

<https://doi.org/10.1038/s42003-024-07014-7>

Spt5 orchestrates cryptic transcript suppression and transcriptional directionality

Check for updates

Haejin An^{1,2}, Hyeokjun Yang^{1,2} & Daeyoung Lee¹

Spt5 is a well-conserved factor that manipulates multiple stages of transcription from promoter-proximal pausing (PPP) to termination. Recent studies have revealed an unexpected increase of antisense transcripts near promoters in cells expressing mutant Spt5. Here, we identify Spt5p-restricted intragenic antisense transcripts and their close relationship with sense transcription in yeast. We confirm that Spt5 CTR phosphorylation is also important to retain Spt5's facility to regulate antisense transcription. The genes whose antisense transcription is strongly suppressed by Spt5p share strong endogenous sense transcription and weak antisense transcription, and this pattern is conserved in humans. Mechanistically, we found that Spt5p depletion increased histone acetylation to initiate intragenic antisense transcription by altering chromatin structure. We additionally identified termination factors that appear to be involved in the ability of Spt5p to restrict antisense transcription. By unveiling a new role of Spt5 in finely balancing the bidirectionality of transcription, we demonstrate that Spt5-mediated suppression of DSIF complex regulated-unstable transcripts (DUTs) is essential to sustain the accurate transcription by RNA polymerase II.

Spt5 is a well-conserved and essential transcription elongation factor that travels with RNA polymerase II (RNAPII) and participates in various steps throughout the transcription cycle^{1,2}. One of the most important Spt5-mediated transcription steps is promoter-proximal pausing (PPP). Spt5 forms the 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole sensitivity-inducing factor (DSIF) complex to stabilize paused RNAPII and release it when CDK9 phosphorylates Spt5². PPP was traditionally believed to exist only in organisms with negative elongation factor (NELF), but recent developments in next-generation sequencing (NGS) technology revealed that a primitive form of pausing is conserved in yeast^{3–5}. These findings enhanced research interest on the meaning of PPP and the role of Spt5 in yeast.

Spt5 regulates RNAPII distribution on the sense strand, and was recently found to restrict antisense transcription near the transcription start site (TSS) in *Schizosaccharomyces pombe* (*S. pombe*)⁶. Research in human cancer cells further showed that inhibiting Spt5 N-terminal phosphorylation evokes antisense transcription from divergent promoters⁷. Despite these interesting findings, however, the role of Spt5 in antisense transcription and the function of the generated transcripts remains unknown. Antisense transcripts in *Saccharomyces cerevisiae* (*S. cerevisiae*) are typically linked to cryptic transcription originating from cryptic promoters⁸. A cryptic transcript can emerge

from both the sense and antisense strands, particularly from lengthy genes with low transcriptional activity^{9–11}. Cryptic transcripts can be categorized by their stability and degradation process as cryptic unstable transcripts (CUTs)^{10,12}, stable unannotated transcripts (SUTs)¹¹, and Xrn1-sensitive unstable transcripts (XUTs)¹³. The new cryptic transcripts were recently shown to be regulated by gene-loop formation or transcriptome surveillance^{14,15}, suggesting that termination factors might also participate in regulating cryptic transcription.

Cryptic transcription is physically suppressed by preventing the transcriptional machinery from accessing the cryptic promoter^{8,16–19}. The major controller of this chromatin accessibility is the Set2p-Rpd3S pathway¹⁹. During transcription elongation, the lysine methyltransferase, Set2p, binds to Ser2-phosphorylated C-terminal domain of RNAPII (Ser2p) and methylates H3K36^{9,20,21}. The histone deacetylase, Rpd3S, is subsequently recruited to deacetylate the transcribed coding region, which disrupts chromatin accessibility and inhibits cryptic transcription^{22–24}. Since histone acetylation levels are co-transcriptionally modulated through histone exchange, this process must be prevented to enable the consistent inhibition of cryptic transcription^{8,19,25}. Recent structural studies suggested that Spt5p can bind and interact with histones, shedding some light on the role of Spt5p in regulating cryptic transcription^{26,27}. A reporter screening suggested that Spt5p and Bur1/2p may suppress the cryptic initiation^{17,28,29}. However, there

¹Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea.

²These authors contributed equally: Haejin An, Hyeokjun Yang. ✉e-mail: daeyoung@kaist.ac.kr

are still missing links between these transcription factors and the process of cryptic transcription.

In the present study, we show that Spt5p and its C-terminal region (CTR) phosphorylation are essential for both suppressing intragenic antisense transcription and enabling successful sense transcriptional elongation in yeast. This suggests that transcription must be precisely modulated in a bidirectional manner to maintain proper gene expression. For genes expressed at higher levels on the sense strand versus the antisense strand, transcription from the latter was found to be suppressed by Spt5p, and this mechanism was conserved in human cells. The increased antisense transcription seen upon Spt5p depletion was found to arise from hyperacetylation of histones and downregulation of H3K36me3 and H3K79me3. Finally, our results suggest that termination factors may contribute to the Spt5p-mediated regulation of cryptic transcription.

Results

Spt5p depletion evokes transcription elongation defects and antisense transcription throughout the genome

Recent studies revealed that Spt5 contributes to protecting RNAPII from degradation^{1,30}. Spt5 must be acutely depleted for the protein levels of RNAPII to be maintained, and this discovery has facilitated efforts to investigate the direct roles of Spt5 in transcription. Thus, we generated auxin-inducible degron (AID^{*})-tagged strains in *S. cerevisiae*. This system enables the acute removal of the target protein using AID^{*}, which contains the minimal size required for its function³¹. We found that Spt5p was completely degraded after 1 h of auxin treatment in cells harboring Spt5p-AID^{*}. Interestingly, longer-term depletion of Spt5p for 4 h was associated with notable decreases in the levels of several essential proteins, suggesting that cellular vitality was inhibited (Supplementary Fig. 1a). Therefore, to exclude secondary effects, we utilized auxin treatment for 1 h to deplete Spt5p in the majority of our experiments.

Spike-in normalized quantifications of sense-strand transcripts from precision run-on sequencing (PRO-seq) libraries generated using Spt5p-depleted and control cells showed that Spt5p regulates transcription at various stages from elongation to termination (Fig. 1a). The PRO-seq signal was largely diminished across all regions of the regulated genes upon 4 h depletion of Spt5p, whereas under the 1 h depletion of Spt5p, the RNAPII signal decreased downstream of the late gene body (GB) but increased near the early GB (Fig. 1a, b). A loss of RNAPII processivity was previously associated with accumulation of RNAPII on the early GB in mammalian SPT5 knockdown (KD) cells^{6,30,32}. Thus, our results suggest that the ability of Spt5p to maintain RNAPII transcription across genes is conserved between *S. cerevisiae* and mammals. The pausing index was increased in most genes, reflecting the decrease in the GB signal (Supplementary Fig. 1c). Unlike the situation in the GB, the signal in the promoter region (PR) was largely unchanged under Spt5p depletion (Supplementary Fig. 1d). Chromatin immunoprecipitation sequencing (ChIP-seq) against RNAPII revealed that the diminished PRO-seq signal resulted from the loss of chromatin-associated RNAPII (Fig. 1b). Consistent with the decrease in RNAPII processivity, the mRNA levels of the altered genes were dramatically decreased upon 1 h of auxin treatment (Fig. 1b, Supplementary Fig. 1d). The group of genes showing significant downregulation in our mRNA-seq results and decreased GB signal in our PRO-seq results exhibited a high degree of overlap (Supplementary Fig. 1e). These results suggest that Spt5p plays an essential role in transcription elongation.

In addition to the effects of Spt5p depletion on sense-strand transcription, we found a surprising increase in antisense transcription under this condition. Although increases in antisense transcripts near promoters were previously reported in Spt5 mutants^{5,7}, we observed a much broader and general increase in intragenic antisense transcription in our system (Fig. 1c, d, Supplementary Fig. 1f). This increased antisense transcription was still observed under long-term (4 h) depletion of Spt5p, despite the dramatic reductions in the RNAPII levels. Overall, the PRO-seq density change was dramatically higher on the antisense strand compared to the sense strand under Spt5p depletion (Fig. 1e). We ruled out the possibility

that auxin caused this transcriptional change by conducting PRO-seq in wild-type cells containing OsTIR1 (Supplementary Fig. 2a, b). These findings suggest that the regulation of antisense transcription by Spt5p is crucial, as increased levels of antisense transcription overcome the effects of RNAPII degradation.

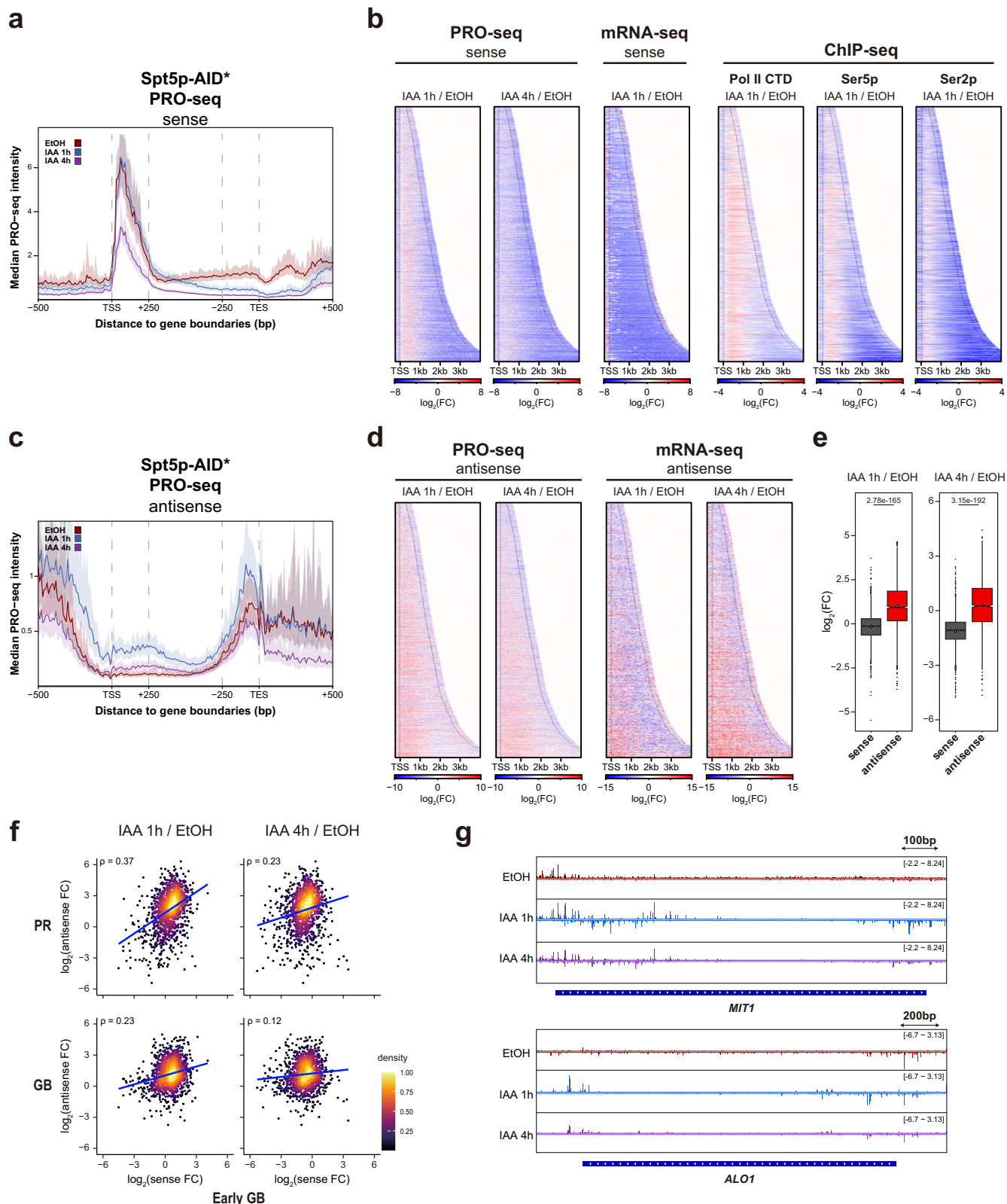
Comparing the PRO-seq intensities between sense and antisense transcription (y-axis of Fig. 1a, c), yeast transcription is dominant in the sense strand as previously reported¹⁸. Since Spt5p modulates transcription in both directions, we wondered if Spt5p's antisense transcription restriction supports this transcriptional direction preference. Thus, we examined the relationship between fold changes in antisense PRO-seq density and transcriptional elongation defects such as accumulation of RNAPII in the early GB and loss of RNAPII near the TES. Our results revealed that the increased antisense transcription upon Spt5p depletion was positively correlated with increased sense PRO-seq intensity in the early GB and negatively correlated with the sense signal upstream of the TES. This correlation was maintained under long-term Spt5p depletion (Fig. 1f, Supplementary Fig. 1g). Our results indicate that elongation defects on the sense strand are highly correlated with increased antisense transcription under Spt5p depletion and further suggest that Spt5p harmoniously balances sense and antisense transcription to ensure proper gene expression. The representative genome browser view gives a clear snapshot of both elongation defects and increased antisense transcription in a single gene (Fig. 1g).

