausing site, as also found in budding yeast. Consistent with previous reports, the PRO-seq signals corresponding to the 1st pausing site for mammalian shift-to-3' genes were located near the entrance of the +1 nucleosome (Fig. 5E) (*5, 15, 16*). The PRO-seq signals corresponding to the 2nd pausing site, however, were found between the 1st pausing site and the +1 dyad (Fig. 5E). In control analysis, we detected almost no difference in the PRO-seq pattern around the +1 dyad for no-shift genes (fig. S4F). From these results, we conclude that Ino80 plays a conserved regulatory role in proper localization of RNAPII pausing at the 1st pausing site in a nucleosome-context-dependent manner in various organisms, from budding yeast to mouse (Fig. 6).

Discussion

We found highly confined PRO-seq signals immediately downstream of the observed TSSs in *S. cerevisiae*, which yielded a pattern that was highly similar to that representing promoter-proximal pausing in metazoans. However, the slowed promoter-proximal elongation in *S. cerevisiae* occurred more broadly and RNAPII seemed to be paused farther downstream than in metazoans. Indeed, most of the 1st pausing sites were located downstream of the +1 nucleosome. This finding was unlike that in higher eukaryotes, where most pausing was found to occur

upstream of the +1 dyad (*5, 15, 16*). One of the reasons for this discrepancy might be that NELF, which is involved in the pausing of RNAPII at a more promoter-proximal region (*50*), is not conserved in yeasts, including *S. cerevisiae* and *S. pombe* (*51, 52*). The hypothesis is consistent with the PRO-seq distribution reported for *S. pombe*, which showed a near-overlapping association of RNAPII pausing with the +1 dyad (*12*).

Surprisingly, Ino80 knockdown yielded opposite shifts of the RNAPII pausing sites in budding yeast and mESCs. One possible explanation is that Ino80 plays different roles in these species, since the Ino80 complex contains species-distinct components that could target Ino80 to specific contexts. Alternatively, our observation that the nucleosome distribution around the pausing site was similar in the two organisms led us to postulate that these discrepancies could be due to differences in the chromatin architectures around promoter regions in budding yeast versus mESCs (*35, 36*). In *S. cerevisiae*, the +1 nucleosome generally includes a TSS for most genes. Thus, most of the 1st pausing sites are downstream of the +1 nucleosome. The loss of Ino80p induced the main pausing site to shift toward the +1 nucleosome, accompanied by both a moderate increase in nucleosome fuzziness and a decrease in the +1 nucleosome occupancy. This suggests that Ino80p may play a critical role in the passage of elongating RNAPII from the 2nd to 1st pausing site and helps establish intact pausing through its activity to position the +1 nucleosome. Supporting our hypothesis, previous studies showed an increase of nucleosome fuzziness in Ino80p mutants (*21, 41*) and demonstrated that Ino80p functions to pull the +1 nucleosome toward the NFR for a subset of genes (*32*). In higher eukaryotes, the +1 nucleosome is located farther downstream of TSS than in budding yeast. Consistent with this, the median 1st pausing site for Ino80-dependent genes in mESCs was found to be upstream of the +1 nucleosome (Fig. 5E; The median of the *siEgfp* treated samples was 84bp upstream of the +1 dyad). On one hand, this could reflect that Ino80 can promote RNAPII pausing in a nucleosome-independent manner. Alternatively, a previous study implied that the positioned +1 nucleosome

recruits more NELF and enhances promoter-proximal pausing (*53*). Thus, mammalian Ino80-dependent nucleosome positioning may also work to establish RNAPII pausing through recruiting *trans*-activating pausing factors. Consistent with this scenario, RNAPII resumed elongation to the 2nd pausing site upon the loss of Ino80 (Fig. 5E). The delocalization of the +1 nucleosome due to Ino80 depletion probably allows further elongation of RNAPII, and the 2nd pausing that occurs around 30bp upstream of the +1 dyad (Fig. 5E; The median of the *siIno80* treated samples) might be due to a physical blocking of the +1 nucleosome. This location seems to be close to the location of SHL(-2), at which transcribing RNAPII can be stalled *in vitro* (*54*). Based on the present and previous findings, we suggested that the nucleosome-dependent intrinsic pausing at the 2nd pausing site is conserved from budding yeast to mESCs, and that Ino80 plays a key role in actively determining the main pausing site through its ability to modulate +1 nucleosome positioning in a context-dependent manner (Fig. 6). In support of our hypothesis, we note that a recent study using human NELF-C-AID suggested that NELF loss results in a similar downstream shift of RNAPII pausing accumulated at the 2nd pausing site, which is tightly associated with the +1 nucleosome (*16*).

Supporting the idea that the between-species discrepancies in RNAPII pausing observed herein are due to differences in chromatin architecture, a defect in DSIF was also reported to cause different phenotype in *S. cerevisiae* versus *M. musculus*. The deletion of Spt4p in budding yeast showed the prominent increase in PRO-seq signal at the 5' ends of genes (*12*), whereas Spt5 depletion in mouse triggered the accumulation of PRO-seq signals downstream of the promoter-proximal regions (*55*). Interestingly, the directions in which elongating RNAPII accumulated in the above-described mutants are similar to that observed following Ino80-KD in the present work, further supporting that Ino80 and DSIF may play synergistic roles in both species. In addition, Spt4 deletion in *S. pombe* has resulted in PRO-seq distribution resembling a defect upon Spt5 loss in mouse (*12*). Intriguingly, a previous study found a similar but distinct nucleosome distribution

in budding and fission yeasts (*56, 57*). Further experiments seeking to gain additional contextual information or identify unknown factors in *S. pombe* might help explain these discrepancies in two divergent yeasts.

Based on our collective results, slowed elongating RNAPII within the promoter-proximal regions is an evolutionarily conserved mechanism from yeast to human. Chromatin remodelers could not only regulate promoter opening and productive elongation but also govern promoter-proximal RNAPII pausing through its role to tune nucleosome architecture. In the future, it will be interesting to study the detailed mechanism of chromatin remodeling in the early transcription elongation.

Materials and Methods

Yeast strains and cell culture

The budding yeast strains used in this study are listed in table S2. AID-tagged proteins were conditionally depleted using 250mM auxin (Sigma, I2886) stock dissolved in ethanol at a final concentration of 0.5mM, as previously described (*29*). Briefly, Ino80-AID cells were grown to mid-log phase at 30°C in YPD containing ethanol. The ethanol was removed by media exchange, cells were incubated with auxin (where indicated) for 3 hrs to allow conditional depletion. The auxin was removed by media exchange, and cells were incubated in auxin-free medium an additional 3 hrs. At the indicated time points, Ino80-AID cells were harvested and subjected to PRO-seq and PRO-cap. The workflow is schematically presented in fig. S1C. The efficiency of Ino80p knockdown was confirmed by Western blotting. To generate the deletion strains, we conducted standard LiAc transformation using PCR-based gene targeting. These cells were incubated to mid-log phase at 30°C in YPD and were subjected to PRO-seq.

mESC culture

E14Tg2a mESCs were maintained under feeder-free conditions at 37°C with 5% CO_2 in humidified air. Briefly, mESCs were cultured on gelatin-coated dishes in Glasgow's minimum essential medium (GMEM) containing 10% knockout serum replacement (Gibco, 10828-028), 1% non-essential amino acids (Gibco, 11140-050), 1% sodium pyruvate (Gibco, 11360-070), 0.1 mM β-mercaptoethanol (Gibco, 21985-023), 1% FBS (Hyclone, SH30917.03), 0.5% antibiotic-antimycotic (Thermo, 15140122), and 1,000 units/ml LIF (Millipore, ESG1106).

RNA interference

The siRNAs against EGFP (5'-GUUCAGCGUGUCCGGCGAG-3') and Ino80 (5'-GGCUUAUCUGUAAAGGCACAAUUGA-3') were synthesized and annealed by Bioneer. mESCs were seeded to plates, incubated for 24 hrs, and transfected with the indicated siRNAs (final concentration, 50nM) using DharmaFECT I (Dharmacon, T-2001-03) according to the manufacturer's protocol. Briefly, 50nM of siRNAs and DharmaFECT I diluted in Opti-MEM (Gibco, 31985062) were separately incubated for 5 min at 25˚C, mixed and incubated for 20 min at 25˚C, and then used for transfection. The culture medium was replaced at 24 hrs of transfection and cells were harvested at 48 hrs of transfection.

Western blot analysis of protein depletion

Whole-cell lysates of Ino80-AID cells were prepared using a standard bead-beating protocol, and proteins were eluted by boiling at 100˚C for 5 min in 2x SDS sample buffer (20% glycerol, 0.4% bromophenol blue, 100mM Tris-Cl, pH6.8, 4% SDS, and 200mM β-mercaptoethanol). The utilized antibodies included anti-FLAG M2 (Sigma A8592; used at 1:3000) and anti-β-actin (Santa Cruz sc-47778 HRP; used at 1:5000). The AID-tagged Ino80 cells were a gift from the Friedman lab as previously described (*29*).

mESCs were washed with PBS (Welgene, LB004-02) and detached from the dishes by incubation with trypsin-EDTA (Gibco, 25300-054) at 37˚C for 2 min. The trypsin was inactivated by the addition of GMEM with 1% FBS and 0.5% antibiotic-antimycotic, and the cells were harvested,

washed with PBS, and resuspended in EBC300 (120 mM NaCl, 0.5% NP-40, and 50 mM Tris-Cl, pH 8.0) containing protease inhibitors. Whole-cell lysates were prepared by vigorously vortexing the cell mixture for 30 min followed by centrifugation for 30 min at 4˚C. Proteins were eluted by being boiled at 100˚C for 5 min with 5x SDS sample buffer. The utilized antibodies included anti-Ino80 (Abcam, ab118787; used at 1:1000) and anti-tubulin (Cell Signaling, 2144S; used at 1:4000).

Spotting assay

Cells were spotted onto the indicated plate in serial five-fold dilutions starting at 0.5OD, and then incubated at 30˚C for ~2 days.

Yeast cell permeabilization

Yeast cells were permeabilized as described (*6*) with some of previously reported modifications (*58*). Briefly, exponentially growing yeast cells were harvested by centrifugation at 2000rpm for 3min at 4˚C. Cells were washed once with ice-cold DEPC-H_2O. Cell pellets were resuspended in 10ml of 0.5% sarkosyl (Sigma, L5777) and incubated for 20min on ice. Cells were spun down at 400xg for 5min at 4˚C and resuspended in storage buffer (10mM Tris-Cl, pH8.0, 25% glycerol, 5mM $MgCl_2$, 0.1mM EDTA, and 5mM DTT) to an OD of 5 per 200µl. The solutions were flash-frozen using LN_2 and stored at -80˚C.

