

Transcription elongation is finely tuned by dozens of regulatory factors

Mary Couvillion[†], Kevin M Harlen[†], Kate C Lachance[†], Kristine L Trotta, Erin Smith, Christian Brion, Brendan M Smalec, L Stirling Churchman*

Blavatnik Institute, Department of Genetics, Harvard Medical School, Boston, United States

Abstract Understanding the complex network that regulates transcription elongation requires the quantitative analysis of RNA polymerase II (Pol II) activity in a wide variety of regulatory environments. We performed native elongating transcript sequencing (NET-seq) in 41 strains of *Saccharomyces cerevisiae* lacking known elongation regulators, including RNA processing factors, transcription elongation factors, chromatin modifiers, and remodelers. We found that the opposing effects of these factors balance transcription elongation and antisense transcription. Different sets of factors tightly regulate Pol II progression across gene bodies so that Pol II density peaks at key points of RNA processing. These regulators control where Pol II pauses with each obscuring large numbers of potential pause sites that are primarily determined by DNA sequence and shape. Antisense transcription varies highly across the regulatory landscapes analyzed, but antisense transcription in itself does not affect sense transcription at the same locus. Our findings collectively show that a diverse array of factors regulate transcription elongation by precisely balancing Pol II activity.

Editor's evaluation

In this manuscript the authors have conducted native elongation transcript sequencing on yeast strains deleted for one of 41 different transcription, chromatin modifying and RNA processing factors. They find that a large fraction of these deletions affect transcription elongation and RNA Pol II pausing indicating that elongation is carefully regulated by many factors.

*For correspondence: churchman@genetics.med.harvard.edu

[†]These authors contributed equally to this work

Competing interest: See page 20

Funding: See page 20

Preprinted: 15 August 2021

Received: 25 March 2022

Accepted: 15 May 2022

Published: 16 May 2022

Reviewing Editor: Jerry L Workman, Stowers Institute for Medical Research, United States

© Copyright Couvillion, Harlen, Lachance et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Transcription is a highly regulated and conserved process that consists of three phases: initiation, elongation, and termination (*Shandilya and Roberts, 2012; Svejstrup, 2004*). Post-initiation regulation is critical for co-transcriptional RNA processing, shaping the chromatin landscape, and preventing run-on transcription into downstream genes (*Herzel et al., 2017; Holmes et al., 2015; Proudfoot et al., 2002; Rando and Winston, 2012*). Transcription elongation is controlled across gene bodies by a wide variety of factors, including transcription factors, chromatin modifiers, chromatin assembly factors and chaperones, RNA processing factors, and histone variants. Understanding how these factors act separately and in concert to influence RNA polymerase II (Pol II) activity will shed light on how transcription elongation and co-transcriptional processes are coordinated.

Transcription is a discontinuous process: periods of productive elongation are frequently interrupted by pauses. Pol II pausing was first observed in vitro in *Escherichia coli* polymerase transcribing the *lac* operon and lambda DNA (*Dahlberg and Blattner, 1973; Gilbert et al., 1974; Kassavetis and Chamberlin, 1981; Kingston and Chamberlin, 1981; Lee et al., 1976; Maizels, 1973*). Observations of Pol II pausing in vivo provided the first evidence of promoter proximal pausing (*Gariglio et al., 1981*). These findings were extended by chromatin immunoprecipitation (ChIP) studies, which

identified paused polymerase near the 5' ends of certain *Drosophila* and mammalian genes (**Bentley and Groudine, 1986; Eick and Bornkamm, 1986; Gilmour and Lis, 1986; Krumm et al., 1992; Nepveu and Marcu, 1986; Rougvie and Lis, 1988; Spencer and Groudine, 1990; Strobl and Eick, 1992**).