Although Spt4p forms the DSIF complex with Spt5p, the two contrast in that, unlike Spt5p depletion, *spt4Δ* was found to increase the RNAPII signal in the PR^{5,33}. To compare outcomes of Spt5p depletion and *spt4Δ*, we generated a Spt4p-AID^{*} strain and observed the effects of acute Spt4p depletion. In contrast to the effects of Spt5p depletion, prolonged depletion of Spt4p barely affected the protein stability of RNAPII (Supplementary Fig. 2c). Unlike the reported impact of *spt4Δ*, PRO-seq signal was not altered by short-term depletion of Spt4p (Supplementary Fig. 2d). However, long-term depletion of Spt4p increased the PRO-seq intensity in the overall coding region, especially at the pausing peak, as seen for *spt4Δ*. Therefore, we propose that Spt4p indirectly affects PPP through a yet unknown mechanism. This is supported by previous reports that Spt4p undergoes complex interactions with diverse factors including Spt5p^{30,33,34}. We also noted that antisense transcription was increased under both long-term depletion of Spt4p and in *spt4Δ* cells⁵. This suggests that the increased antisense transcription seen with *spt4Δ* is secondary effect of prolonged depletion. Together, the present and previous findings suggest that although Spt4p depletion alone does not substantially affect transcriptional elongation, Spt4p may collaborate with various factors to regulate bidirectional transcription.

DUTs are novel cryptic transcripts restricted by Spt5p

As Spt5p depletion had wide-ranging effects on transcription, we sought to characterize the Spt5p-regulated transcripts. We divided the entire yeast genome into overlapping 36-bp bins, counted the reads that mapped completely to each bin, and excluded bins that overlapped with the sense strand of a coding gene (From -300 bp of the TSS to +300 bp of the TES). We identified significantly up- or down-regulated transcripts using a false discovery rate (FDR) <0.05 and absolute value of log₂FC > 1. The identified transcripts, which we designated 'DSIF complex regulated-unstable transcripts (DUTs)', included 1,164 up-regulated transcripts and 476 down-regulated transcripts (Fig. 2a). When we investigated the regions harboring DUTs, we found that more than half of the up-regulated DUTs were located within intragenic regions or proximal to coding regions, whereas the down-regulated DUTs were typically found in intergenic regions (Fig. 2b). Since the up-regulated DUTs primarily locates within the coding regions, we decided to focus on those within intragenic regions.

To explore whether Spt5p depletion modified the expression of other well-known cryptic transcripts, we compared the overlapped transcripts with the DUTs. Notably, the known cryptic transcripts did not show remarkable overlaps with the DUTs (Fig. 2c). In addition, they remained largely unaltered upon short-term Spt5p depletion and were decreased upon



longer-term depletion; exceptions to this were Set2p-regulated antisense transcripts (SRATs)³⁵ and Suu72p-restricted transcripts (SRTs)¹⁴, which exhibited slight elevations under Spt5p depletion (Fig. 2d). The DUTs and other cryptic transcripts were not altered in auxin-treated control cells (Supplementary Fig. 2e). These findings demonstrate that DUTs are distinctive Spt5p-regulated cryptic transcripts.

To explore whether DUTs are derived from de novo transcription, we generated PRO-cap libraries in Spt5p-AID* cells, targeting cap-containing transcripts produced by active RNAPII. We used the above-described method to define DUTs, except that we excluded only bins that overlapped within +/-36 bp from the TSS on both strands. We also increased the thresholds to FDR < 0.01 and

Fig. 1 | Spt5p is a critical regulator of proper sense and antisense transcription. Median PRO-seq intensity for sense (a) and antisense (c) strands across all filtered genes ($n = 1807$) in Spt5p-AID* cells. IAA (indole-3-acetic acid; auxin) at 1 mM was used to trigger the rapid depletion of Spt5p. Three conditions are illustrated in the plots, based on the time elapsed after the initiation of IAA treatment: control (red), short (blue), and long (violet). **b** Heatmaps depict the \log_2FC of the signal on the sense strand for PRO-seq (left), mRNA-seq (middle), and ChIP-seq (right) ranked by gene length, as obtained upon Spt5p depletion. The TSS and TES of each gene are represented by black dots. **d** Heatmaps depict the \log_2FC of the antisense signals for PRO-seq (left) and mRNA-seq (right). **e** Boxplots represent the \log_2FC for PRO-seq

signal on both sense and antisense strands in intragenic regions; p -values were calculated using the Wilcoxon test. **f** Scatter density plots depict the correlation between \log_2FC of the sense early GB density and \log_2FC of the antisense PR or GB density. Early GB represents the region between 'TSS + 250 bp' and 'TSS + 500 bp' for genes longer than 1 kb in *S. cerevisiae*. The antisense data are distributed on the y axis, while the sense PRO-seq data are distributed on the x axis. The blue line denotes the trend line. Pearson correlation coefficient (ρ) is shown at the top left. **g** Genome browser view of PRO-seq signals for representative genes in Spt5p mutants. The antisense signal is represented reversely to the sense transcription.

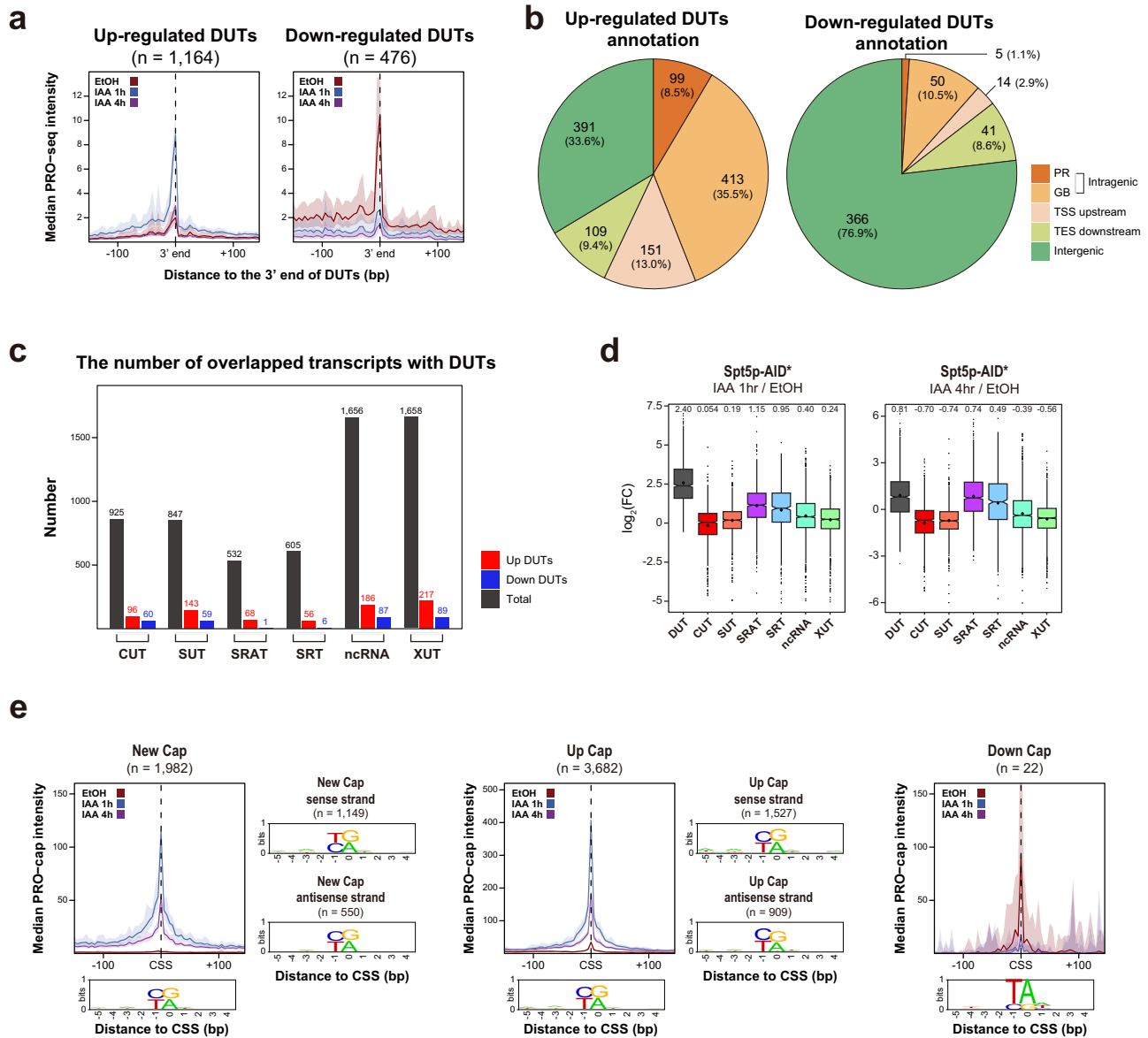


Fig. 2 | DUTs are newly defined cryptic transcripts located mainly within genes. **a** Metagene profiles show the median intensity obtained under Spt5p KD for the PRO-seq signal in the 3' end of up-regulated ($n = 1164$) and down-regulated ($n = 476$) DUTs. **b** Pie graphs illustrate the number of up- and down-regulated DUTs belonging to each genomic region. **c** Bar graphs illustrate the number of transcripts overlapping with DUTs for the following cryptic transcript types; CUT (Cryptic Unstable Transcript), SUT (Stable Unannotated Transcript), SRAT (Set2-Repressed Antisense Transcript), SRT (Ssu72-Restricted Transcript), ncRNA (non-

coding RNA), and XUT (Xrn1-sensitive Unstable Transcript). **d** Boxplots depict the \log_2FC of the PRO-seq obtained under Spt5p depletion for each cryptic transcript. The number at the top of the boxplot presents the median value of \log_2FC . **e** Median plots depict the intensity of the PRO-cap signal from the CSS (Cryptic TSS) in the New Cap ($n = 1982$), Up Cap ($n = 3682$), and Down Cap ($n = 22$) groups. The sequence logos of the CSS in each group are illustrated. The logos for cryptic initiation sites that overlapped with filtered genes are depicted next to the meta-gene plot.