Isolation of nuclei

mESCs were transfected with the indicated siRNAs, and nuclei were isolated as described (*4, 44*) with some modifications. Briefly, ~20 x 10^6 plated mESCs were washed once with PBS and detached by incubation with trypsin-EDTA at 37˚C for 2min. The trypsin was inactivated by the addition of GMEM with 1% FBS and 0.5% antibiotic-antimycotic, and the cells were harvested and washed twice with ice-cold PBS. The cells were resuspended in 5ml of ice-cold swelling buffer (20mM Tris-Cl, pH7.5, 2mM $MgCl_2$, 3mM $CaCl_2$, and 2U/ml RNase inhibitor) for 5min on ice. Lysis buffer (5ml; 20mM Tris-Cl, pH7.5, 2mM $MgCl_2$, 3mM $CaCl_2$, 0.5% NP-40, 10%

glycerol, and 2U/ml RNase inhibitor) was added and the cell pellets were resuspended by gentle pipetting using an end-cut tip. The cells were centrifuged at 1000xg for 5min at 4˚C and the cell pellets were resuspended in 1ml of freezing buffer (50mM Tris-Cl, pH8.3, 40% glycerol, 5mM $MgCl_2$, and 0.1mM EDTA). The pelleted nuclei were transferred into a new 1.5ml tube and were resuspended in freezing buffer at ~5 x 10^6 nuclei per 100µl. The solutions were flash-frozen using LN_2 and stored at -80˚C.

PRO-seq and PRO-cap library preparation

Nuclear run-on reactions and RNA extractions were performed based on the published protocol (*6*) with minor modifications that were previously reported (*44, 58*). Briefly, the flash-frozen yeast cells were quickly thawed on ice. For the yeast spike-in control, 0.125OD of permeabilized *S. pombe* (ED665) cells were added to each 5OD of permeabilized *S. cerevisiae* sample before the nuclear run-on reaction was performed. Combined yeast cells were spun down at 400xg for 5min at 4˚C. Nuclear run-on reactions were conducted with 25µM biotin-11-UTP (PerkinElmer, NEL543001EA), 25µM biotin-11-CTP (PerkinElmer, NEL542001EA), 125µM ATP (Roche, 11140965001), and 125µM GTP (Roche, 11140957001) in run-on reaction buffer (20mM Tris-Cl, pH7.7, 200mM KCl, 5mM $MgCl_2$, 2mM DTT, and 0.4U/µl RNase inhibitor) with 0.5% sarkosyl. The reaction mixtures were incubated at 30˚C for 5min. For the isolated nuclei of mESCs, nuclear run-on reactions were performed with 25µM biotin-11-UTP (PerkinElmer, NEL543001EA), 25µM biotin-11-CTP (PerkinElmer, NEL542001EA), 125µM ATP (Roche, 11140965001), and 125µM GTP (Roche, 11140957001) in run-on reaction buffer (5mM Tris-Cl, pH8.0, 150mM KCl, 2.5mM $MgCl_2$, 0.5mM DTT, and 0.4U/µl RNase inhibitor) with 0.5% sarkosyl. The reaction mixtures were incubated at 37˚C for 5 min.

RNA was extracted from the run-on-reacted cell pellets using either a standard hot-phenol method (for yeast samples) or TRIzol LS (Ambion, 10296028; for mESC samples). Next, the respective library was generated followed using the published PRO-seq or PRO-cap protocols (*6*)

for the steps spanning RNA fragmentation by base hydrolysis to full-scale PCR amplification. Note that there were a few differences in the applied reagents: we used Superscript IV reverse transcriptase (Invitrogen, 18091050) instead of Superscript III (Invitrogen, 56575); we used 25mM of each dNTP (Thermo Scientific, R1121) instead 12.5mM of each dNTP (Roche, 03622614001); and we used Phusion High-Fidelity DNA Polymerase (Thermo Scientific, F530L) instead of Phusion polymerase (NEB, M0530). DNA libraries of ~140 bp to 350 bp were selected by agarose gel extraction (Zymo Research, D4007) according to the manufacturer's protocol and sequenced using an Illumina HiSeq X Ten, HiSeq 4000 and NovaSeq 6000.

Sequence alignment and data processing (PRO-seq and PRO-cap)

Sequence alignment and data processing were performed based on the publicly available alignment pipelines of GitHub, as used in the previous study (*58*) with minor modifications. Briefly, raw sequencing reads were processed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) as follows: Adaptor sequences (5'-TGGAATTCTCGGGTGCCAAGG-3') were removed, the reads were trimmed to a maximum length of 36nt and, for PRO-seq, the reads were reverse-complemented. Next, reads that mapped to rRNA sequences were depleted using SortMeRNA (*59*), and reads that were not mapped to rRNA sequences were uniquely aligned to the genome using Bowtie, with the algorithm allowing for two mismatches (*60*): The processed reads of yeast samples that were generated with the spike-in approach were mapped to a combined genome consisting of *S. cerevisiae* (sacCer3) and *S. pombe* (SpombeASMv2), and then uniquely aligned reads from each genome were parsed for downstream analysis. The processed reads of mESCs samples were mapped to the *M. musculus* mm10 genome. The coverage of the aligned reads was generated using the genomecov function of BEDtools (*61*). Only the most 3' nucleotide of each read was calculated for PRO-seq, and only the most 5' nucleotide of each read was calculated for PRO-cap. For the spike-in control, the recorded coverage in the bedGraph file was normalized by the relative number of reads mapped

to a *S. pombe* genome (table S1). bedGraph files were converted to BigWig files by bedGraphToBigWig (*62*) and the downstream analysis was performed based on the publicly available custom R scripts on GitHub, as previously reported (*58*).

Protein-coding gene sets based on the annotated data in the Saccharomyces Genome Database (SGD; N = 6,692) were initially used for *S. cerevisiae* samples. The observed TSS was defined as the single nucleotide with the most PRO-cap reads within the 250bp upstream and downstream of the annotated TSS, in a similar manner to that used in the previous report (*12*). Genes with no PRO-cap signal, genes that had PRO-seq reads lower than 10, and genes shorter than 300bp were filtered out; in the end, 5,715 genes were used out of 6,692 SGD genes. For mESCs, protein-coding gene sets based on the RefSeq annotation were downloaded from UCSC Genome Browser. Genes that had PRO-seq reads lower than 10 and those shorter than 1kb were discarded; in the end, 31,173 genes were selected out of 37,802 genes. The annotated TSS ± 1kb was tiled in a 50-bp window by shifting 5 bp, and the PRO-seq coverage for each window was measured. The window of the most PRO-seq reads was selected and the 5' position of the selected window was used as the PRO-seq peak, in a manner similar to that described in a previous study (*4*). To identify paused gene sets, mappable reads within the promoter-proximal regions and gene-body regions were calculated. Significant enrichments of the signals from the promoter-proximal regions compared to those from the gene-body regions were evaluated using Fisher's exact test with Bonferroni's correction. For *S. cerevisiae*, the regions from the observed TSS to downstream 250bp (TSS to TSS+250 bp) were used as the promoter-proximal regions and those from TSS + 250bp to the annotated transcription end site (TES) were used to calculate the gene-body signals. For mESCs, the regions from 100bp downstream to 200bp upstream of the observed PRO-seq peaks were used as the promoter-proximal regions, and gene-body signal was measured from the regions of 1kb downstream of the observed PRO-seq peak to the annotated TES. A gene was identified as being paused if the *P* value was lower than 0.01 and as being not paused if the *P*

572 value was higher than 0.99, as determined using combined replicates. A gene exhibiting *P* value
573 < 0.05 or *P* value > 0.95 for both biological replicates was further assigned as a high-confidence
574 paused or high-confidence not paused gene, respectively. Pausing site was defined as the single-
575 nucleotide of the maximum PRO-seq read within the promoter-proximal regions, as described in
576 a recent study (*16*). If multiple sites with the maximum values were detected, the closest one to
577 the TSS was selected.
578 All average profiles centered on the indicated point in this work were generated using a
579 bootstrapped estimation. Briefly, 1,000 random gene sets were taken as each representing 10% of
580 the total genes, and the median and confidence intervals of each averaged subsample were
581 calculated. In the relevant figures, the thick line represents the median value and shaded regions
582 indicate the 12.5th and 87.5th percentiles.
583 **Analysis of publicly distributed ChIP-seq and MNase-seq data**
584 Raw sequencing reads of the indicated accession numbers were downloaded from NCBI GEO,
585 unless otherwise noted. For MNase-seq, raw reads were uniquely mapped to the *S. cerevisiae*
586 sacCer3 genome or to the *M. musculus* mm10 genome using Bowtie, which trimmed the 3' bases
587 to 36 bp (if the raw reads were longer than 36 bp), allowed two mismatches, and restricted the
588 maximum insert size to 200bp. BEDtools was used to covert the aligned BAM files to BED
589 formats. The BED files were then processed by iNPS (*63*) to determine the nucleosome positions.
590 Briefly, the "MainPeak" nucleosome that was the closest to either the observed TSS in *S.*
591 *cerevisiae* or the observed PRO-seq peak in mESCs was assigned as the +1 nucleosome. The +1
592 dyad was defined as the mid-point between the start and the end inflection, and 75 bp around the
593 +1 dyad was referred to as the +1 nucleosome position. To discard false-positive nucleosome
594 positions, nucleosomes that did not overlap the H3K4me3 ChIP-seq enrichment calculated from
595 the existing data (*34, 64*) were discarded, as previously reported (*16*). The Gaussian smoothing
596 values were used to process BigWig files for average profiles. For ChIP-seq, a combined genome

consisting of *S. cerevisiae* (sacCer3) and *S. pombe* (SpombeASMv2) was used, and unique reads from each genome were parsed for downstream analysis. MACS2 (*65*) was used to convert the aligned BAM files to bedGraph formats. For the spike-in control, the recorded coverage in the bedGraph file was normalized by the relative number of reads mapped to a *S. pombe* genome. The bedGraph files were converted to BigWig files by bedGraphToBigWig (*62*).

Statistical analysis

Statistical analyses were performed using R 3.6.3. In boxplots, whiskers represent 1.5 x interquartile range. *P* values for boxplot were calculated using stat_compare_means (method = "wilcox") function in ggpubr library. Symbols of ns, *, **, ***, **** represent $P > 0.05$, <= 0.05, <= 0.01, <= 0.001, <= 0.0001 respectively. *P* values for Venn diagram were calculated by hypergeometric distribution using phyper function in stats library.

References and Notes

1. L. Core, K. Adelman, Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* **33**, 960-982 (2019).
2. T. Wada, T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G. A. Hartzog, F. Winston, S. Buratowski, H. Handa, DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12**, 343-356 (1998).
3. Y. Yamaguchi, T. Takagi, T. Wada, K. Yano, A. Furuya, S. Sugimoto, J. Hasegawa, H. Handa, NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41-51 (1999).
4. L. J. Core, J. J. Waterfall, J. T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848 (2008).
5. H. Kwak, N. J. Fuda, L. J. Core, J. T. Lis, Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-953 (2013).
6. D. B. Mahat, H. Kwak, G. T. Booth, I. H. Jonkers, C. G. Danko, R. K. Patel, C. T. Waters, K. Munson, L. J. Core, J. T. Lis, Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**, 1455-1476 (2016).
7. L. S. Churchman, J. S. Weissman, Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368-373 (2011).
8. T. Nojima, T. Gomes, A. R. F. Grosso, H. Kimura, M. J. Dye, S. Dhir, M. Carmo-Fonseca, N. J. Proudfoot, Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526-540 (2015).
9. L. J. Core, J. J. Waterfall, D. A. Gilchrist, D. C. Fargo, H. Kwak, K. Adelman, J. T. Lis, Defining the status of RNA polymerase at promoters. *Cell Rep* **2**, 1025-1035 (2012).
10. T. Henriques, D. A. Gilchrist, S. Nechaev, M. Bern, G. W. Muse, A. Burkholder, D. C. Fargo, K.

Adelman, Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol Cell* **52**, 517-528 (2013).