The advent of high-throughput and high-resolution sequencing technologies has led to the development of sequencing methods such as NET-seq and precision run-on sequencing (PRO-seq) that measure Pol II density genome-wide at nucleotide resolution. Collectively, these techniques have highlighted the control of transcription elongation by regulatory factors. These approaches and other nascent RNA sequencing methods visualize the production of transcripts from RNA polymerases across the genome (**Churchman and Weissman, 2011; Core et al., 2008; Kwak et al., 2013; Mayer et al., 2015; Nojima et al., 2015; Schwalb et al., 2016**), and therefore are capable of revealing the immediate and direct effects of a perturbation on transcription. In addition, these assays capture unstable transcripts such as antisense RNAs, which can be critical to transcription regulation but are invisible by many techniques (**Camblong et al., 2007; Hongay et al., 2006; Martens et al., 2004; Uhler et al., 2007**). The strand-specificity and high resolution of these methods are transforming our understanding of transcription elongation and regulation.

NET-seq, PRO-seq, and other high-resolution methods have revealed both regions of high Pol II density, such as promoter proximal pausing, and specific sites of Pol II pausing across gene bodies (**Churchman and Weissman, 2011; Ferrari et al., 2013; Kindgren et al., 2020; Kwak et al., 2013; Larson et al., 2014; Mayer et al., 2015; Nojima et al., 2015; Vvedenskaya et al., 2014; Weber et al., 2014**). Regions or peaks of high Pol II density, such as promoter proximal pauses, are created in part by a high density of pause sites that together create barriers to elongation and provide an opportunity for regulation and coordination of co-transcriptional events (**Bentley, 2014; Mayer et al., 2017; Noe Gonzalez et al., 2021; Rougvie and Lis, 1988**). Myriad factors control Pol II peaks in vivo. For example, in yeast prominent peaks of Pol II density occur near polyadenylation [poly(A)] sites (**Harlen et al., 2016**). Loss of Rtt103, a termination factor, causes a dramatic peak in Pol II density directly downstream of poly(A) sites (**Harlen et al., 2016**). On the other hand, specific sites of Pol II pauses are reminiscent of pausing observed at single nucleotide positions in vitro (**Herbert et al., 2008; Kassavetis and Chamberlin, 1981; Kingston and Chamberlin, 1981; Mayer et al., 2017**). These in vitro pauses arise from intrinsic properties of the polymerase itself, interactions with the DNA template, and the presence of bound proteins (e.g. histones and transcription factors) (**Herbert et al., 2006; Hodges et al., 2009; Kassavetis and Chamberlin, 1981; Kireeva et al., 2005; Kireeva and Kashlev, 2009; Shaevitz et al., 2003**). NET-seq analysis of Pol II pause sites in yeast and mammalian cells has revealed a similar connection to DNA sequence and histones, but has not been explored across different regulatory landscapes (**Churchman and Weissman, 2011; Gajos et al., 2021**).

Pol II transcribes much of the genome in all eukaryotes, yet only a fraction of its transcripts mature into stable, protein-coding RNA products (**Bertone et al., 2004; Cheng et al., 2005; David et al., 2006; Hangauer et al., 2013; Kapranov et al., 2007; Mercer et al., 2011; Nagalakshmi et al., 2008; Smolle and Workman, 2013; Steinmetz et al., 2006**). A major contributor to unstable noncoding RNA products is antisense transcripts, i.e., RNAs transcribed from the strand opposite the sense strand of a protein-coding gene. Originally identified in bacteria (**Spiegelman et al., 1972**), antisense transcripts were soon discovered in eukaryotes as well (**Anderson et al., 1981; Bibb et al., 1981**). Since its discovery, antisense transcription has been detected opposite the vast majority of annotated genes in yeast (**Xu et al., 2011**), arising initially as a natural consequence of open chromatin regions (**Jin et al., 2017**). Antisense transcription regulates gene expression at a number of yeast genes (**Camblong et al., 2007; Hongay et al., 2006; Houseley et al., 2008; Lenstra et al., 2015; Martens et al., 2004; Uhler et al., 2007**); however, a general genome-wide function has not been identified (**Murray and Mellor, 2016**). To better understand pervasive antisense transcription and its role in regulation, it is important to determine whether it is tunable by regulatory factors, which would help distinguish whether the levels of antisense transcription are tightly set or whether antisense transcription is simply a nuisance that the cell works to minimize.