absolute value of $\log_2FC > 2$, to select for more significant changes. We then defined the selected bins as 'New Cap' ($n = 1982$) if they were not counted in control replicate, and as 'Up Cap' ($n = 3682$) or 'Down Cap' ($n = 22$) if both replicates were counted. We plotted the PRO-cap results obtained from cryptic TSSs (CSSs) under both EtOH and auxin treatments. Metagene plots showed that the capped transcripts were properly selected. No New Cap signal was seen in the control treatment groups (Fig. 2e, metaplot). Then, we sought to elucidate the origin of the New Cap. Among them, 1149 originated from the sense strands of genes, 550 originated from antisense strands of genes, and 333 (16.4%) were found outside of genes. A similar pattern was observed for the Up Cap. In addition, the sequence preference for initiation sites was markedly similar between new and up-regulated capping. The same sequence preference was previously reported for the observed TSS⁵. In contrast, the start sites of Down Cap showed a slightly different preference, with T preferred at the -1 position and A at the $+1$ position (Fig. 2e, Sequence Logos)³⁶. This suggests that Spt5p suppresses the propensity of RNAPII to cryptically initiate at a position resembling the TSS.

Phosphorylation of Spt5p CTR regulates bidirectional transcription

Spt5p can play diverse roles according to its post-translational modification state. For example, phosphorylation of the Spt5 CTR domain acts as a signal for the release of paused RNAPII during PPP in metazoans³⁷. In fission yeast, Cdk9 inhibition is discovered to increase intragenic antisense transcription³⁸. To determine if the ability of Spt5p to modulate sense-antisense transcription is related to the phosphorylation state of its CTR, we obtained from S. Buratowski's group a Bur1p^{CDK9}-irreversibly sensitized (IS) mutant for which kinase activity is inhibited upon CMK treatment³⁹. Using this mutant, we found that phosphorylated Spt5p CTR was undetectable after 1 min of CMK treatment, whereas the level of Spt5p was maintained. The level of RNAPII CTD Ser2p was also slightly reduced under Bur1p inhibition, but recovered by 20 min of CMK treatment (Supplementary Fig. 3a). We determined that the change of Ser2p was weak enough to be ignored, compared to the significant reduction of phosphorylated Spt5p.

Since the effect of the CMK treatment was immediate³⁹, we performed PRO-seq of cells treated with CMK for a short duration. As seen for auxin treatment, bidirectional transcription was not altered by CMK treatment of control cells (Supplementary Fig. 2a, b). The signal in the early GB was increased upon Bur1p inhibition, as seen for Spt5p depletion, whereas the late GB signal was maintained (Fig. 3a, b). Moreover, the signal intensity at the GB increased and became widely diffuse as the duration of CMK treatment increased (Fig. 3h up, Supplementary Fig. 3b). This phenotype is distinct to drastic increase of pausing peak in flavopiridol-treated mammalian cells, but recapitulates previous observations made in CDK9-inhibited fission yeast and NELF-depleted *Drosophila*^{4,40}. The previous and present findings support the idea that the role of Bur1p^{CDK9} in transcription elongation is conserved among eukaryotes but become complicated by the presence of NELF⁴⁰.

Antisense transcription was also enhanced upon Bur1p inhibition as upon Spt5p depletion and this effect increased with the duration of CMK treatment. However, the changes observed under Bur1p inhibition were weaker than those seen under Spt5p depletion, and were mainly localized near the TSS (Fig. 3c, d, Supplementary Fig. 3d). The overall PRO-seq density was highly elevated on the antisense strand compared to the sense strand (Fig. 3e). Although we identified fewer differentially expressed genes (DEGs) under Spt5p CTR phosphorylation blockade than under Spt5p depletion, the DEGs largely overlapped between these conditions (Supplementary Fig. 3f). Moreover, among the types of cryptic transcripts, DUTs were increased the most upon Bur1p inhibition, while SRATs and SRTs exhibited slight increases (Fig. 3f). These results confirm that the transcriptional changes seen under Bur1p inhibition are caused by the dephosphorylation of Spt5p. We also observed a positive correlation between increased early GB density for sense transcripts and the fold change (FC) for antisense transcripts. The FC

of antisense transcription showed that there was a higher correlation near the TSS versus the GB (Fig. 3g). This correlation could also be observed in a single gene (Fig. 3h, lower). Together, these results demonstrate that Spt5p CTR phosphorylation appears to regulate the ability of Spt5p to maintain the balance between sense and antisense transcription.

Endogenous expression levels of both sense and antisense transcripts impact the elevated levels of antisense transcription

Next, we looked for possible representative features of genes exhibiting increased antisense transcription under Spt5p depletion. Given the observed correlation between elongation defects and antisense transcription in our system, we hypothesized that endogenous transcription levels could be involved in determining antisense transcription upon Spt5p depletion and dephosphorylation. To test this hypothesis, we divided the coding genes ($n = 1807$) into quartiles based on their endogenous levels of sense (Q1~Q4) and antisense transcription (q1~q4) derived from mRNA-seq RPKM, and then compared the antisense transcription FC of each quartile in both mutants (Fig. 4a, b). Although antisense transcription was more highly increased near the TSS compared to the GB under Bur1p inhibition, we assessed the signal throughout all intragenic regions. Antisense transcription exhibited the greatest increase among genes that were highly expressed genes from the sense strand, and this degree of enhancement gradually decreased with the sense-strand expression level. Conversely, antisense transcription exhibited the lowest increase among genes that were highly expressed from the antisense strand, and this degree of enhancement gradually increased with the endogenous antisense expression level under Spt5p depletion and Bur1p inhibition. Thus, the examined genes showed different tendencies depending on their endogenous bidirectional expression levels. This correlation was sustained throughout the genes; the absolute value of the correlation coefficient was highest near the TSS and gradually decreased to its lowest level downstream of the TES (Supplementary Fig. 4a). These gradients suggest that the endogenous sense and antisense transcription processes undergo remarkable interactions near the TSS, where active RNAPII complexes are the most densely distributed.

To confirm these positional interactions, we classified 1807 genes into four groups based on their endogenous sense and antisense transcription levels, considering only genes in the top or bottom 30% of each: G1 (high sense, high antisense; $n = 145$), G2 (high sense, low antisense; $n = 130$), G3 (low sense, high antisense; $n = 224$), and G4 (low sense, low antisense; $n = 152$) (Fig. 4c). As expected, genes in G2 showed the highest increases in antisense transcription under both Spt5p depletion and Bur1p inhibition (Fig. 4d). The same pattern was observed in our mRNA-seq data with the exception of Spt4p (Supplementary Fig. 4b). No significant change of antisense transcription was seen in wild-type cells treated with auxin or CMK (Supplementary Fig. 4c). The difference in early GB density was more pronounced for G2 genes compared to the other groups (Fig. 4e), suggesting that the G2 genes may experience difficulties in transcription elongation. The average gene length was similar among the group, suggesting that the increased antisense transcription of G2 genes were not artifacts of a shortened gene length (Supplementary Fig. 4d).

Surprisingly, we found that the increases observed in antisense transcription differed slightly between the Spt5p depletion and Bur1p inactivation conditions. The difference between each quartile was more pronounced for the sense RPKM compared to the antisense RPKM for Spt5p-AID* cells (Fig. 4a). Conversely, the disparity was more remarkable for antisense RPKM compared to the sense RPKM for Bur1p-IS cells (Fig. 4b). Similarly, the FC in G4 genes was higher than that in G1 genes upon Spt5p depletion, whereas the FC in G1 genes was higher than that in G4 genes upon Bur1p inhibition (Fig. 4d). These differences may reflect the distinct regulation stages acted upon by Spt5p and Bur1p. We speculate that Spt5p may determine directionality of transcription while Bur1p ensures the accuracy of this process (see 'Discussion').

Next, we wondered whether the antisense transcription suppression by SPT5 could be observed in other organisms containing NELF, such as humans. We re-analyzed the human DLD-1 PRO-seq data obtained by Ali