11. C. S. Maxwell, W. S. Kruesi, L. J. Core, N. Kurhanewicz, C. T. Waters, C. L. Lewarch, I. Antoshechkin, J. T. Lis, B. J. Meyer, L. R. Baugh, Pol II docking and pausing at growth and stress genes in C. elegans. *Cell Rep* **6**, 455-466 (2014).
12. G. T. Booth, I. X. Wang, V. G. Cheung, J. T. Lis, Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Res* **26**, 799-811 (2016).
13. E. J. Grayhack, X. J. Yang, L. F. Lau, J. W. Roberts, Phage lambda gene Q antiterminator recognizes RNA polymerase near the promoter and accelerates it through a pause site. *Cell* **42**, 259-269 (1985).
14. S. S. Teves, C. M. Weber, S. Henikoff, Transcribing through the nucleosome. *Trends Biochem Sci* **39**, 577-586 (2014).
15. C. M. Weber, S. Ramachandran, S. Henikoff, Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell* **53**, 819-830 (2014).
16. Y. Aoi, E. R. Smith, A. P. Shah, E. J. Rendleman, S. A. Marshall, A. R. Woodfin, F. X. Chen, R. Shiekhattar, A. Shilatifard, NELF Regulates a Promoter-Proximal Step Distinct from RNA Pol II Pause-Release. *Mol Cell* **78**, 261-274.e265 (2020).
17. P. J. Skene, A. E. Hernandez, M. Groudine, S. Henikoff, The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *Elife* **3**, e02042 (2014).
18. J. Poli, S. M. Gasser, M. Papamichos-Chronakis, The INO80 remodeller in transcription, replication and repair. *Philos Trans R Soc Lond B Biol Sci* **372**, (2017).
19. M. Papamichos-Chronakis, S. Watanabe, O. J. Rando, C. L. Peterson, Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**, 200-213 (2011).
20. K. Yen, V. Vinayachandran, B. F. Pugh, SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. *Cell* **154**, 1246-1256 (2013).
21. M. Tramantano, L. Sun, C. Au, D. Labuz, Z. Liu, M. Chou, C. Shen, E. Luk, Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. *Elife* **5**, (2016).
22. C. Jeronimo, S. Watanabe, C. D. Kaplan, C. L. Peterson, F. Robert, The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations. *Mol Cell* **58**, 1113-1123 (2015).
23. M. Udugama, A. Sabri, B. Bartholomew, The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor. *Mol Cell Biol* **31**, 662-673 (2011).
24. N. Krietenstein, M. Wal, S. Watanabe, B. Park, C. L. Peterson, B. F. Pugh, P. Korber, Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* **167**, 709-721.e712 (2016).
25. Y. Xue, S. K. Pradhan, F. Sun, C. Chronis, N. Tran, T. Su, C. Van, A. Vashisht, J. Wohlschlegel, C. L. Peterson, H. T. M. Timmers, S. K. Kurdistani, M. F. Carey, Mot1, Ino80C, and NC2 Function Coordinately to Regulate Pervasive Transcription in Yeast and Mammals. *Mol Cell* **67**, 594-607.e594 (2017).
26. E. S. Choi, Y. Cheon, K. Kang, D. Lee, The Ino80 complex mediates epigenetic centromere propagation via active removal of histone H3. *Nat Commun* **8**, 529 (2017).
27. A. Lafon, S. Taranum, F. Pietrocola, F. Dingli, D. Loew, S. Brahma, B. Bartholomew, M. Papamichos-Chronakis, INO80 Chromatin Remodeler Facilitates Release of RNA Polymerase II from Chromatin for Ubiquitin-Mediated Proteasomal Degradation. *Mol Cell* **60**, 784-796 (2015).
28. J. Poli, C. B. Gerhold, A. Tosi, N. Hustedt, A. Seeber, R. Sack, F. Herzog, P. Pasero, K. Shimada, K. P. Hopfner, S. M. Gasser, Mec1, INO80, and the PAF1 complex cooperate to limit transcription replication conflicts through RNAPII removal during replication stress. *Genes Dev* **30**, 337-354 (2016).
29. A. Klein-Brill, D. Joseph-Strauss, A. Appleboim, N. Friedman, Dynamics of Chromatin and Transcription during Transient Depletion of the RSC Chromatin Remodeling Complex. *Cell Rep* **26**, 279-292.e275 (2019).

30. M. Morawska, H. D. Ulrich, An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* **30**, 341-351 (2013).
31. H. Kwak, J. T. Lis, Control of transcriptional elongation. *Annu Rev Genet* **47**, 483-508 (2013).
32. S. Kubik, M. J. Bruzzone, D. Challal, R. Dreos, S. Mattarocci, P. Bucher, D. Libri, D. Shore, Opposing chromatin remodelers control transcription initiation frequency and start site selection. *Nat Struct Mol Biol* **26**, 744-754 (2019).
33. S. Kubik, M. J. Bruzzone, D. Shore, Establishing nucleosome architecture and stability at promoters: Roles of pioneer transcription factors and the RSC chromatin remodeler. *Bioessays* **39**, (2017).
34. L. M. Soares, P. C. He, Y. Chun, H. Suh, T. Kim, S. Buratowski, Determinants of Histone H3K4 Methylation Patterns. *Mol Cell* **68**, 773-785.e776 (2017).
35. M. Radman-Livaja, O. J. Rando, Nucleosome positioning: how is it established, and why does it matter? *Dev Biol* **339**, 258-266 (2010).
36. C. Jiang, B. F. Pugh, Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* **10**, 161-172 (2009).
37. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).
38. D. N. Bagchi, A. M. Battenhouse, D. Park, V. R. Iyer, The histone variant H2A.Z in yeast is almost exclusively incorporated into the +1 nucleosome in the direction of transcription. *Nucleic Acids Res* **48**, 157-170 (2020).
39. A. Tosi, C. Haas, F. Herzog, A. Gilmozzi, O. Berninghausen, C. Ungewickell, C. B. Gerhold, K. Lakomek, R. Aebersold, R. Beckmann, K. P. Hopfner, Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154**, 1207-1219 (2013).
40. W. Yao, S. L. Beckwith, T. Zheng, T. Young, V. T. Dinh, A. Ranjan, A. J. Morrison, Assembly of the Arp5 (Actin-related Protein) Subunit Involved in Distinct INO80 Chromatin Remodeling Activities. *J Biol Chem* **290**, 25700-25709 (2015).
41. W. Yao, D. A. King, S. L. Beckwith, G. J. Gowans, K. Yen, C. Zhou, A. J. Morrison, The INO80 Complex Requires the Arp5-Ies6 Subcomplex for Chromatin Remodeling and Metabolic Regulation. *Mol Cell Biol* **36**, 979-991 (2016).
42. S. M. Vos, L. Farnung, M. Boehning, C. Wigge, A. Linden, H. Urlaub, P. Cramer, Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607-612 (2018).
43. F. X. Chen, A. R. Woodfin, A. Gardini, R. A. Rickels, S. A. Marshall, E. R. Smith, R. Shiekhattar, A. Shilatifard, PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* **162**, 1003-1015 (2015).
44. L. Hou, Y. Wang, Y. Liu, N. Zhang, I. Shamovsky, E. Nudler, B. Tian, B. D. Dynlacht, Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proc Natl Acad Sci U S A* **116**, 14583-14592 (2019).
45. H. Ehara, T. Kujirai, Y. Fujino, M. Shirouzu, H. Kurumizaka, S. I. Sekine, Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. *Science* **363**, 744-747 (2019).
46. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
47. S. Wu, Y. Shi, P. Mulligan, F. Gay, J. Landry, H. Liu, J. Lu, H. H. Qi, W. Wang, J. A. Nickoloff, C. Wu, Y. Shi, A YY1-INO80 complex regulates genomic stability through homologous recombination-based repair. *Nat Struct Mol Biol* **14**, 1165-1172 (2007).
48. X. Shen, G. Mizuguchi, A. Hamiche, C. Wu, A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541-544 (2000).
49. B. Lai, W. Gao, K. Cui, W. Xie, Q. Tang, W. Jin, G. Hu, B. Ni, K. Zhao, Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* **562**, 281-285 (2018).

50. J. Li, Y. Liu, H. S. Rhee, S. K. Ghosh, L. Bai, B. F. Pugh, D. S. Gilmour, Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol Cell* **50**, 711-722 (2013).
51. T. Narita, Y. Yamaguchi, K. Yano, S. Sugimoto, S. Chanarat, T. Wada, D. K. Kim, J. Hasegawa, M. Omori, N. Inukai, M. Endoh, T. Yamada, H. Handa, Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Mol Cell Biol* **23**, 1863-1873 (2003).
52. Y. Yamaguchi, H. Shibata, H. Handa, Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim Biophys Acta* **1829**, 98-104 (2013).
53. S. Jimeno-González, M. Ceballos-Chávez, J. C. Reyes, A positioned +1 nucleosome enhances promoter-proximal pausing. *Nucleic Acids Res* **43**, 3068-3078 (2015).
54. T. Kujirai, H. Ehara, Y. Fujino, M. Shirouzu, S. I. Sekine, H. Kurumizaka, Structural basis of the nucleosome transition during RNA polymerase II passage. *Science* **362**, 595-598 (2018).
55. J. Fitz, T. Neumann, R. Pavri, Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. *Embo j* **37**, (2018).
56. G. Moyle-Heyrman, T. Zaichuk, L. Xi, Q. Zhang, O. C. Uhlenbeck, R. Holmgren, J. Widom, J. P. Wang, Chemical map of Schizosaccharomyces pombe reveals species-specific features in nucleosome positioning. *Proc Natl Acad Sci U S A* **110**, 20158-20163 (2013).
57. A. B. Lantermann, T. Straub, A. Strålfors, G. C. Yuan, K. Ekwall, P. Korber, Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nat Struct Mol Biol* **17**, 251-257 (2010).
58. G. T. Booth, P. K. Parua, M. Sansó, R. P. Fisher, J. T. Lis, Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. *Nat Commun* **9**, 543 (2018).
59. E. Kopylova, L. Noé, H. Touzet, SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211-3217 (2012).
60. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
61. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
62. W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207 (2010).
63. W. Chen, Y. Liu, S. Zhu, C. D. Green, G. Wei, J. D. Han, Improved nucleosome-positioning algorithm iNPS for accurate nucleosome positioning from sequencing data. *Nat Commun* **5**, 4909 (2014).
64. H. Marks, T. Kalkan, R. Menafra, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A. F. Stewart, A. Smith, H. G. Stunnenberg, The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590-604 (2012).
65. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

Acknowledgments

General: We would like to thank Nir Friedman for providing the AID strain. We acknowledge

Gregory T. Booth and John T. Lis for making publicly available the custom script for PRO-seq

779 analysis in GitHub. We also thank the Lee lab members, H. Jo for obtaining the AID strain from
780 N. Friedman, H. Yang for assistance with mESC culture and Y. Chun for critically reading the
781 paper. **Funding:** This work was supported by a National Research Foundation (NRF) of Korea
782 Grant funded by the Ministry of Science and ICT (MSIT) (2018R1A5A1024261, SRC), and the
783 Collaborative Genome Program for Fostering New Post-Genome Industry of the NRF funded by
784 the MSIT (2018M3C9A6065070). **Author Contributions:** Y.C. and D.L. conceived the study
785 and designed the experiments. Y.C. performed all the PRO-seq experiments and computational
786 analysis in this study. Y.C. and D.L. wrote the paper. S.H. conducted the mESC cultures and
787 siRNA transfections. T.K. supported the computational analysis. **Competing interests:** We have
788 no conflicts of interest to disclose.
789
790
791
792
793
794
795
796
797
798
799
800
801
802