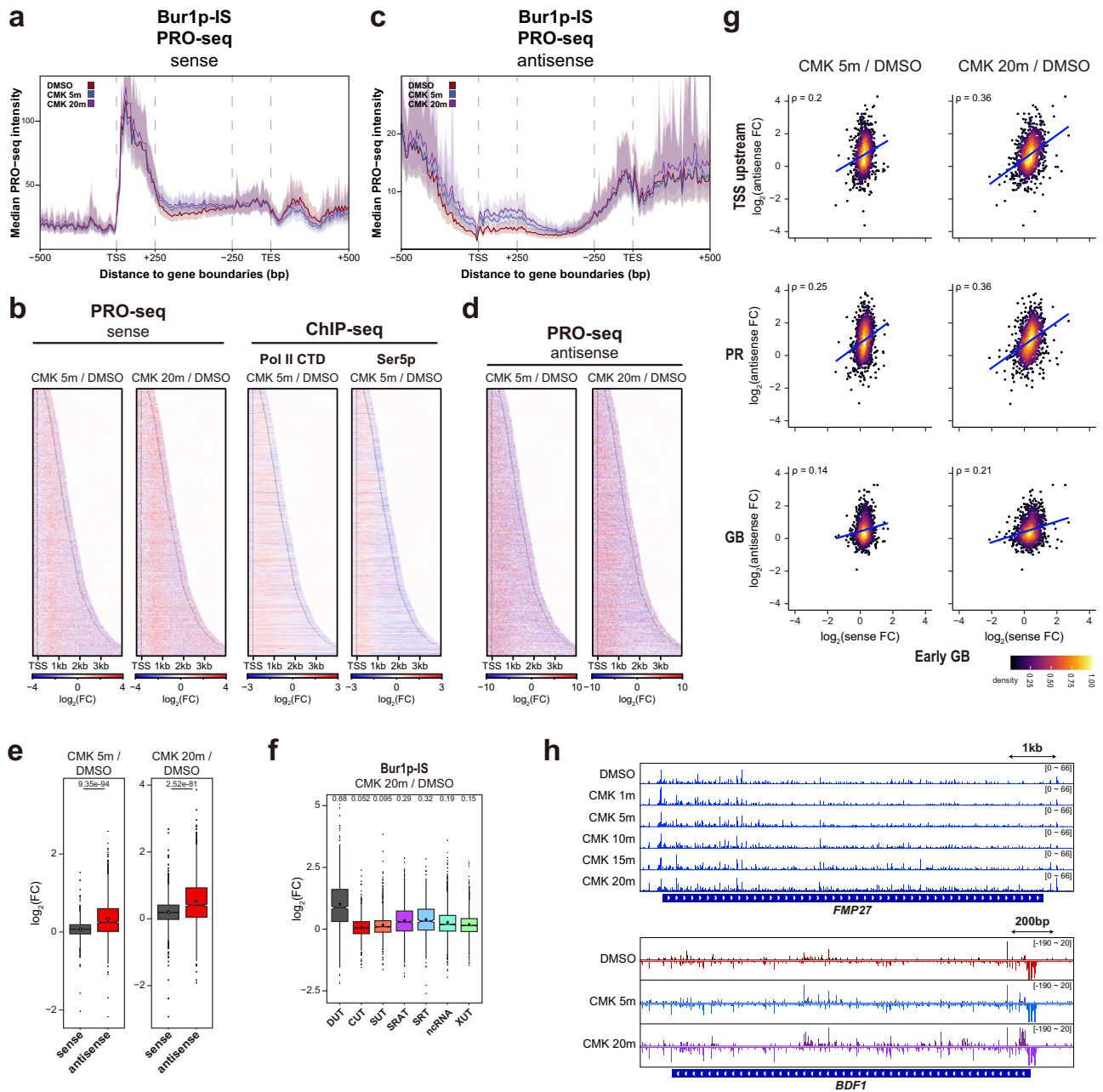


Fig. 3 | Phosphorylation of the Spt5p CTR is crucial for suppressing antisense transcription. Median PRO-seq intensity of sense (a) and antisense (c) strand signals across all filtered genes ($n = 1819$) in Bur1p-IS cells. CMK at 20 μM was used to inhibit Spt5p CTR phosphorylation. Three conditions are illustrated in the plots, based on the time after initiation of CMK treatment: control (DMSO, red), short (5 min, blue), and long (20 min, violet). Heatmaps depicting the $\log_2\text{FC}$ of PRO-seq (b, left; d) and ChIP-seq (b, right) signal intensities ranked by gene length for the sense (b) and antisense (d) strands. The TSS and TES of each gene is represented by black dots. e Boxplots show the $\log_2\text{FC}$ for PRO-seq signal intensity for the sense and antisense strands in intragenic regions. p -values were calculated using the Wilcoxon test. f Boxplots depict the $\log_2\text{FC}$ of PRO-seq signal intensity for each cryptic

transcript under Bur1p inhibition. The number at the top of the boxplot presents the median value of $\log_2\text{FC}$. g Scatter plots compare the $\log_2\text{FC}$ of sense PRO-seq signal intensity in the early GB versus antisense PRO-seq signal intensity in the TSS upstream, PR, and GB. The antisense data are distributed on the y axis, while the sense PRO-seq data are distributed on the x axis. The blue line denotes the trend line. Pearson correlation coefficient (ρ) is positioned at the top left. h Representative example genome browser track showing a sense transcription wave in *FMP27* and elevated antisense PRO-seq signal in *BDF1* upon Bur1p inhibition. The antisense signal is represented reversely to the sense transcription. Transcription changes upon Spt5p depletion in *FMP27* and *BDF1* are represented in Supplementary Fig. 3g.

Shilatfard's group under SPT5 depletion and CDK9 inhibition^{30,41} and found that antisense transcription was increased in SPT5 KD cells treated with inhibitors of VCP and CDK9-inhibited cells by flavopiridol (Supplementary Fig. 4e). Surprisingly, the signal for antisense transcription was also increased when NELF was acutely degraded. Our findings suggest that various factors involved in early elongation contribute to inhibiting antisense transcription.

To examine whether the property described in Fig. 4a–d was also conserved, we generated correlation plots between antisense FC and endogenous sense/antisense transcription levels. As previously observed in yeast, antisense FC was positively correlated with sense transcription and negatively correlated with antisense transcription (Supplementary Fig. 4f). In addition, we classified the genes into the four abovementioned groups (Fig. 4f). Antisense transcription showed the greatest elevation in G2 for

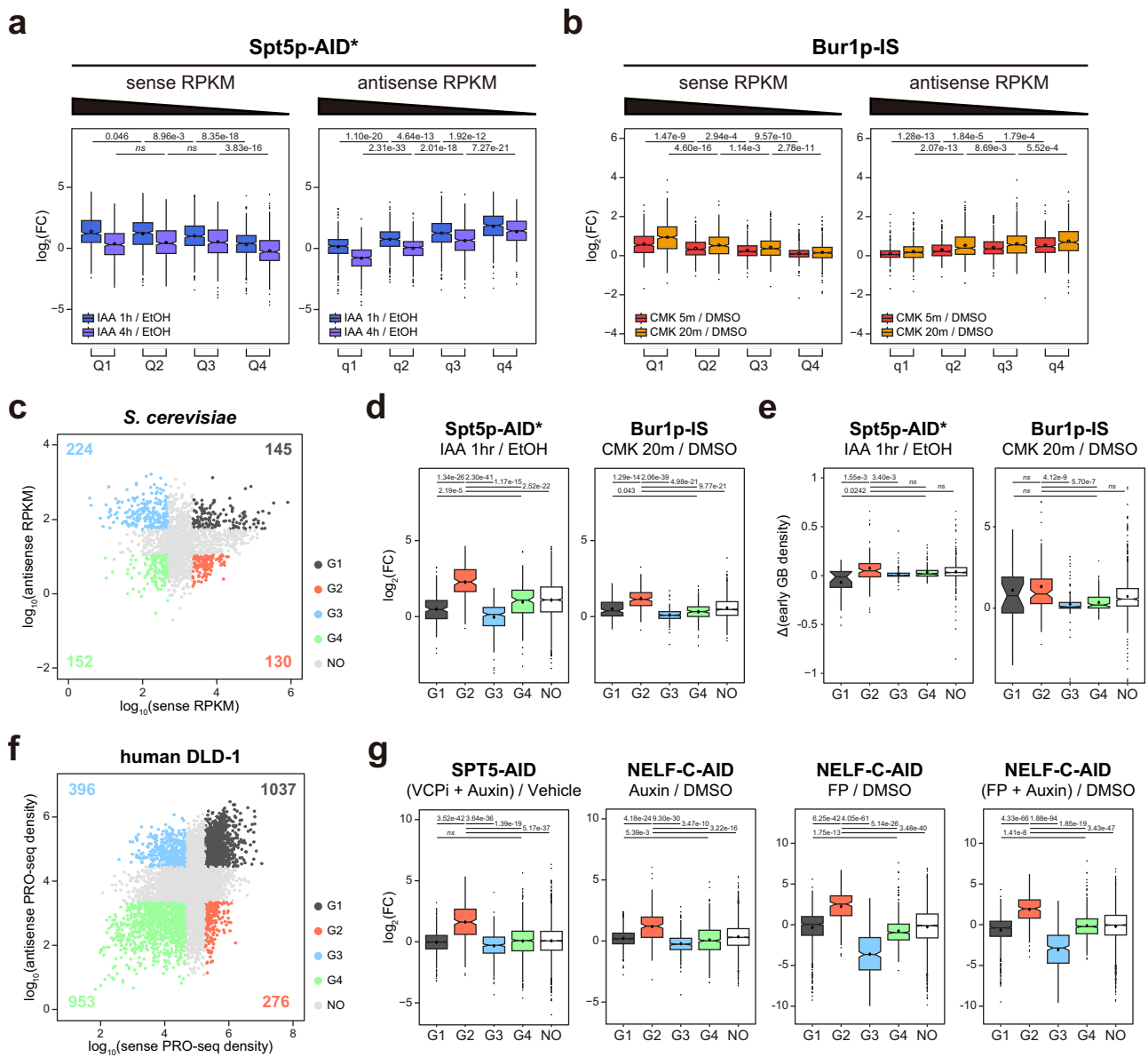


Fig. 4 | G2 genes that show high-level expression from the sense strand and low-level expression from the antisense strand exhibit greater increases of antisense transcription in the Spt5p and Bur1p mutants. Boxplots represent PRO-seq intensity ratios of each quartile for Spt5p-AID* (a) and Bur1p-IS (b) cells. Quantiles are grouped from left to right by decreasing sense RPKM (Q1–Q4) (left) and decreasing antisense RPKM (q1–q4) (right). Genes of *S. cerevisiae* ($n = 1807$) (c) and human DLD-1 ($n = 8622$) (f) are scattered according to their sense (x axis) and antisense (y axis) expression levels. The gene expression levels are presented as RPKM for mRNA-seq (c) and intragenic density for PRO-seq (f). The top or bottom 30% genes of each quadrant are grouped as G1 (black; high sense, high antisense), G2 (orange; high sense, low antisense), G3 (light blue; low sense, high antisense), and G4

(green; low sense, low antisense). The number of genes in each group is presented. The remaining genes are designated ‘No’ (white). **d** Boxplots represent \log_2FC of antisense PRO-seq signal for each group. The signals from Spt5p-AID* cells treated for 1 h with IAA and Bur1p-IS cells treated with CMK for 20 min as compared with their respective control *S. cerevisiae* cells. **e** Boxplots depict the signal density at the early GB for each mutant. **g** Boxplots represent \log_2FC of antisense PRO-seq results for each group. The signals from SPT5-AID cells treated with auxin plus a VCP inhibitor and in NELF-C-AID cells treated with auxin and/or FP are compared with their respective human cells. Human PRO-seq data were downloaded using the appropriate accession number (SPT5-AID, GSE168827 and NELF-C-AID, GSE144786).

every tested mutant (Fig. 4g). These findings indicate that SPT5-mediated antisense transcription suppression is evolutionarily conserved, further suggesting that this suppression system is important for successful transcription.

Spt5p-mediated transcription directionality determination is related to histone acetylation levels

Intragenic cryptic transcription is known to be suppressed by the ability of Set2p-Rpd3S pathway to maintain hypo-acetylated states^{9,22,23}. Multiple transcription factors in fission yeast, Cdk9 and Spt6, have been

revealed to regulate antisense transcription through this pathway^{38,42}. Therefore, we decided to assess possible enrichments of histone modifications in genes exhibiting highly increased antisense transcription in our Spt5p and Bur1p mutant. We compared the signal intensity for genes of each group (G1~G4) using published yeast histone ChIP-seq data (GSE61888)⁴³. We found that H3K36me3 and H3K79me3 were highly enriched in G2 compared to the other groups (Supplementary Fig. 5a). H3K79me3 is a known marker for low histone exchange^{44,45}. Therefore, we implemented ChIP-seq of histone markers which are presumably related to cryptic transcription. Strikingly, we found that the overall levels

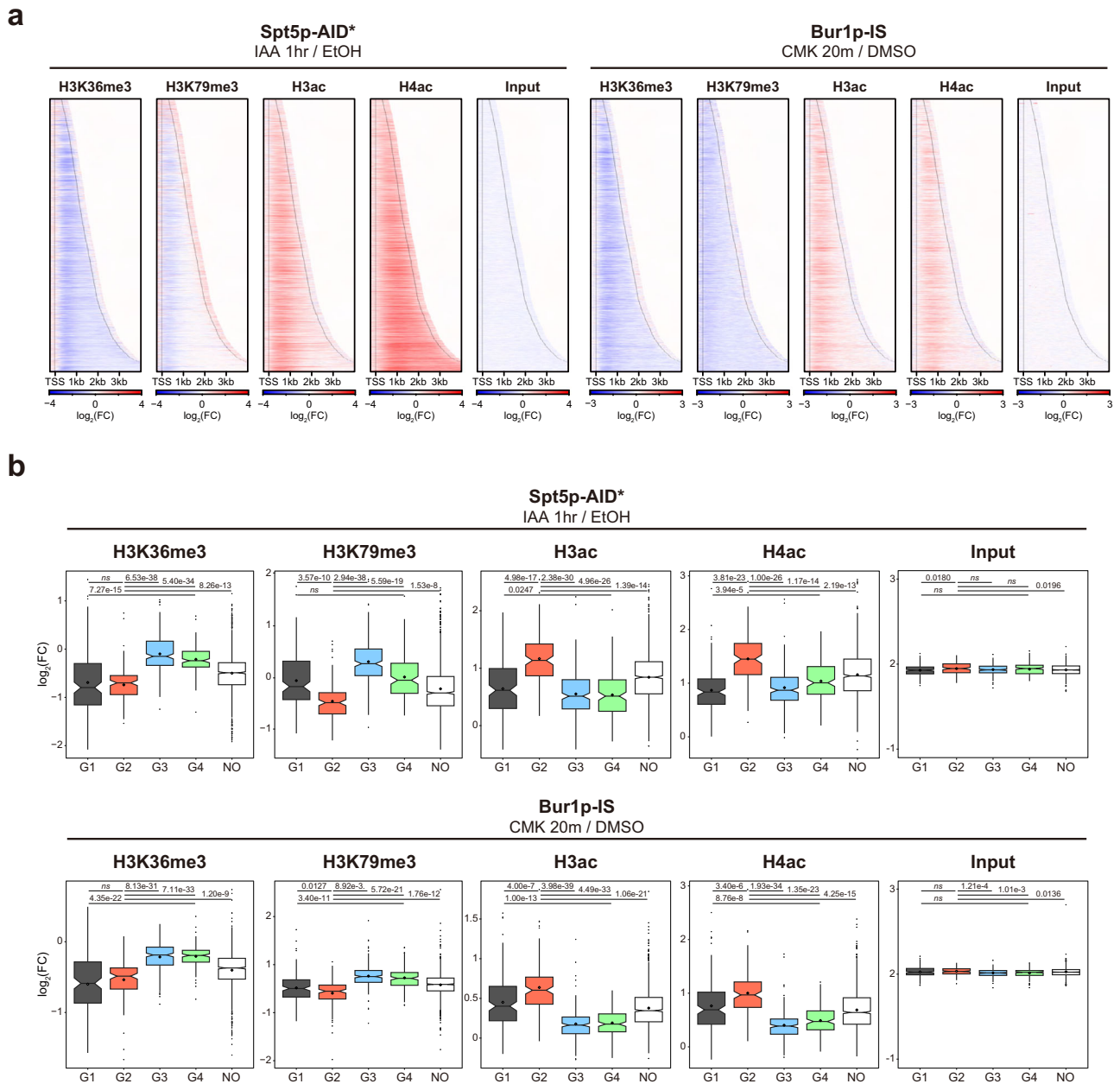


Fig. 5 | Spt5p determines transcription directionality by controlling histone modification. a Heatmaps illustrate the $\log_2\text{FC}$ of each histone ChIP-seq signal (H3K36me3, H3K79me3, pan-acetyl H3, and pan-acetyl H4) normalized by H3 level and ranked by gene length in Spt5p-AID* and Bur1p-IS cells. The TSS and TES of

each gene is represented by black dots. **b** Boxplots depict the $\log_2\text{FC}$ of histone ChIP-seq and Input results for each group described in Fig. 4. Spt5p-AID* cells were treated with or without 1 mM IAA (upper) and Bur1p-IS strain cells were treated with or without 20 μM CMK (lower). *p*-values were calculated using Wilcoxon test.

of H3K36me3 throughout the genes were decreased upon Spt5p depletion or Bur1p inhibition. The levels of H3K79me3 were also notably reduced, especially near the TSS, under Spt5p depletion and Bur1p inactivation, whereas the levels of H3 and H4 acetylation were dramatically increased (Fig. 5a). These results suggest that Spt5p helps maintain proper chromatin context during RNAPII transcription, which may prevent the de novo generation of DUTs.

We next examined whether these histone modification changes were involved in the elongation defect and/or elevated antisense transcription seen in Spt5p-depleted cells. Therefore, we generated correlation plots for the histone ChIP-seq signal throughout the genes, particularly focusing on the signal in the PR for H3K79me3, where the difference in H3K79me3 levels was concentrated in Spt5p-AID* cells. Our analysis showed that genes exhibiting greater increases in antisense transcription under either Spt5p

depletion or Bur1p inhibition showed more pronounced losses of H3K36me3 and H3K79me3 and more obvious gains of acetylation. Likewise, genes showing greater alterations in histone acetylation exhibited increased signal accumulation in the early GB. The alterations in H3K79me3 levels were somewhat correlated with the signal accumulation in the early GB, especially in Spt5p-AID* cells, whereas the changes in H3K36me3 levels showed less correlation (Supplementary Fig. 5b, c). In the same manner, a higher FC of antisense transcription correlated positively with H3 and H4 acetylation upon Spt5p depletion or inhibition of its CTR phosphorylation in G2 genes. The H3K79me3 levels were lowest in G2 genes, and genes of G1 and G2 displayed similar FC levels for H3K36me3 in Spt5p-AID* and Bur1p-IS cells (Fig. 5b). These results suggest that Spt5p depletion might lead to histone hyperacetylation independently of the Set2-Rpd3S pathway.

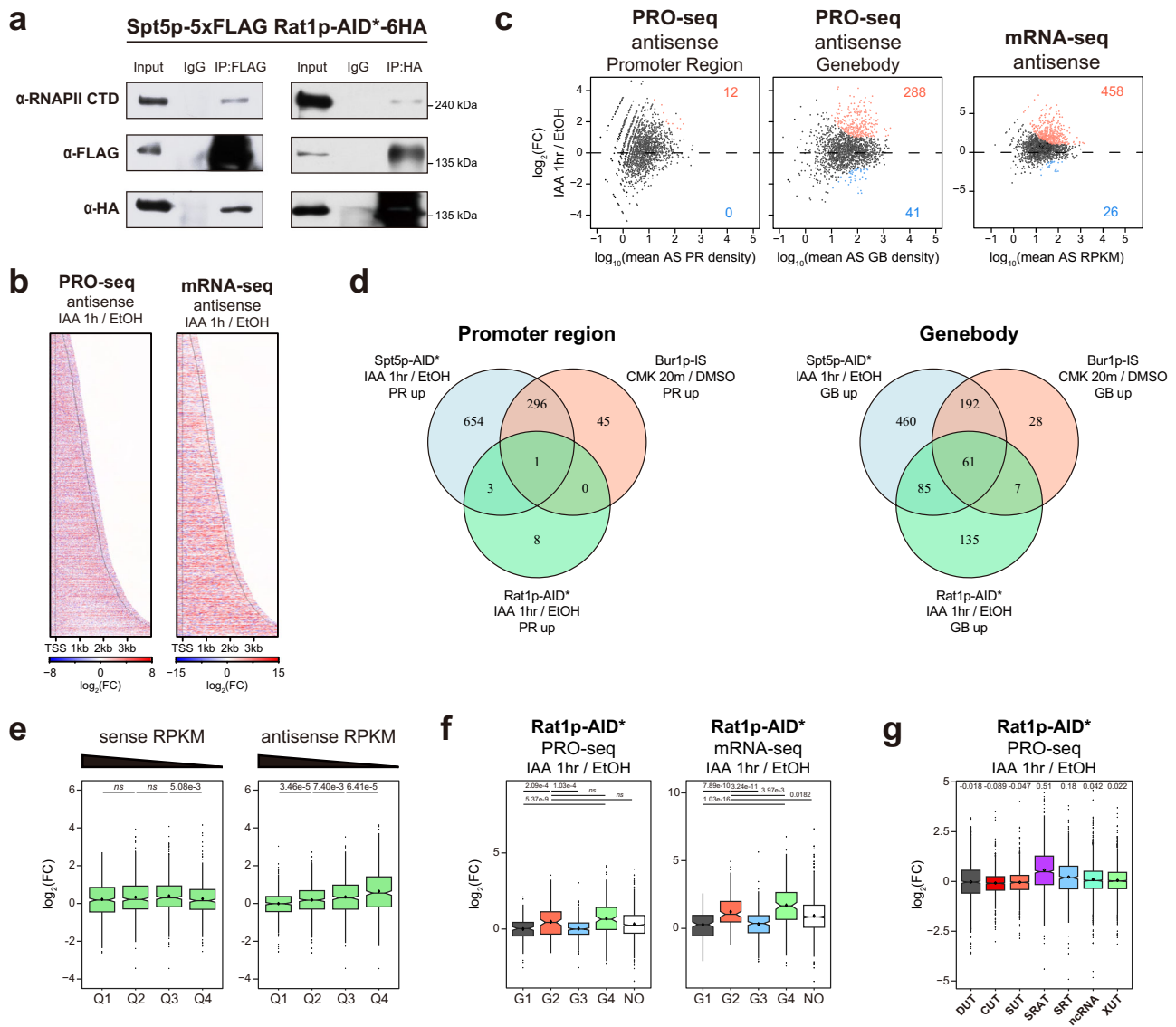


Fig. 6 | Rat1p restricts antisense transcription by interacting with Spt5p. **a** Co-immunoprecipitation of Spt5p-5xFLAG Rat1p-AID*-6HA cell lysate with anti-FLAG and anti-HA. **b** Heatmaps depict \log_2FC results from antisense PRO-seq (left) and mRNA-seq (right) upon Rat1p depletion, ranked by gene length across all filtered genes ($n = 1807$). IAA at 1 mM was applied for 1 h to degrade Rat1p. **c** MA plots illustrating changes observed under treatment with IAA versus EtOH for the antisense PRO-seq signal. Genes are illustrated as dots. Red indicates significantly up-regulated genes with $\log_2FC > 1$ and adjusted p -value < 0.05 ; blue indicates significantly down-regulated genes with $\log_2FC < -1$ and adjusted p -value < 0.05 . The numbers of DEGs are depicted in the plot. **d** Venn-diagram indicating the number of

overlapped genes between DEGs showing significantly increased antisense PRO-seq signals in the PR and GB. Image sizes do not reflect gene numbers. **e** Boxplots representing the PRO-seq intensity ratios in each quartile for Rat1p-AID* cells. Quantiles are grouped by decreasing sense RPKM from left to right (Q1–Q4) (left) and decreasing antisense RPKM from left to right (q1–q4) (right). p -values were calculated using the Wilcox test. **f** Boxplots depict the \log_2FC for PRO-seq signals (left) and mRNA-seq signals (right) upon Rat1p depletion in each group described in Fig. 4. **g** Boxplots depict the \log_2FC of PRO-seq signals upon Rat1p KD for each cryptic transcript. The number at the top of the boxplot is the median value of \log_2FC .