Figures and Tables

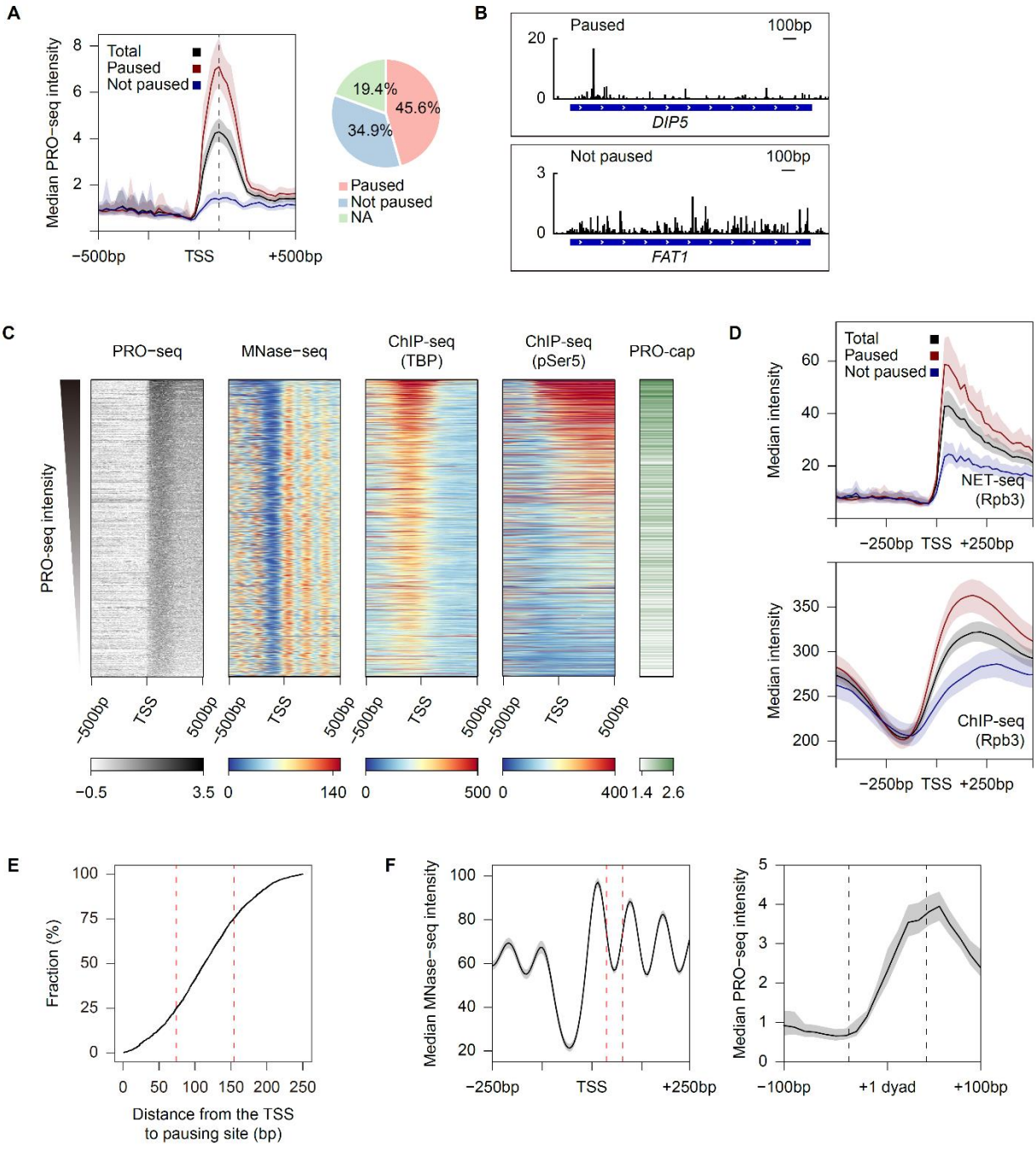


Fig. 1. PRO-seq reveals a non-uniform distribution of transcription elongation genome-wide in *S. cerevisiae* resembling promoter-proximal pausing in metazoans.

(**A**) Average profile showing the median PRO-seq intensities of paused and not paused genes. Pie chart indicates the percentage of paused (N = 2,599) and not paused (N = 1,990) genes among the total filtered protein-coding (N = 5,315) genes. NA indicate genes which were not classified as either being paused or not paused. (**B**) Genome browser view of PRO-seq signals for representative paused or not paused genes. (**C**) Heatmaps showing PRO-seq, PRO-cap, existing MNase-seq (GSM3304635), TBP ChIP-seq (GSM3452564) and pSer5 ChIP-seq (GSM3452562) signals. Genes were sorted by the PRO-seq signals at their promoter-proximal regions. The PRO-cap intensity reflected read counts within 250bp upstream and downstream of the TSS. (**D**) Average Rpb3 NET-seq (GSM617027) and ChIP-seq (GSM2813906) profiles centered on the TSS. Processed sequencing files downloaded from the NCBI Gene Expression Omnibus (GEO) were used for this analysis. (**E**) Cumulative curve analyzing the distance from the TSS to pausing site for paused genes. The red dotted lines represent the 25th (74 bp) and 75th (154 bp) percentiles. (**F**) Average profiles showing median MNase-seq intensity (GSM3304635) at the TSS (Left) and the median PRO-seq intensity at the +1 nucleosome dyad (Right). The +1 dyad was defined by the improved nucleosome-positioning algorithm, iNPS from an existing MNase-seq data (GSM3304635) (*63*). The two red dotted lines represent the 25th (74 bp) and 75th (154 bp) percentiles of the pausing site (Left) and the two black dotted lines represent the position of the +1 nucleosome (Right; 75 bp upstream and downstream of the +1 dyad).

All PRO-seq data were generated from auxin-untreated Ino80-AID cells, and combined biological replicates were used. For average profiles, medians reflect either the 20-bp bin (PRO-seq) or the 10-bp bin (MNase-seq). For heatmaps, signals reflect the 10-bp bin around the indicated site. The PRO-seq, PRO-cap, and ChIP-seq intensities are presented in spike-in-normalized reads per million. MNase-seq data are presented in Gaussian smoothing-normalized reads per million.

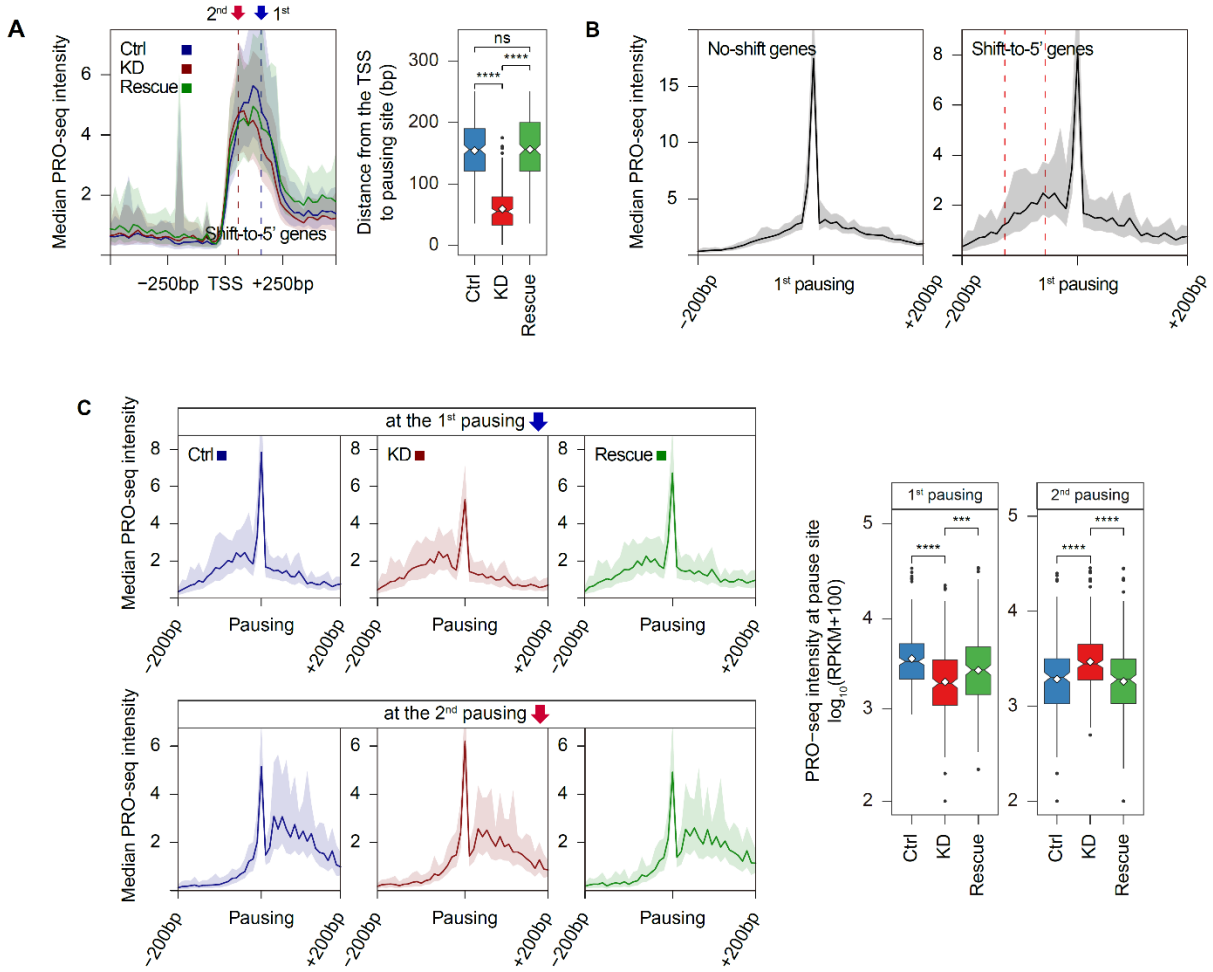


Fig. 2. Pausing-site determination by Ino80p occurs at genes with alternative pausing sites. (**A**) Average profile indicating median PRO-seq intensities in Ino80-AID cells treated with auxin for 0 hr (Ctrl) and 3 hrs (KD) and rescued for 3 hrs after auxin removal (Rescue) for shift-to-5' genes (N = 221). The arrows and below dotted lines represent the median of the 1st (blue, 55bp) and 2nd (red, 156bp) pausing sites. Medians reflect 20-bp bins. Boxplot shows the distance between the TSS and pausing site. (**B**) Average profiles representing median PRO-seq intensities in the control sample for either no-shift genes (N = 1,211) or shift-to-5' genes (N = 221). The two red dotted lines (Right) represent the 25th (-56 bp) and 75th (-126 bp) percentiles of the 2nd pausing site relative to the 1st pausing site. Medians reflect 10-bp bins. (**C**) Average profiles displaying median PRO-seq intensities around either the 1st or the 2nd pausing site. Medians reflect 10-bp

bins. Boxplot shows the PRO-seq intensity at the indicated single nucleotide in $\log_{10}$(RPKM+100).

All data was generated using combined biological replicates. PRO-seq intensity was calculated using spike-in-normalized reads per million. Asterisks represent statistically significant differences, as calculated using the Wilcoxon test.

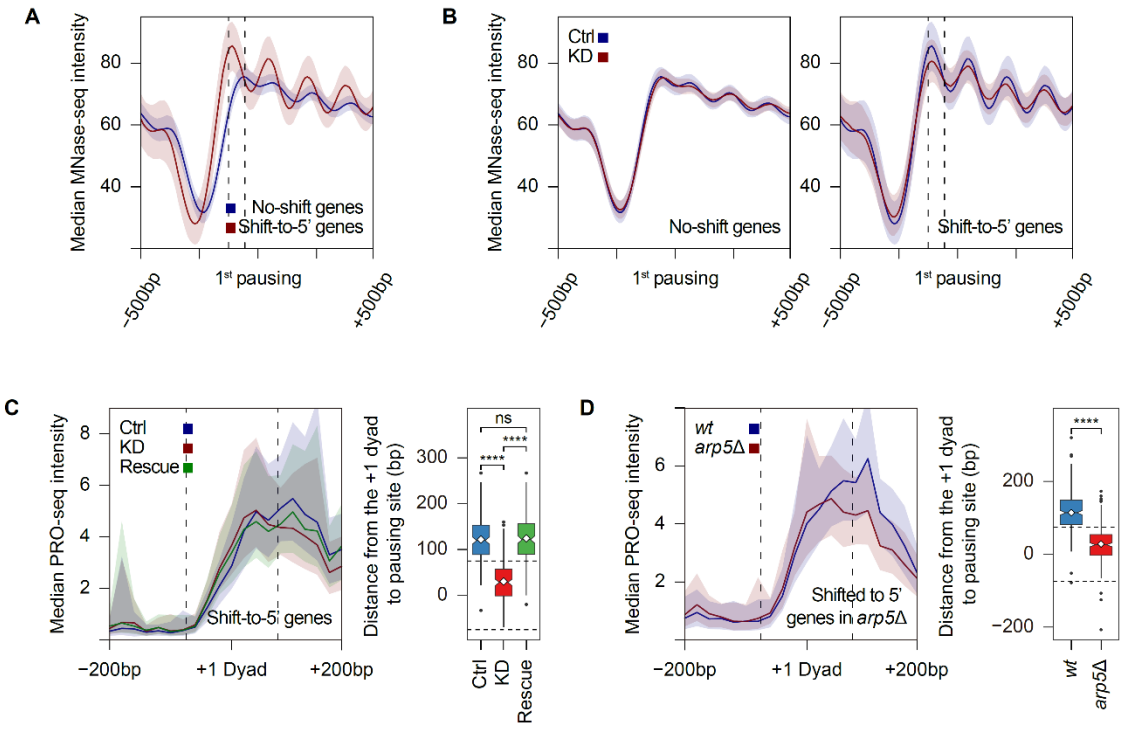
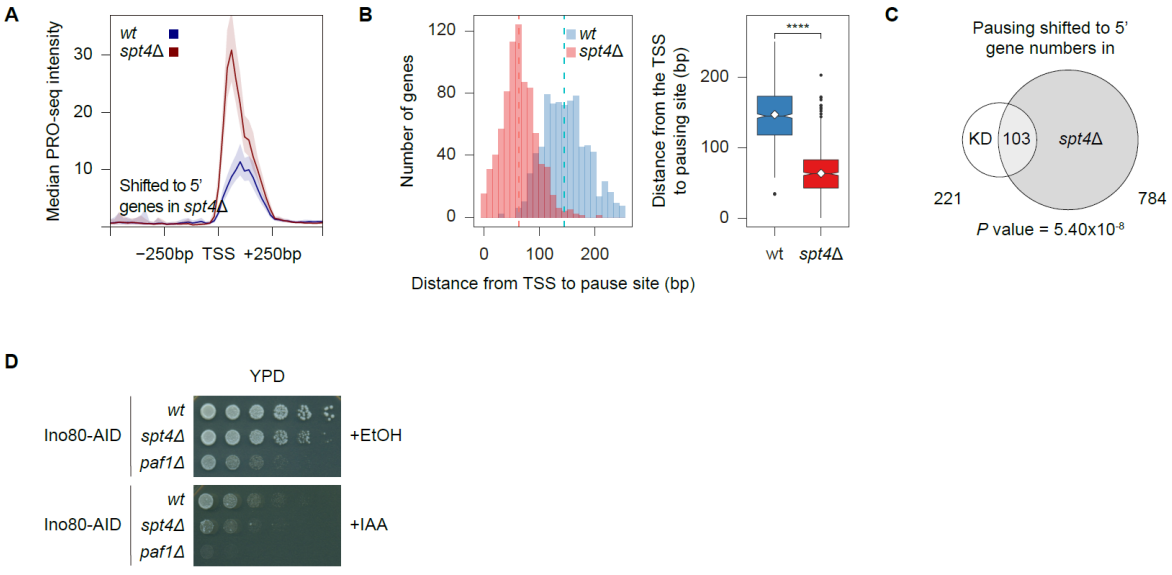


Fig. 3. Chromatin remodeling activity of Ino80p in nucleosome positioning around the 1st pausing site is critical for proper localization of pausing.

(**A**) Average profiles of median MNase-seq intensities in control samples (GSM3304635) for no-shift genes (N = 1,211) and shift-to-5' genes (N = 221) centered on the 1st pausing site. (**B**) Average profiles of median MNase-seq intensities in control sample (GSM3304635) and Ino80p-KD sample (GSM3304637) for no-shift genes (Left) and shift-to-5' genes (Right) at the 1st pausing site. For Fig. 3, A and B, the two dotted lines indicate the 25th (-56 bp) and 75th (-126 bp)

855 percentiles of the 2nd pausing site relative to the 1st pausing site. (**C** and **D**) Average profiles of
856 median PRO-seq intensity for the indicated samples around the +1 dyad (defined in Fig. 1F) and
857 boxplots of the distance from the +1 dyad to pausing site. Note that only nucleosomes overlapped
858 with H3K4me3 ChIP-seq enrichment (GSM2507874) were used, in an effort to exclude false-
859 positive nucleosomes. For (**C**), shift-to-5' genes upon Ino80p-KD (N = 160) were used for
860 analyses. For (**D**), pausing site shifted toward 5' genes in *arp5*Δ (N = 264) were used for
861 analyses. The two dotted lines in profiles and boxplots represented the position of the +1
862 nucleosome (75 bp upstream and downstream of +1 dyad).
863 All PRO-seq data were generated using combined biological replicates. PRO-seq and MNase-seq
864 intensity were calculated using spike-in-normalized reads per million and Gaussian smoothing-
865 normalized reads per million, respectively. For average profiles, medians reflect either the 20-bp
866 bin (PRO-seq) or the 10-bp bin (MNase-seq). Asterisks represented statistically significant
867 differences, as calculated using the Wilcoxon test.
868

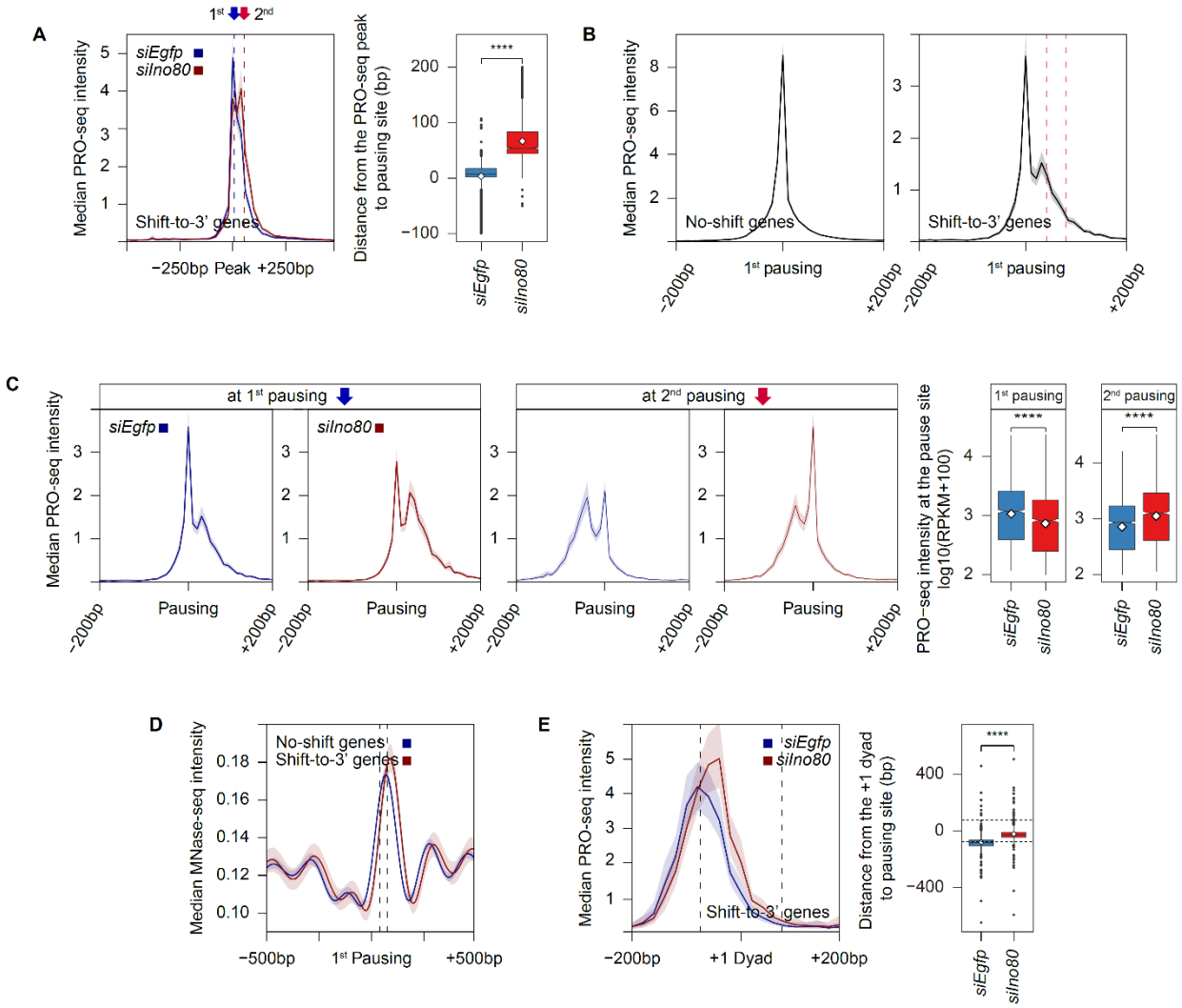


869

Fig. 4. Loss of the conserved pausing factor Spt4p shifts pausing upstream in a manner similar to Ino80p knockdown.

(**A**) Average profiles representing median PRO-seq intensities of *wt* and *spt4*Δ. Medians reflect 20-bp bins. (**B**) Histograms and boxplots demonstrating the distance from the TSS to pausing site in *wt* and *spt4*Δ. Asterisks represent statistically significant differences, as calculated using the Wilcoxon test. For (**A**) and (**B**), genes whose pausing sites were shifted toward 5' (> 30bp) in *spt4*Δ (N = 784) were used for analyses. (**C**) Venn diagram depict overlap between shift-to-5' genes upon Ino80p-KD (N = 221) and genes showing an upstream shift (> 30bp) of RNAPII pausing in *spt4*Δ (N = 784). *P* value was calculated using the hypergeometric distribution. (**D**) Spotting assays analyzing genetic interactions between *INO80* and *SPT4* or *PAF1*. Cells were spotted onto YPD plates containing either ethanol or auxin (0.5mM) with 5-fold serial dilutions and incubated at 30°C.