Rat1p^{Xrn2} contributes to suppressing antisense transcription via an interaction with Spt5p

As mentioned above, SRTs were slightly increased under both Spt5p depletion and Bur1p inhibition (Figs. 2b and 3g). SRTs are antisense transcripts expressed from divergent promoters that are endogenously suppressed by gene loops formed by Ssu72p¹⁴. Since Ssu72p is a transcription termination factor⁴⁶ and Spt5p can participate in transcription termination⁴⁷, we speculated that termination factors may regulate antisense transcription. Therefore, we assessed candidate termination factors for possible involvement in regulating antisense transcription. First, we tested Rrp6p, which serves as a nuclear subunit of the 3'-5' exosome and contributes to 'reverse torpedo' termination⁴⁸. Since *rrp6Δ* is frequently reported to stimulate CUTs^{10,12}, we generated degron strains, confirmed the

degradation of Rrp6p (Supplementary Fig. 6a), and then assessed the effect of Rrp6p depletion in our system. However, Rrp6p depletion for 1 h was insufficient to alter the levels of antisense transcription and known cryptic transcription (Supplementary Fig. 6b, c). This suggests that CUTs are generated by secondary phenomena in *rrp6Δ* cells.

Next, we focused on Xrn2 (Rat1p in *S. cerevisiae*), which acts as a 5'-3' riboexonuclease in degrading uncapped nascent RNAs and causing 'torpedo' transcription termination⁴⁹. Recent mass spectrometry data revealed that Xrn2 is among the highest-level binders of Spt5 in fission yeast⁵⁰. Our co-immunoprecipitation results showed that Rat1p physically binds Spt5p in *S. cerevisiae* (Fig. 6a). We additionally generated strains in which Rat1p proteins were rapidly degraded (Supplementary Fig. 6a). Surprisingly, we found that antisense transcription levels were increased upon Rat1p

depletion. This pattern was more notable in the GB compared to PR (Fig. 6b, c) and in mRNA-seq rather than PRO-seq. These results suggested that Rat1p plays an additional role at the GB, beyond its contribution to torpedo termination after the TES.

To clarify the interaction between Rat1p and Spt5p, we compared the DEGs showing up-regulated antisense transcripts in each PRO-seq library. In the PR, where less antisense transcription was observed upon Rat1p depletion, the DEGs of Rat1p-AID* showed little overlap with the other DEG sets of Spt5p-AID* and Bur1p-IS. Conversely, the DEGs identified under Rat1p KD in the GB highly overlapped with the DEGs identified under Spt5p depletion (~50.7%), and the antisense transcripts of these 61 genes were heightened in all three experimental settings. (Fig. 6d). Interestingly, the increases showed different patterns: The antisense transcripts increased throughout the genes upon Spt5p depletion, increased near the TSS upon Bur1p inhibition, and increased near the TES upon Rat1p depletion. On the other hand, no antisense transcription was observed in Rrp6p-AID* or control cells (Supplementary Fig. 6d, f).

We next questioned whether the properties of increased antisense transcripts in the Rat1p depletion were the same as seen under the Spt5p depletion and Bur1p inhibition. Similar to the patterns seen in Spt5p-AID* cells, antisense transcription increased as the endogenous antisense transcription levels decreased. However, the sense RPKM was not correlated with the increases in antisense transcription (Fig. 6e, Supplementary Fig. 6e). Genes of G4 showed the greatest increases in antisense transcription upon Rat1p depletion, which contrasted with our findings under Spt5p depletion and Bur1p inhibition (Fig. 6f). In addition, DUTs were not notably altered in the Rat1p mutant, while SRATs and SRTs were slightly increased (Fig. 6g). These results suggest that the Rat1p-mediated regulation of antisense transcription could be partially, but not exclusively, related to the Spt5p-mediated pathway.

Discussion

In this study, we unexpectedly discovered a new role for Spt5p and its CTR phosphorylation in generally restricting intragenic antisense transcription. More specifically, we show that Spt5p contributes to preventing RNAPII from initiating at the cryptic site. Spt5p suppresses the antisense transcription of genes that show high-level expression on the sense strand and low-level expression on the antisense strand. Antisense transcription from the divergent promoters was observed in control experiments (Fig. 1c). Although the antisense transcription within genes was weaker than the endogenous antisense transcription upstream of divergent promoters, the large fold changes were generally seen upon Spt5p depletion. The presence of antisense transcripts on a gene is expected to impede the normal transcriptional function of RNAPII. Here, we observed an increase in newly initiated transcripts within the GB in Spt5p-depleted cells (Fig. 2), indicating that antisense transcription and/or new initiation will prohibit RNAPII-mediated elongation. However, we cannot exclude the possibility that the elongation defects in the Spt5p mutant itself result in the mislocalization of RNAPII.

If the absolute quantity of increased antisense transcription were the same for all genes, it would be reasonable for genes with low-level transcription on the antisense strand to show high fold changes. However, we found that genes with high-level transcription on the sense strand also showed the high fold-change increases in antisense transcription upon Spt5p depletion. This indicates that the more forceful progression of RNAPII on the sense strand can suppress antisense transcription. The loss of efficient elongation in Spt5p mutants overlooks the rationale for blocking antisense transcription. Since these genes are highly transcribed by RNAPII, mislocalized RNAPII may bind more often to these genes compared to other genes. The unphosphorylation of Spt5p was more pronounced in G1 than in G4 due to the high-level enrichments of RNAPII and Spt5p (Fig. 4). Thus, Spt5p phosphorylation might serve as quality control stage for guiding RNAPII progression in the correct direction.

We found that histone acetylation was significantly increased upon Spt5p depletion and Bur1p inhibition. The enrichments in H3K36me3 and

H3K79me3 were higher in genes of G2 compared to the other groups. Therefore, these histone modifications have a high potential to suppress antisense transcription, as previously reported⁵¹. H3K36me3 was also known to regulate antisense transcription by association with transcription factors in fission yeast^{38,42}. However, the increases in antisense transcription were correlated with increases in histone acetylation rather than decreases in H3K36me3 and H3K79me3, and thus differed slightly from the observations previously reported for *GALI:ADH1_T*⁵². This correlation indicates that histone acetylation might be the direct cause of increased cryptic transcription by exposing the cryptic promoters to be transcribed^{8,16–19}. That is, cryptic transcription is effectively restricted by the Set2p-Rpd3S pathway at steady state^{18,19}. However, Spt5p depletion and CTR dephosphorylation can acetylate histones through a Set2p-independent pathway to increase intragenic antisense transcription (Fig. 7). In that Spt5p can be involved in chromatin preservation both directly and through interaction with the chromatin remodeler^{27,53}, we speculate that hyperacetylation upon Spt5p depletion is caused by instability of nucleosome assembly. Interestingly, Spt5 and its CTR phosphorylation can negatively regulate heterochromatin formation in fission yeast⁵⁴. Our observation may explain how antisense transcripts regulated by Spt5 can control proper heterochromatin formation since heterochromatin factors are known to suppress antisense transcription^{55,56}.

Rat1p depletion increased antisense transcripts mostly in GB and those transcripts were only correlated with endogenous antisense RPKM rather than with sense RPKM (Fig. 6). We have discovered that Rat1p's antisense transcription regulation is related to Spt5p. However, the detailed mechanism remains yet to be disclosed. A recent structural study revealed that Rat1p can compete with Spt5p to bind with the same binding site on RNAPII⁵⁷. This implies that Rat1p depletion can postpone the Spt5p dislodging from RNAPII or conversely, Spt5p depletion can bring forward Rat1p-RNAPII binding. This newly reported interaction between Spt5p and Rat1p can be a hint to understand the role of Rat1p in antisense transcription regulation.

We discovered that Spt5p-mediated cryptic antisense transcription is conserved from yeast to humans (Fig. 4), suggesting that this process is important for ensuring a proper transcriptional balance. It is plausible that more diverse factors are involved in the transcriptional processes of higher eukaryotes compared to yeast. We observed that NELF depletion also induced antisense transcription, indicating that transcription factors contribute to suppressing antisense transcription. In summary, Spt5-related chromatin remodelers, elongation factors, and RNAPII itself may contribute to fine-tuning transcription on the opposite strand of coding genes.

Methods

Yeast strains and culture

The yeast strains used in this paper are listed in Supplementary Table 1. Yeast cells were grown at 30 °C in YPD medium until they reached mid-log phase (0.5 OD/ml). For degradation of AID*-tagged target proteins, cells were incubated in YPD medium containing auxin at a final concentration of 1 mM (Sigma, I2886). The AID* strains were treated with auxin for 1 h for rapid depletion, unless otherwise stated. The kinase-mutated IS strain lost its kinase function when exposed to CMK at a final concentration of 20 μM (MedChem Express, HY-52101). The IS mutant strains were a gift from Prof. Stephen Buratowski's lab.

Western blot analysis

Yeast cells were lysed with a standard bead-beating protocol; and proteins were eluted by boiling the sample for 5 min in 2× SDS sample buffer (20% glycerol, 0.4% bromophenol blue, 100 mM Tris-Cl, pH 6.8, 4% SDS, and 200 mM β-mercaptoethanol). The utilized antibodies and their concentrations are listed in Supplementary Table 2.

Co-immunoprecipitation

Mid-log phase cells (40 OD) were harvested and resuspended in NP-40 lysis buffer (0.1% NP-40, 200 mM NaCl, 2 mM EDTA, 6 mM Na₂HPO₄, 4 mM

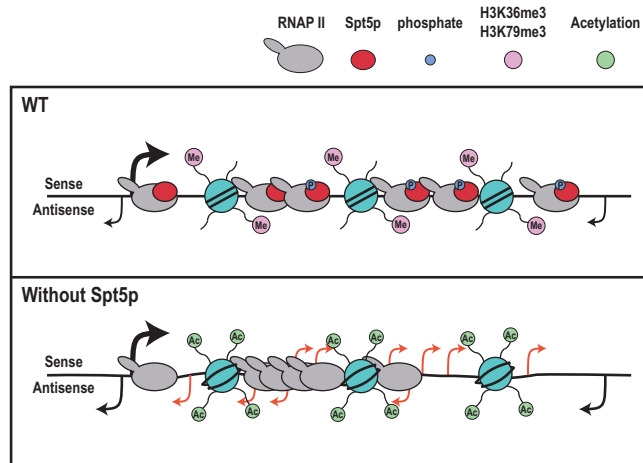


Fig. 7 | Proposed function of Spt5p in suppressing antisense transcription. Spt5p enables proper transcriptional elongation by RNAPII and prevents antisense transcription within genes. H3K36me3 and H3K79me3 are highly enriched on genes exhibiting high-level transcription from the sense strand and low-level transcription from the antisense strand. In the absence of Spt5p, RNAPII encounters difficulties in progression and experiences premature termination. This causes rapid histone turnover and consequent removal of H3K36/79me3 and incorporation of acetylated histones, yielding a highly loosened DNA structure. The black arrows indicate the RNAPII signal observed in PRO-seq, with the signal intensity represented by the arrow size. The increased cryptic transcription within genes is indicated by orange arrows. RNAPII having the potential to complete mature elongation is depicted as a gray circle. Other colors are as follows: Spt5p (red), phosphate (light blue), H3K36me3 / H3K79me3 (light pink), and histone acetylation (green).