All PRO-seq data were generated using combined biological replicates. PRO-seq intensities were calculated using spike-in-normalized reads per million.



885

886 **Fig. 5. The ability of Ino80 to regulate RNAPII pausing site is conserved in mESCs.**

887 (**A**) Average profiles representing median PRO-seq intensities in mESCs treated with either

888 *siEgfp* or *siIno80* for 48 hrs at shift-to-3' genes (N = 2,324) around the PRO-seq peak. Medians

889 reflect 20-bp bins. The arrows and below dotted lines represent the median of the 1st (blue, 7bp)

890 and 2nd (red, 54bp) pausing sites. Boxplot showed the distance between pausing site and the PRO-

891 seq peak. (**B**) Average profiles of median PRO-seq intensities in mESCs treated with *siEgfp* at

892 shift-to-3' genes (N = 2,324) or no-shift genes (N = 16,903). Medians reflect 10-bp bins. The two

893 red dotted lines (Right) represent the 25th (39bp) and 75th (76bp) percentiles of the 2nd pausing site

relative to the 1st pausing site. (**C**) Average profiles displaying median PRO-seq intensities around the 1st or 2nd pausing site. Medians reflect 10-bp bins. Boxplots showed the PRO-seq intensities at the indicated single nucleotides in $\log_{10}$(RPKM+100). (**D**) Average profiles of median MNase-seq intensities obtained from untreated mESCs (GSM2906312 and GSM2906313) for no-shift genes and shift-to-3' genes at the 1st pausing site. The two black dotted lines indicate the 25th (39bp) and 75th (76bp) percentiles of the 2nd pausing site relative to the 1st pausing site. Medians reflect 10-bp bins. (**E**) Average profiles of median PRO-seq intensities around the +1 dyad, which was determined by MNase-seq used in Fig. 5D, for shift-to-3' genes. The two dotted lines represent the position of the +1 nucleosome (75 bp upstream and downstream of the +1 dyad). Only nucleosomes overlapped with H3K4me3 enrichment calculated from an existing ChIP-seq data (GSM590111) (*64*) were used, in an effort to exclude false-positive nucleosomes (N = 629). Boxplot indicates the distance from the +1 dyad to pausing site. Asterisks represent statistically significant differences, as calculated using the Wilcoxon test. All PRO-seq data were generated using combined biological replicates. PRO-seq and MNase-seq intensities were calculated using spike-in-normalized reads per million and Gaussian smoothing-normalized reads per million, respectively.

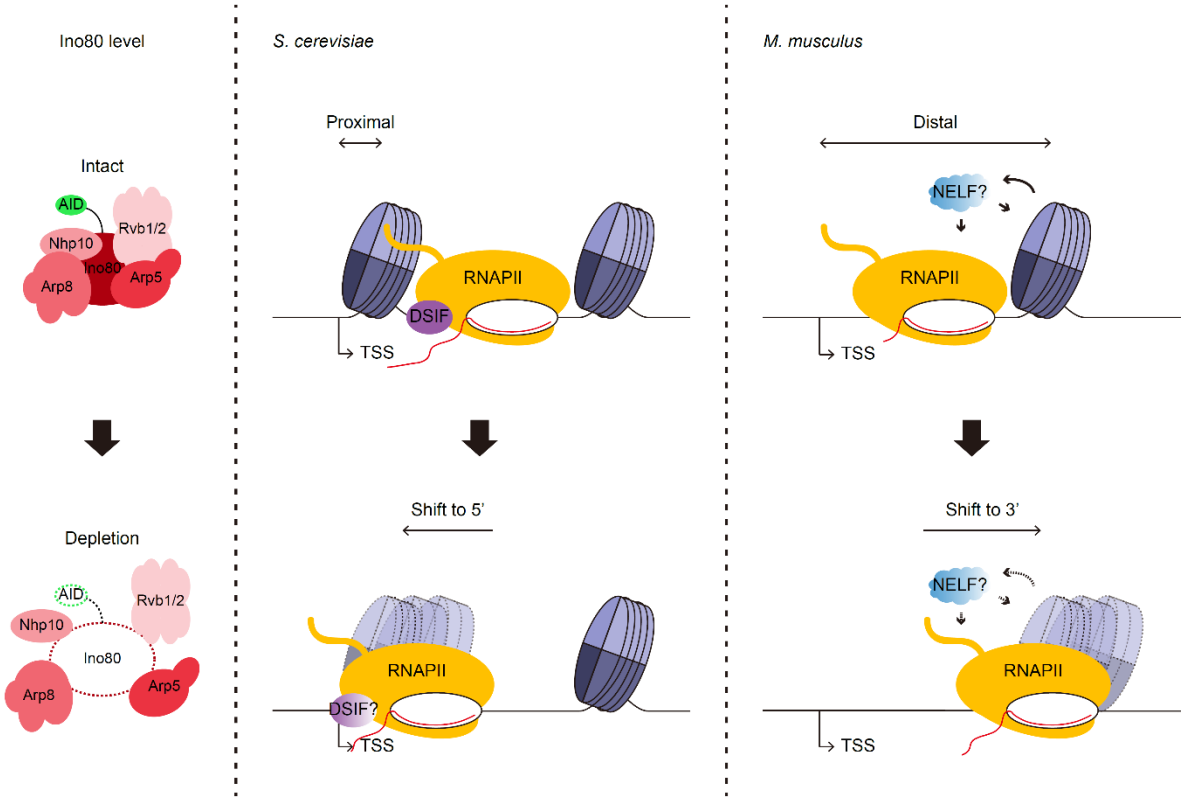


Fig. 6. Model of the function of Ino80 in mediating alternative pausing site determination. Model depicts a conserved regulatory role of Ino80 in proper localization of RNAPII pausing through its activity to modulate the +1 nucleosome in budding yeast and mESCs. See Discussion for details.

Supplementary Materials

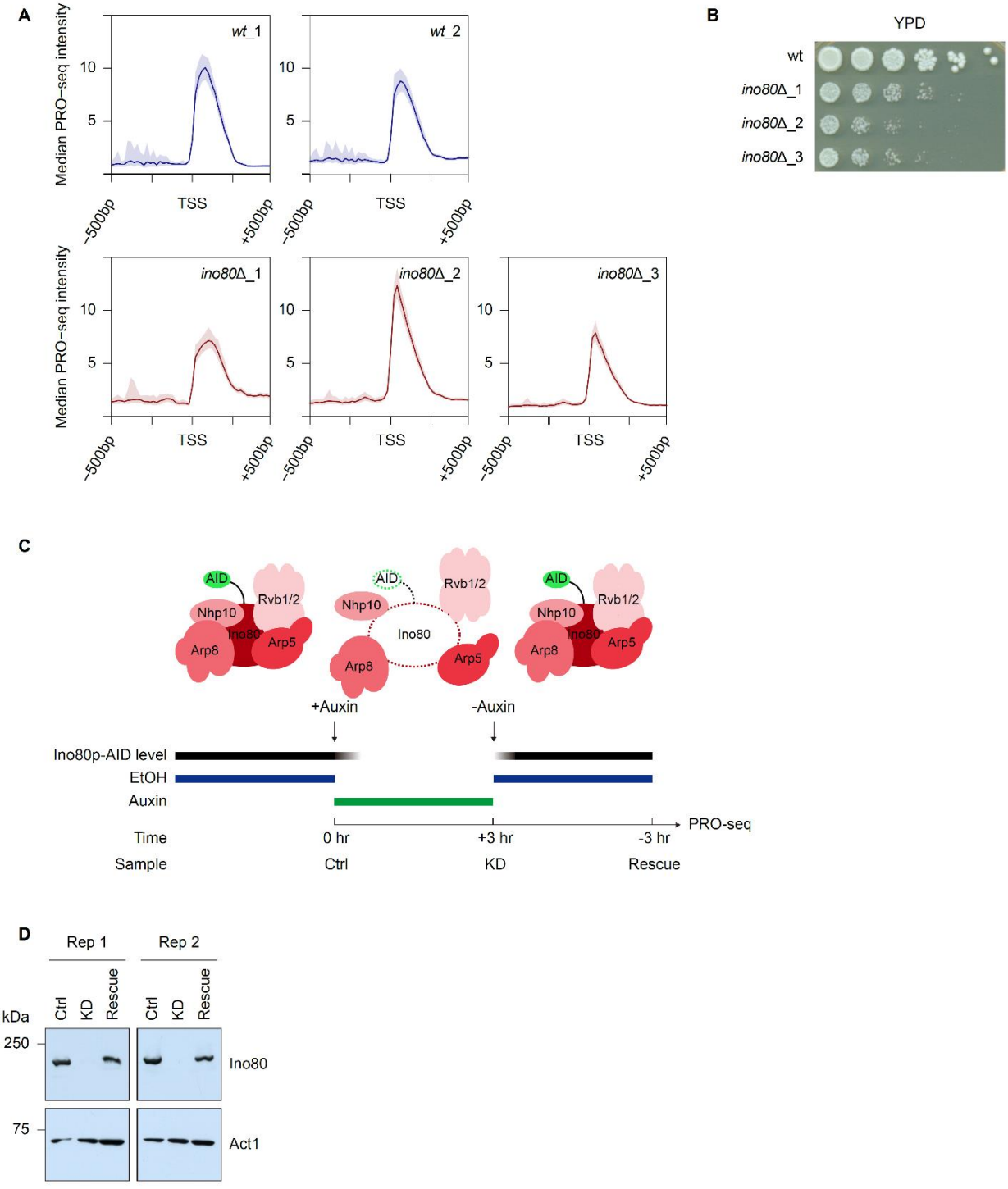


fig. S1. Loss of Ino80p causes variation in fitness and the promoter-proximal PRO-seq pattern.