NaH₂PO₄, and protease inhibitors) and lysed via a standard bead-beating protocol, and 0.1% of the whole-cell lysate was used as the input. Immunoprecipitation was carried out at 4 °C for 2 h using Dynabead protein A (10002D, ThermoFischer) or protein G (10004D, ThermoFischer) along with 25 U of benzonase (E1014, Millipore) and 1.5 mM MgCl₂. The beads were washed three times and boiled in 2X SDS sample buffer to elute the proteins, which were resolved and quantified using Western blot analysis.

Yeast cell preparation for PRO-seq

Mid-log phase cells (5 OD) were harvested by centrifugation, cell pellets were washed with ice-cold DEPC-H₂O, the cells were permeabilized with 10 ml of 0.5% sarkosyl (Sigma, L5777) for 20 min at 4 °C. The permeabilized cells were gently centrifuged at 400 × g for 5 min at 4 °C and then stored in 200 μl of storage buffer (10 mM Tris-Cl, pH 8.0, 25% glycerol, 5 mM MgCl₂, 0.1 mM EDTA, and 5 mM DTT). The samples were snap-frozen and stored at −80 °C. For the yeast spike-in control, 1.25 OD of *S. pombe* (ED665) cells underwent permeabilization and were stored in 500 μl of storage buffer. Prior to the subsequent PRO-seq steps, 50 μl of spike-in stock (0.125 OD) was added to 5 OD of permeabilized *S. cerevisiae* sample. To exclude the possibility that artifacts were caused by the chemical treatment, we generated and assessed PRO-seq libraries with OsTIR1-containing parental strains for auxin (U2721; represented as OsTIR1 in this paper) and wild-type cells (w303a; represented as wild-type or WT in this paper) for CMK.

PRO-seq and PRO-cap

Nuclear run-on reactions and RNA extractions were conducted as previously described⁵⁸, with the minor adjustments reported in previous study⁴. The combined yeast cells were gently centrifuged at 400 × g for 5 min at 4 °C, and the supernatant was completely removed. The obtained cell pellets were resuspended in 300 μl of nuclear run-on reactions containing 25 μM biotin-11-UTP (PerkinElmer, NEL543001EA), 25 μM biotin-11-CTP (PerkinElmer, NEL542001EA), 125 μM ATP (Roche, 11140965001), and 125 μM GTP (Roche, 11140957001) in run-on reaction buffer (20 mM Tris-HCl,

pH 7.7, 200 mM KCl, 5 mM MgCl₂, 2 mM DTT, and 0.4 U/μl RNase inhibitor) containing 0.5% sarkosyl. The run-on reaction was performed at 30 °C for 5 min and then gently centrifuged at 400 × g for 5 min at 4 °C to ensure complete removal of the supernatant. RNA was extracted from the cell pellets using hot acidic phenol. The respective libraries were generated using the published PRO-seq or PRO-cap protocols⁵⁸. Minor variations in reagents were as previously reported³. To reduce the bias from PCR duplicates, we introduced the 10-nucleotide UMI (Unique Molecular Identifier) sequence to the 3' adaptor (VRA3-UMI, 5'-/5Phos/NNNNNNNNNNGAUCGUCGGACUGUAGAACUCUGAAC/Inverted dT/−3') for PRO-seq experiments and 5' adaptor (RA5-UMI, 5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNNNNNN-3') for PRO-cap experiments. DNA libraries of approximately 100 bp to 350 bp were purified by agarose gel extraction (Zymo Research, D4007) and sequenced on an Illumina HiSeq X.

mRNA-seq

Mid-log yeast cells (10 OD) were harvested and washed once with cold water. For the spike-in control, 0.5 OD of *S. pombe* (ED665) cells were mixed with 10 OD of *S. cerevisiae* cells. Total RNA was extracted from cell pellets using a standard hot-phenol RNA extraction method. DNA was eliminated from the total RNA using a TURBO DNA-free kit (Invitrogen, AM1907) at 37 °C for 1 h. mRNA was isolated from the total RNA using an NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB, E7490) according to the provided protocol. Libraries were prepared using an NEBNext Ultra™ II Directional RNA Library Prep Kit for Illumina (NEB, E7760). The resulting libraries were sequenced using the Illumina HiSeqX platform, employing the paired-end method with 150-bp reads.

Chromatin immunoprecipitation and library generation

ChIP was conducted as reported by Strahl-Bolsinger et al.⁵⁹, with several modifications. In brief, 4 OD of *S. pombe* (ED665) cells were mixed with 80 OD of *S. cerevisiae* cells for spike-in normalization. Whole-cell lysates were subjected to 12 cycles of sonication for 20 s using a Branson sonifier SFX550 at 35% amplitude. Adequate time intervals were maintained between each sonication cycle. The immunoprecipitated and input DNA were eluted in elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) at 65 °C for 1 h. Subsequently, RNase A (ThermoFisher, EN0531) and proteinase K (Promega, V3021) were added to the eluted DNA for RNA degradation and de-crosslinking. The extracted DNA was purified using a QIAquick PCR purification kit (Qiagen, 28106). Libraries for RNAPII ChIP-seq were generated using the NEXTflex™ Illumina ChIP-Seq Library Prep kit (BIOO, 5143-02), following the manufacturer's protocols. Sequencing was performed on a HiSeqX using the paired-end method with 150 bp reads. For histone ChIP-seq libraries, an NEBNext Ultra™ II DNA Library Prep Kit for Illumina (NEB, E7645) was utilized, as described by the manufacturer. Sequencing was performed on a NovaSeq 6000 using the paired-end method with 150-bp reads.

Data analysis for PRO-seq and PRO-cap

BigWig files were generated as previously described^{3,4} with minor modification to remove PCR duplicates using UMI-tools⁶⁰. Briefly, 10-bp UMI sequences were removed and kept in the FASTQ name using the function 'umi_tools extract' in UMI-tools during adaptor trimming. Duplicates with the same UMI sequences were removed from the BAM files using the 'umi_tools dedup' function. However, we found that the deduplicated replicates of PRO-cap were not suitable for further analysis due to low reproducibility. Therefore, we simply removed UMI sequences of FASTQ and applied spike-in normalization for PRO-cap data processing. To analyze PRO-seq data in *SPT4* KO, we used published data (GSE76142)⁵. For human PRO-seq, mapped genome of BigWig files between NELF-C-AID (hg19, GSE144786) and SPT5-AID (hg38, GSE168827) is different. Therefore, we downloaded the FASTQ files from GEO. We aligned the human DLD-1 PRO-seq data to hg38 for the experiments and to mm10 for the spike-in normalization.

To define the observed TSS with our PRO-cap in *S. cerevisiae*, we followed the method described in our previous paper³ with the use of an additional filter to investigate antisense transcription. We used the reads obtained from ethanol-treated PRO-seq and PRO-cap experiments performed in Spt5p-AID* cells. We further filtered out genes that overlapped with other genes or were located within 300 bp of another gene on the both strands to exclude read throughs from neighboring genes. In total, 1807 *Saccharomyces Genome Database* (SGD) genes were analyzed as follows. For genes of human DLD-1, we downloaded protein-coding genes with RefSeq annotation from the UCSC Genome Browser. For those with multiple isoforms, we selected the BED region from the isoform with the highest expression level. After we filtered out genes shorter than 1 kb, those located within 1 kb of another gene, and those with PRO-seq read counts lower than 10, we retained 8622 genes for further analysis in DLD-1 cells.

To define DUTs (i.e., the cryptic transcripts suppressed by Spt5p), we divided the total yeast genome (sacCer3) into 36-bp bins at 1-bp intervals and counted the exactly matching reads. To consider strandedness, we categorized the bins as plus strand and minus strand. We excluded bins that overlapped with the sense strand of the analyzed genes ($n = 1807$, from TSS -300 bp to TES $+300$ bp). Bins that overlapped with regions near to filtered-out genes ($n = 3664$, from TSS -300 bp to TES $+300$ bp) were excluded, regardless of strandedness. Next, we counted the mapped reads that exactly matched each remaining bin. Bins with read counts of 0 in all samples and with a non-zero read count for only one sample were excluded. We then used the remaining bins and calculated the adjusted p -value and \log_2FC using DESeq2⁶¹. Up-regulated transcripts were defined as those with the FDR < 0.05 and $\log_2FC < -1$ and down-regulated transcripts are defined as the FDR < 0.05 and $\log_2FC > 1$. Then, they are combined if they are within a distant of 36 bp. This strategy yielded 1165 up-regulated DUTs and 476 down-regulated DUTs.

For the New Cap, Up Cap, and Down Cap analyses, we followed the above-described DUT-defining methodology up to the step at which divided bins by their strands. We excluded bins near filtered-out genes and those near the observed TSSs (from TSS -36 bp to TSS $+36$ bp) of the analyzed genes. We quantified the mapped reads that exactly matched each remaining bin. Bins with read counts of 0 in all samples except one were excluded. The threshold of significantly changed transcripts was increased to FDR < 0.01 and absolute value of $\log_2FC > 2$. From among the bins, we identified those with no read count obtained with EtOH as New Cap ($n = 2032$). The remaining bins were categorized as Up Cap ($n = 3857$) for those with $\log_2FC > 2$ and Down Cap ($n = 24$) for those with $\log_2FC < -2$. For this analysis, we excluded bins with no read count for all replicates and with read counts for only one replicate. To obtain the BED files for published cryptic transcripts, we simply downloaded them from the supplemental data in the reference paper^{10,13,14,35}. We conducted downstream analysis using publicly available R scripts obtained from GitHub, with minor modification⁴. In this research, the gene region was defined as follows: TSS upstream (TSS -250 bp to TSS), PR (TSS to TSS $+250$ bp), GB (TSS $+250$ bp to TES), TES upstream (TES -250 bp to TES), and TES downstream (TES to TES $+250$ bp). The early GB region was taken as the region from TSS $+250$ bp to TSS $+500$ bp for genes longer than 1 kb (in yeast) or 2 kb (in humans).

Data analysis for mRNA-seq

The raw reads were mapped to the combined yeast genome (sacCer3 of *S. cerevisiae* and SpombeASMv2 of *S. pombe*) using STAR⁶² (version 2.7.6a) with clipping of 50 bp from the 3' end. The aligned BAM files were sorted and indexed using SAMtools⁶³ (version 1.14). BEDTools⁶⁴ (version 2.30.0) was used to convert the indexed BAM files to BED formats. The genomecov function of BEDTools was used to generate the coverage of the aligned reads. The BedGraph files were normalized using the spike-in reads aligned to *S. pombe*. The normalized BedGraph files were converted to BigWig files using bedGraphToBigWig⁶⁵. Unlike the PRO-seq strategy, the read counts in mRNA-seq experiments were calculated for all intragenic regions, without being separated into the PR and GB, using the multicov function of

BEDTools. Significant DEGs were identified using DESeq2 with the following thresholds: adjusted p -value < 0.05 and absolute $\log_2FC > 1$ ⁶¹.