928 (**A**) Average profiles indicating the median PRO-seq intensities (Spike-in-normalized reads per
929 million) in *wt* and *ino80*Δ. Medians reflect 20-bp bins. (**B**) Spotting assays assessing growth
930 variation in *ino80*Δ cells of the same batch used to generate the triplicate PRO-seq data presented
931 in fig. S1A. Cells were spotted onto YPD plates with 5-fold serial dilutions and incubated at
932 30°C. (**C**) Schematic illustration of experimental outline. Ino80-AID cells (*29*) were grown to
933 mid-log phase in YPD containing ethanol (Ctrl). The ethanol was washed away and the cells were
934 incubated with auxin (0.5mM) for 3 hrs (KD). The auxin was washed away and the cells were
935 incubated without auxin for additional 3 hrs (Rescue). PRO-seq was performed at each indicated
936 time point. (**D**) Western blot of whole-cell lysates from Ino80-AID cells under the Ctrl, KD, and
937 Rescue conditions showed that Ino80p is almost completely degraded after 3 hrs of auxin
938 incubation and restored after 3 hrs of auxin withdrawal. Western blot analysis was performed
939 against Act1p served as a loading control.
940

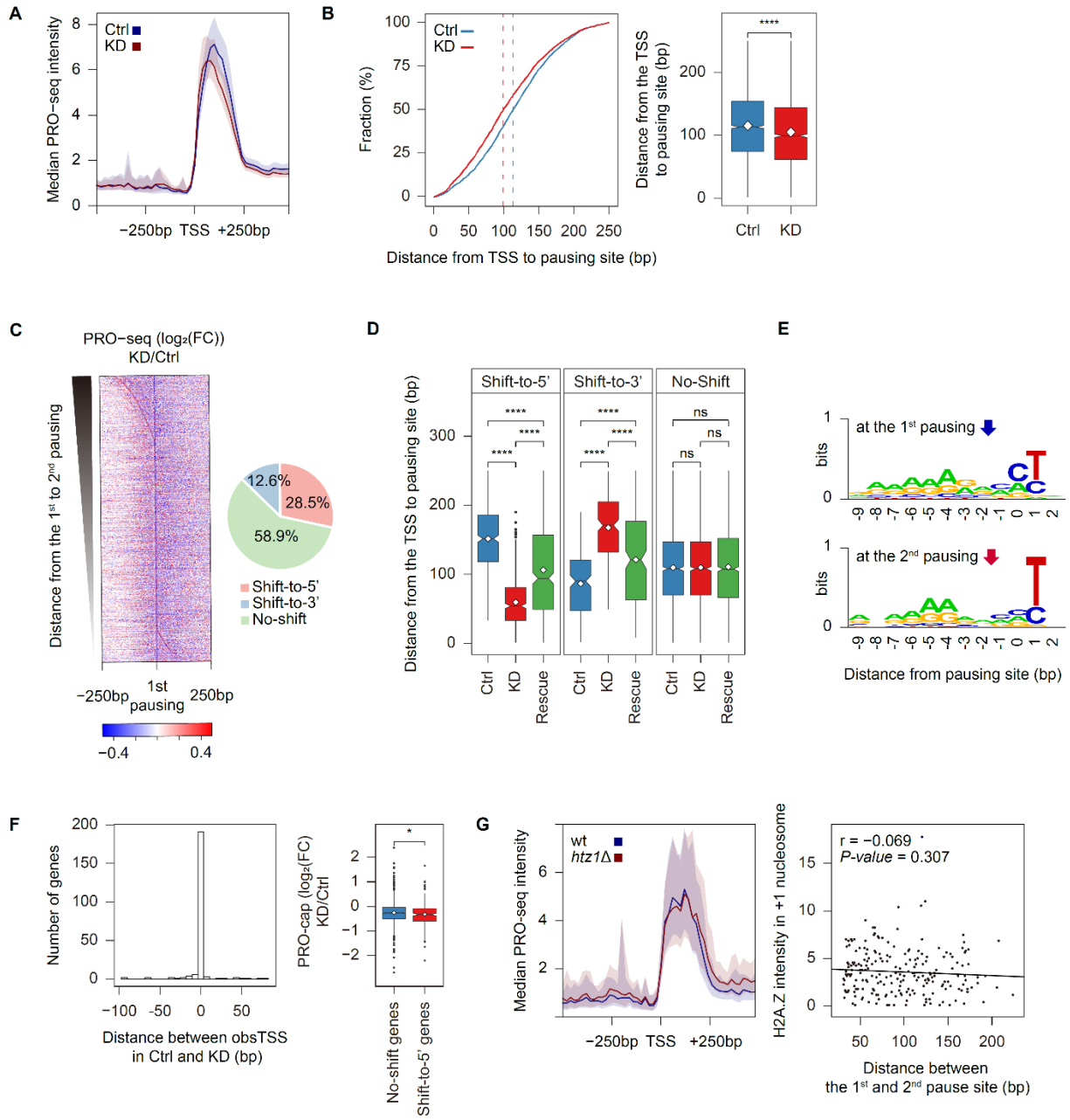


fig. S2. Ino80p knockdown transits the RNAPII pausing from the 1st to the 2nd pausing site and it is independent of both TSS usage and H2A.Z^{Htz1}.

(**A**) Average profile indicating median PRO-seq intensities in Ino80-AID cells treated with auxin for 0 hr (Ctrl) and 3 hrs (KD), as assessed at the genes defined as being high-confidence paused (N = 2,599). (**B**) Cumulative curve and boxplot demonstrate the distance from the TSS to pausing

site for paused genes in the control and knockdown conditions. The dotted lines represent the median value (113bp for Ctrl and 99bp for KD). (**C**) Heatmap representing the PRO-seq signal upon Ino80p-KD as a $\log_2$ fold change relative to the control sample at the 1st pausing site. Genes were sorted by the distance to the 2nd pausing site. Pie chart shows the percentage of genes whose pausing site was shifted toward 5' (N = 742), toward 3' (N = 327), or not shifted (N = 1,530) among total paused genes (N = 2,599) upon Ino80p-KD. (**D**) Boxplot indicating the distance from the TSS to pausing site for genes shifted more than 30bp toward upstream (N = 442) or downstream (N = 178), or for not shifted genes (N = 1,530). Upon rescue of Ino80p, the pausing site was restored to the 1st pausing site in both types of shifted genes. (**E**) Sequence logos around either the 1st or the 2nd pausing site were generated using WebLogo (*37*). (**F**) Histogram analyzing the distance between the observed TSS for shift-to-5' genes under control and Ino80p-KD conditions (Left). A large fraction of genes (189 out of 221) showed no change in their major initiation site upon knockdown. Boxplot demonstrates the $\log_2$ fold change in the PRO-cap signal 100bp around the TSS (in RPKM) upon Ino80p-KD relative to control for shift-to-5' genes or no-shift genes (Right). (**G**) Average profiles indicating the median PRO-seq intensities in *wt* (BY4741) and *htz1*Δ cells (Left). Scatter plot displays the correlation (as Pearson's r) between the H2A.Z^{Htz1} intensity (IP/input in reads per million) at the +1 nucleosome and the distance of the pausing site shift (Right). For analysis of H2A.Z intensity, we used an existing MNase-ChIP-seq data set (GSM2790633, GSM2790634, GSM2790635 and GSM2790636).

All PRO-seq data were generated using combined biological replicates. The PRO-seq and PRO-cap intensities were calculated using spike-in-normalized reads per million. For average profiles, medians reflect 20-bp bins. Asterisks represent statistically significant differences, as calculated using the Wilcoxon test.

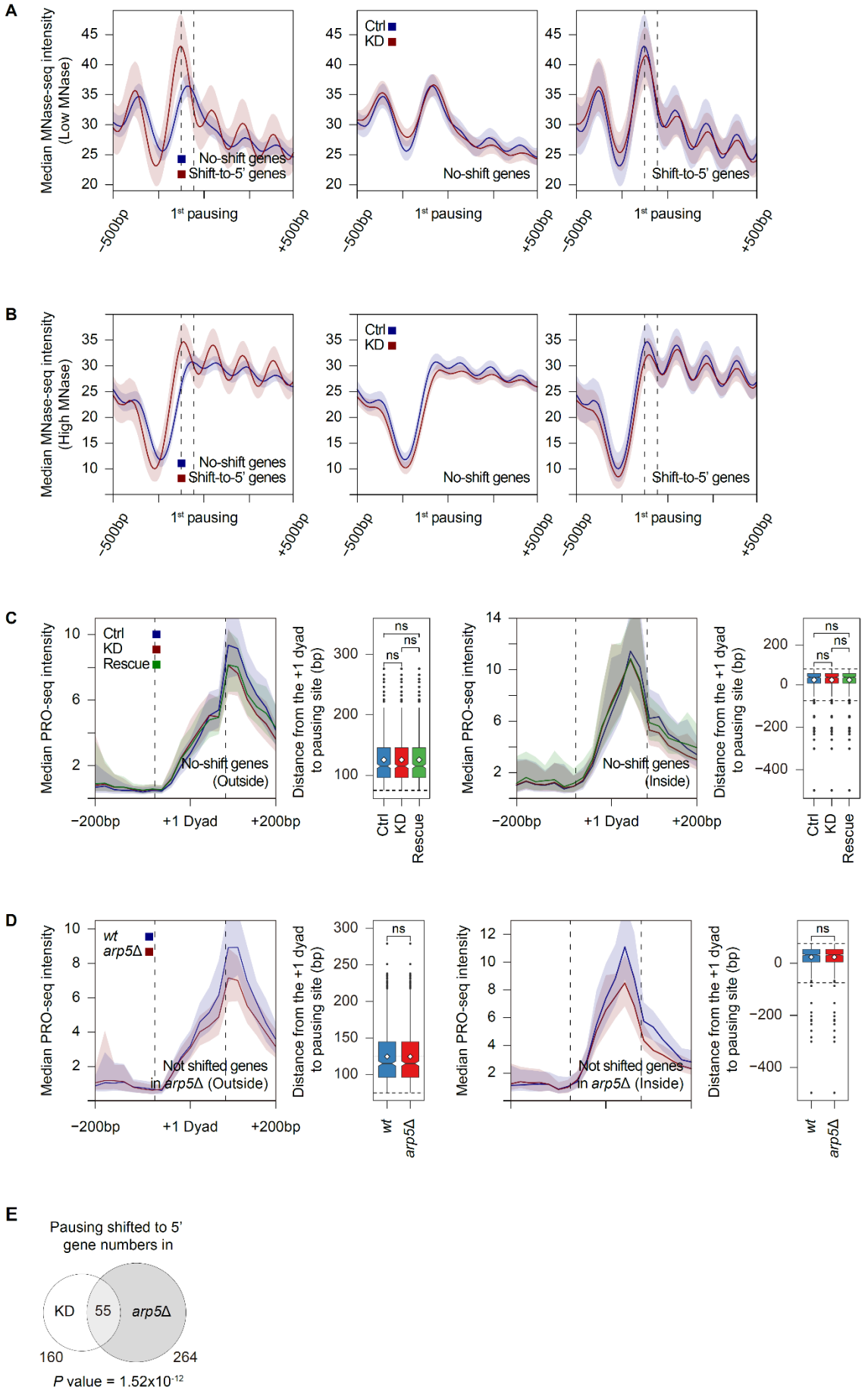


fig. S3. MNase-seq analyses at the pausing site and PRO-seq analyses at the +1 dyad.

(**A**) Average profiles of median MNase-seq intensities in control samples (GSM3177776) for no-shift genes (N = 1,211) and shift-to-5' genes (N = 221) centered on the 1st pausing site (Left). Average profiles of median MNase-seq intensities in control sample (GSM3177776) and Ino80p-KD sample (GSM3177778) for no-shift genes (Middle) and shift-to-5' genes (Right) at the 1st pausing site. (**B**) Same as (**A**). Instead, GSM3177777 was used as control samples and GSM3177779 was used as Ino80p-KD samples. For fig. S3, A and B, the two dotted lines indicate the 25th (-56 bp) and 75th (-126 bp) percentiles of the 2nd pausing site relative to the 1st pausing site. (**C** and **D**) Average profiles of median PRO-seq intensities for the indicated samples around the +1 dyad (defined in Fig. 1F) and boxplots indicate the distance from the +1 dyad to pausing site. Note that only nucleosomes overlapped with H3K4me3 ChIP-seq enrichment (GSM2507874) were assigned to the +1 nucleosome, in an effort to exclude false-positive nucleosomes. No-shift genes upon Ino80p-KD for (**C**) and not shifted genes in *arp5*Δ for (**D**) were divided into two groups, namely those exhibiting a pausing site either outside (N = 463 for Ino80p-KD and N = 518 for *arp5*Δ; Left) or inside (N = 401 for Ino80p-KD and N = 603 for *arp5*Δ; Right) of the +1 nucleosome, which allowed us to more clearly distinguish the location of the PRO-seq peak. The two dotted lines represent the position of the +1 nucleosome (75 bp upstream and downstream of the +1 dyad). Asterisks represent statistically significant differences, as calculated using the Wilcoxon test. (**E**) Overlap between shift-to-5' genes upon Ino80p-KD (N = 160) and genes showing an upstream shift (> 30bp) of RNAPII pausing in *arp5*Δ (N = 264). *P* value was calculated using the hypergeometric distribution.

All PRO-seq data were generated using combined biological replicates. PRO-seq and MNase-seq intensities were calculated using either spike-in-normalized reads per million or gaussian smoothing-normalized reads per million, respectively. For average profiles, medians reflect either a 20-bp bin (PRO-seq) or a 10-bp bin (MNase-seq).

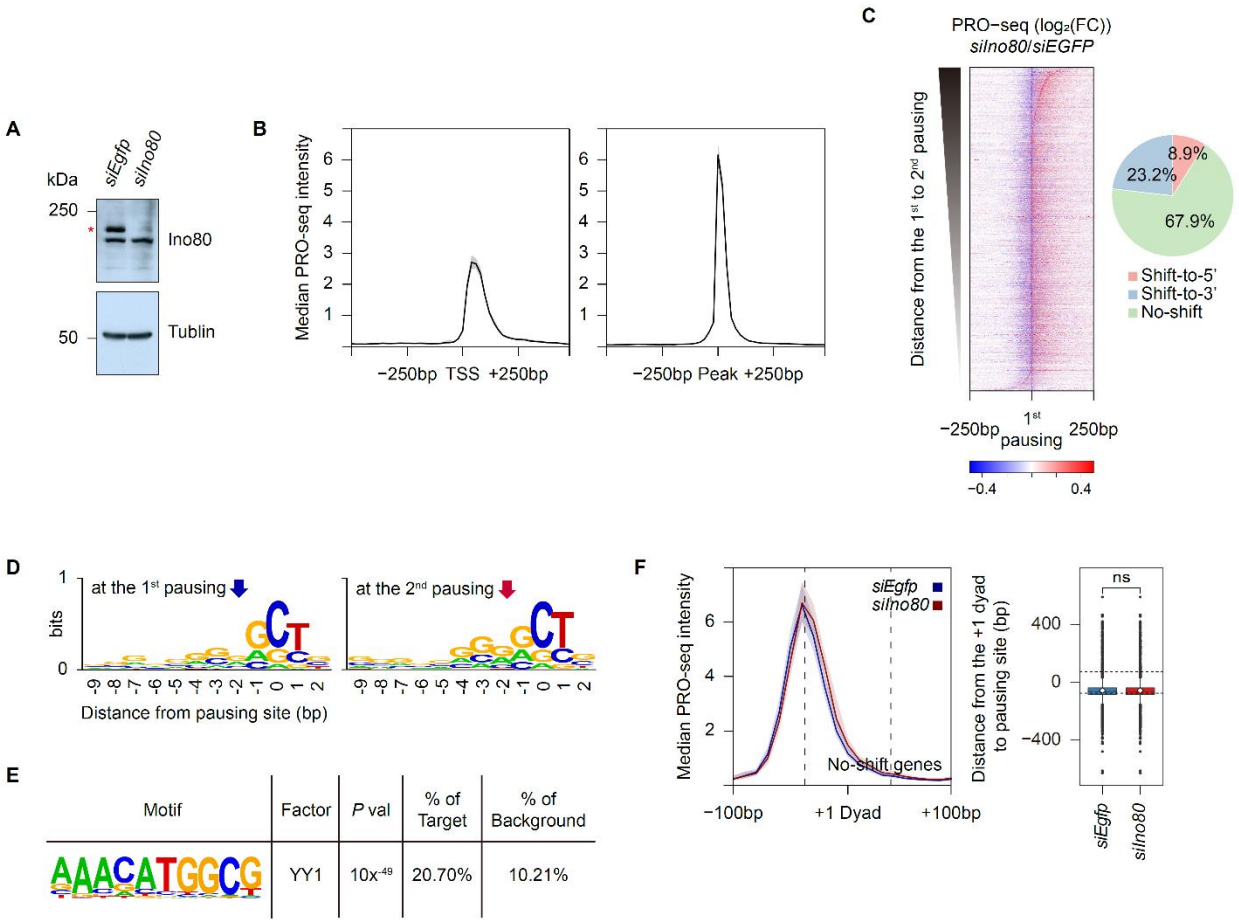


fig. S4. Ino80 knockdown yields a conserved defect of RNAPII pausing in mESCs.

(**A**) Western blot analysis of whole-cell lysates from mESCs treated with *siEgfp* or *siIno80* for 48 hrs shows that Ino80 was almost completely and specifically degraded in the presence of *siIno80* (The red asterisk). Tubulin was detected as a loading control. (**B**) Average profile of median PRO-seq intensities in mESCs treated with *siEgfp* centered on either the annotated TSS or the observed PRO-seq peak. (**C**) Heatmaps representing PRO-seq signals at the 1st pausing site in mESCs treated with *siIno80*, given as a $\log_2$ fold change relative to that in mESCs treated with *siEgfp*. Genes were sorted by the distance to the 2nd pausing site. Pie chart shows the percentage of genes whose pausing sites are shifted toward 5' (N = 2,206), toward 3' (N = 5,767), or not shifted (N = 16,903). (**D**) Sequence logos around either the 1st or 2nd pausing site were generated

using WebLogo (*37*). (**E**) *De novo* Motif analysis using the findMotifsGenome.pl program of HOMER (*46*) found that the YY1 motif was significantly enriched at the promoter regions of shift-to-3' genes (N = 2,324) compared to no-shift genes (N = 16,903). The sequences 250bp upstream and downstream of the observed PRO-seq peak were used as the promoter regions. (**F**) Average profiles of the median PRO-seq intensity around the +1 dyad, which was determined by MNase-seq used in Fig. 5D, for no-shift genes. The two dotted lines represent the positions of the +1 nucleosome (75 bp upstream and downstream of the +1 dyad). Note that only nucleosomes with H3K4me3 ChIP-seq enrichment (GSM590111) were used, in an effort to exclude false-positive nucleosomes (N = 4,564). Boxplot indicates the distance from the +1 dyad to pausing site.

All PRO-seq data were generated using combined biological replicates. PRO-seq intensity was calculated in reads per million. For average profiles, medians reflect 20-bp. For heatmaps, the signals reflect a 10-bp bin around the indicated site. Asterisks represent statistically significant differences, as calculated using the Wilcoxon test.

Experiment	Sample[a]	Total reads	Mapped to rRNA	Uniquely mapped to genome	Spike-in genome	Norm. factor[b]	Spearman correlation (Rho)[c]	
							Promoter proximal[d]	Whole genes
PRO-seq	SC_Ino80-AID_Ctrl_rep1	26142062	11900913	5501869	322452	0.94	0.985	0.992
	SC_Ino80-AID_Ctrl_rep2	25041456	11493690	5282796	322219	0.94		
	SC_Ino80-AID_KD_rep1	25713790	11192062	5434019	341750	1.00	0.984	0.990
	SC_Ino80-AID_KD_rep2	22761013	10326913	4624242	324736	0.95		
	SC_Ino80-AID_Rescue_rep1	24015944	11196976	4997072	253963	0.74	0.979	0.990
	SC_Ino80-AID_Rescue_rep2	28013177	13288181	5407738	275995	0.81		
PRO-cap	SC_Ino80-AID_Ctrl_rep1	29171466	2073847	17697595	621427	0.77	0.995	N/A
	SC_Ino80-AID_Ctrl_rep2	33743267	2528625	20358414	711586	0.88		
	SC_Ino80-AID_KD_rep1	27699506	1537081	18037675	783854	0.96	0.995	
	SC_Ino80-AID_KD_rep2	31144225	2322090	18014706	808134	1.00		
PRO-seq	SC_*wt*_rep1	25155762	8671835	6399398	399563	1.00	0.987	0.992
	SC_*wt*_rep2	26429266	10004941	6435134	351368	0.88		
	SC_*htz1del*_rep1	24700582	10071645	5543700	287164	0.72	0.985	0.992
	SC_*htz1del*_rep2	29874171	11196944	7006309	355166	0.89		
	SC_*arp5del*_rep1	23312472	7394892	6670498	394804	0.99	0.988	0.993
	SC_*arp5del*_rep2	23135952	7778575	6603702	340276	0.85		
	SC_*wt*_rep1	18220821	6492964	4474354	301389	0.55	0.986	0.978
	SC_*wt*_rep2	20853640	5327949	5544028	543728	1.00		
	SC_*spt4del*_rep1	21261858	2758684	11162014	321189	0.59	0.996	0.995
	SC_*spt4del*_rep2	22331041	4475130	11536938	302405	0.56		
	MM_E14Tg2a_*siEgfp*_rep1	41178443	487331	21444972	N/A	N/A	0.979	0.988
	MM_E14Tg2a_*siEgfp*_rep2	60330200	750424	33427705				

	MM_E14Tg2a_*siIno80*_rep1	60798641	732917	33253699			0.980	0.990
	MM_E14Tg2a_*siIno80*_rep2	58179420	840975	34197412				

table S1. Summary of PRO-seq reads and reproducibility obtained in this study.

[a] SC indicates *S. cerevisiae* sample and MM indicates mESCs sample. [b] Norm. factor was calculated by the relative number of reads mapped to a *S. pombe* genome. [c] Reproduciblity of PRO-seq and PRO-cap was calculated by a Spearmans' Rho using spike-in-normalized reads per million at the indicated regions. [d] For *S. cerevisiae*, the regions from the observed TSS to downstream 250bp (TSS to TSS+250 bp) were used as the promoter-proximal regions. For mESCs, the regions from 100bp downstream to 200bp upstream of the observed PRO-seq peaks were used as the promoter-proximal regions.

#	Yeast strain	Genotype	Reference
1	NF191	INO80-IAA*-FLAG, genetic background:DF5, paternal strain: U2721, *his3-Δ200*, *leu2-3,2-112*, *lys2-801*, *trp1-1*(am), *URA3::TIR-9Myc*, *INO80-44AID9Flag::hphNT*	(*29*)
2	SC1242	*spt4*Δ INO80-IAA*-FLAG, genetic background:DF5, paternal strain: U2721, *his3-Δ200*, *leu2-3,2-112*, *lys2-801*, *trp1-1*(am), *URA3::TIR-9Myc*, *INO80-44AID9Flag::hphNT* *spt4Δ::natMX6*	This study
3	SC1163	*paf1*Δ INO80-IAA*-FLAG, genetic background:DF5, paternal strain: U2721, *his3-Δ200*, *leu2-3,2-112*, *lys2-801*, *trp1-1*(am), *URA3::TIR-9Myc*, *INO80-44AID9Flag::hphNT* *paf1Δ::natMX6*	This study
4		BY4741, genetic background S288C, paternal strain: BY4741, MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	
5	SC1086	*arp5*Δ, genetic background S288C, paternal strain: BY4741, MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 arp5Δ::kanMX6*	This study
6	SC1266	*htz1*Δ, genetic background S288C, paternal strain: BY4741, MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 htz1Δ::natMX6*	This study

7	SC1264	*spt4*Δ, genetic background S288C, paternal strain: BY4741, MATa *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 spt4*Δ*::natMX6*	This study

table S2. List of strains used in this study.