Data analysis for ChIP-seq

The raw reads were mapped individually to sacCer3 for experiments and to SpombeASMv2 for spike-in normalization using bowtie2⁶⁶ (version 2.3.4.3). MACS2⁶⁷ was used to convert the aligned BAM files to bedGraph files with the pileup option. The bedGraph formats aligned to sacCer3 were normalized with the spike-in mapped reads. BigWig files were generated using bedGraphToBigWig⁶⁵. We have used the published yeast histone ChIP-seq data (GSE61888) to observe the strong histone marks in genes of G2⁴³. For histone ChIP-seq data upon Spt5p depletion of Bur1p inhibition, the BigWig files were also normalized with H3 values.

Statistics and reproducibility

Every sequencing data except Spt5p-AID* PRO-seq are the average of two biological replicates. Three biological replicates are used for Spt5p-AID* PRO-seq. The amount of yeast cells used for each library are shown in methods of each experiment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The PRO-seq, mRNA-seq, and ChIP-seq data obtained in this research have been deposited to GEO with accession numbers GSE267734 (PRO-seq), GSE267735 (RNA-seq) and GSE267736 (ChIP-seq). The processed BigWig files have also been deposited to GEO. The accession numbers for the genome-wide data downloaded for our analyses are mentioned in the Figure Legends or Methods. Uncropped images of every blot in the manuscript are shown in Supplementary Figs. 8–12. The source data including bed files can be found in Supplementary Data 1–4. Any other data generated in this study are available from the corresponding author on reasonable request.

Received: 11 June 2024; Accepted: 3 October 2024;

Published online: 22 October 2024

References

- Hu, S. et al. SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. *Mol. Cell* **81**, 4425–4439. e4426 (2021).
- Song, A. & Chen, F. X. The pleiotropic roles of SPT5 in transcription. *Transcription* **13**, 53–69 (2022).
- Cheon, Y., Han, S., Kim, T., Hwang, D. & Lee, D. The chromatin remodeler Ino80 mediates RNAPII pausing site determination. *Genome Biol.* **22**, 294 (2021).
- Booth, G. T., Parua, P. K., Sanso, M., Fisher, R. P. & Lis, J. T. Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. *Nat. Commun.* **9**, 543 (2018).
- Booth, G. T., Wang, I. X., Cheung, V. G. & Lis, J. T. Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Res.* **26**, 799–811 (2016).
- Shetty, A. et al. Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol. Cell* **66**, 77–88 e75 (2017).
- Park, K. et al. ZWC complex-mediated SPT5 phosphorylation suppresses divergent antisense RNA transcription at active gene promoters. *Nucleic Acids Res.* **50**, 3835–3851 (2022).
- McDaniel, S. L. & Strahl, B. D. Shaping the cellular landscape with Set2/SETD2 methylation. *Cell Mol. Life Sci.* **74**, 3317–3334 (2017).
- Li, B. et al. Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. *Genes Dev.* **21**, 1422–1430 (2007).
- Xu, Z. et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).

11. Neil, H. et al. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**, 1038–1042 (2009).
12. Wyers, F. et al. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725–737 (2005).
13. van Dijk, E. L. et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**, 114–117 (2011).
14. Tan-Wong, S. M. et al. Gene loops enhance transcriptional directionality. *Science* **338**, 671–675 (2012).
15. Schulz, D. et al. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* **155**, 1075–1087 (2013).
16. Silva, A. C. et al. The replication-independent histone H3-H4 chaperones HIR, ASF1, and RTT106 co-operate to maintain promoter fidelity. *J. Biol. Chem.* **287**, 1709–1718 (2012).
17. Cheung, V. et al. Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLoS Biol.* **6**, e277 (2008).
18. Smolle, M. & Workman, J. L. Transcription-associated histone modifications and cryptic transcription. *Biochim. Biophys. Acta* **1829**, 84–97 (2013).
19. Smolle, M., Workman, J. L. & Venkatesh, S. reSETting chromatin during transcription elongation. *Epigenetics* **8**, 10–15 (2013).
20. Schaft, D. et al. The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. *Nucleic Acids Res.* **31**, 2475–2482 (2003).
21. Krogan, N. J. et al. Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell Biol.* **23**, 4207–4218 (2003).
22. Hainer, S. J. & Martens, J. A. Identification of histone mutants that are defective for transcription-coupled nucleosome occupancy. *Mol. Cell Biol.* **31**, 3557–3568 (2011).
23. Du, H. N. & Briggs, S. D. A nucleosome surface formed by histone H4, H2A, and H3 residues is needed for proper histone H3 Lys36 methylation, histone acetylation, and repression of cryptic transcription. *J. Biol. Chem.* **285**, 11704–11713 (2010).
24. Du, H. N., Fingerman, I. M. & Briggs, S. D. Histone H3 K36 methylation is mediated by a trans-histone methylation pathway involving an interaction between Set2 and histone H4. *Genes Dev.* **22**, 2786–2798 (2008).
25. Smolle, M. et al. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat. Struct. Mol. Biol.* **19**, 884–892 (2012).
26. Farnung, L., Ochmann, M., Engholm, M. & Cramer, P. Structural basis of nucleosome transcription mediated by Chd1 and FACT. *Nat. Struct. Mol. Biol.* **28**, 382–387 (2021).
27. Evrin, C. et al. Spt5 histone binding activity preserves chromatin during transcription by RNA polymerase II. *EMBO J.* **41**, e109783 (2022).
28. Chu, Y., Simic, R., Warner, M. H., Arndt, K. M. & Prelich, G. Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *EMBO J.* **26**, 4646–4656 (2007).
29. Zhou, K., Kuo, W. H., Fillingham, J. & Greenblatt, J. F. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc. Natl Acad. Sci. USA* **106**, 6956–6961 (2009).
30. Aoi, Y. et al. SPT5 stabilization of promoter-proximal RNA polymerase II. *Mol. Cell* **81**, 4413–4424.e4415 (2021).
31. Morawska, M. & Ulrich, H. D. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* **30**, 341–351 (2013).
32. Fong, N., Sheridan, R. M., Ramachandran, S. & Bentley, D. L. The pausing zone and control of RNA polymerase II elongation by Spt5: Implications for the pause-release model. *Mol. Cell* **82**, 3632–3645.e3634 (2022).
33. Uzun, U., Brown, T., Fischl, H., Angel, A. & Mellor, J. Spt4 facilitates the movement of RNA polymerase II through the +2 nucleosomal barrier. *Cell Rep.* **36**, 109755 (2021).
34. Ding, B., LeJeune, D. & Li, S. The C-terminal repeat domain of Spt5 plays an important role in suppression of Rad26-independent transcription coupled repair. *J. Biol. Chem.* **285**, 5317–5326 (2010).
35. Venkatesh, S., Li, H., Gogol, M. M. & Workman, J. L. Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nat. Commun.* **7**, 13610 (2016).
36. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
37. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).
38. Sanso, M. et al. Cdk9 and H2Bub1 signal to Clr6-Cll/Rpd3S to suppress aberrant antisense transcription. *Nucleic Acids Res.* **48**, 7154–7168 (2020).
39. Chun, Y. et al. Selective Kinase Inhibition Shows That Bur1 (Cdk9) Phosphorylates the Rpb1 Linker In Vivo. *Mol. Cell Biol.* **39**, e00602-18 (2019).
40. DeBerardine, M., Booth, G. T., Versluis, P. P. & Lis, J. T. The NELF pausing checkpoint mediates the functional divergence of Cdk9. *Nat. Commun.* **14**, 2762 (2023).
41. Aoi, Y. et al. NELF Regulates a Promoter-Proximal Step Distinct from RNA Pol II Pause-Release. *Mol. Cell* **78**, 261–274.e265 (2020).
42. DeGennaro, C. M. et al. Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. *Mol. Cell Biol.* **33**, 4779–4792 (2013).
43. Weiner, A. et al. High-resolution chromatin dynamics during a yeast stress response. *Mol. Cell* **58**, 371–386 (2015).
44. Gat-Viks, I. & Vingron, M. Evidence for gene-specific rather than transcription rate-dependent histone H3 exchange in yeast coding regions. *PLoS Comput. Biol.* **5**, e1000282 (2009).
45. Lee, S. et al. Dot1 regulates nucleosome dynamics by its inherent histone chaperone activity in yeast. *Nat. Commun.* **9**, 240 (2018).
46. Tan-Wong, S. M., Wijayatilake, H. D. & Proudfoot, N. J. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.* **23**, 2610–2624 (2009).
47. Baejen, C. et al. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol. Cell* **66**, 38–49.e36 (2017).
48. Lemay, J. F. & Bachand, F. Fail-safe transcription termination: Because one is never enough. *RNA Biol.* **12**, 927–932 (2015).
49. Kim, M. et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517–522 (2004).
50. Kuś, K. et al. DSIF factor Spt5 coordinates transcription, maturation and exoribonucleolysis of RNA polymerase II transcripts. Preprint at <https://www.biorxiv.org/content/10.1101/2023.10.16.562519v1> (2023).
51. Brown, T. et al. Antisense transcription-dependent chromatin signature modulates sense transcript dynamics. *Mol. Syst. Biol.* **14**, e8007 (2018).
52. Murray, S. C. et al. Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Res.* **43**, 7823–7837 (2015).
53. Simic, R. et al. Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO J.* **22**, 1846–1856 (2003).
54. MacKinnon, S. et al. Spt5 C-terminal repeat domain phosphorylation and length negatively regulate heterochromatin through distinct mechanisms. *PLoS Genet* **19**, e1010492 (2023).
55. Zofall, M. et al. Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature* **461**, 419–422 (2009).
56. Gullerova, M. & Proudfoot, N. J. Cohesin complex promotes transcriptional termination between convergent genes in *S. pombe*. *Cell* **132**, 983–995 (2008).
57. Zeng, Y., Zhang, H. W., Wu, X. X. & Zhang, Y. Structural basis of exoribonuclease-mediated mRNA transcription termination. *Nature* **628**, 887–893 (2024).

58. Mahat, D. B. et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11**, 1455–1476 (2016).
59. Strahl-Bolsinger, S., Hecht, A., Luo, K. & Grunstein, M. SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev.* **11**, 83–93 (1997).
60. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
62. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
63. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
64. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
65. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
67. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

Acknowledgements

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (No. 022M3A9B6017424) and by NRF grants funded by the Ministry of Science and ICT (MSIT) (2018R1A5A1024261, SRC; 2022R1A2C3003115). We thank Prof. Stephen Buratowski for providing the IS strain. We also thank Youngseo Cheon in our lab for critical reading of this paper.

Author contributions

D.L. designed this study. H.A. completed all experiments in this research. H.Y. implemented all computational data analysis. All authors participated in writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-024-07014-7>.

Correspondence and requests for materials should be addressed to Daeyoung Lee.

Peer review information *Communications Biology* thanks Jerry Workman and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editors: Elah Pick and David Favero. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024