To gain insight into the regulation of the production of coding and non-coding transcripts by Pol II, we used NET-seq to analyze 41 *Saccharomyces cerevisiae* mutant strains lacking known elongation regulators. We investigated how each factor regulates nascent transcription, production of antisense transcripts, and pausing across gene bodies. Surprisingly, across these regulatory contexts, we find

that antisense transcription at a locus does not affect its sense transcription. Metrics describing each transcription phenotype span a broad dynamic range with wild-type activity lying near the center. The loss of each factor revealed distinct sets of pause sites that we used to create machine learning models of Pol II pausing, highlighting which genomic features can classify pause positions. Together, our results show that Pol II transcription elongation is determined by the contrasting impacts of many regulatory factors.

Results

Reverse genetic screen for transcription regulators

To obtain insight into the transcription regulatory network of *S. cerevisiae*, we individually deleted 41 non-essential transcription elongation regulators, including RNA processing factors, transcription elongation factors, histone variants, chromatin modifiers, and chromatin remodelers and chaperones, and assessed the transcriptional effects of each deletion using NET-seq (**Figure 1A**). The wild-type transcription baseline was established using four biological replicates of wild-type cells; the results from the replicates were highly correlated ($R^2 \geq 0.97$; **Figure 1—figure supplement 1A**). All mutant strains were analyzed in at least biological duplicate. Results from strain replicates were highly correlated ($R^2 \geq 0.75$; **Supplementary file 1**). Importantly, all replicates were performed at different times, by different researchers, and in different strain isolates, demonstrating the reproducibility of our results.

Nascent gene expression is uniquely disrupted across deletion strains

Because all of the factors examined in our screen play roles in transcription regulation, we first sought to determine whether each factor regulates different sets of genes, or whether modifications of the transcriptional regulation network affect the transcription of overlapping sets of genes. Based on NET-seq data, we assessed the role of each factor in regulating nascent transcription, a more direct measurement of transcriptional phenotype than can be obtained from RNA-seq data. Nascent transcripts are produced antisense to the coding strand at substantial levels (**Churchman and Weissman, 2011**), so to obtain a complete and accurate view of expression differences, we annotated the antisense version of all genes and included these in differential expression analysis with DESeq2 (**Love et al., 2014**). First, we focused on sense protein-coding genes and inspected how many were differentially expressed across the strains. Interestingly, in some strains (e.g. *rph1Δ* and *nap1Δ*), very few genes were transcribed at significantly altered levels relative to the wild-type, whereas upon loss of Rpb4, a subunit of RNA polymerase, over 10% of all protein-coding genes were differentially transcribed (**Figure 1B, Figure 1—figure supplement 1B; Supplementary file 2**).

We then investigated the degree to which differentially transcribed genes were shared across mutant strains. Over 90% of differentially transcribed genes were identified in fewer than nine deletion strains, and 41% were differentially transcribed in only a single strain (**Figure 1C**). Only a few genes had altered expression in most of the deletion strains; some of these, such as *HSP12* are involved in stress responses, and their regulation may represent the cellular reaction to losing key transcription regulators.

We next asked whether certain biological functions or pathways were commonly affected across the deletion strains using GO enrichment analysis (**Figure 1D, Figure 1—figure supplement 1C; Supplementary file 3 Anders and Huber, 2010; Ashburner et al., 2000; Mi et al., 2019; The Gene Ontology Consortium, 2019**). Over 90% of GO pathways enriched among the differentially transcribed genes were identified in fewer than five deletion strains, with 56% identified in a single strain, emphasizing the largely distinct responses to loss of each factor (**Figure 1—figure supplement 1D**). GO enrichments were not particularly strong or specific overall (**Supplementary file 3**); however, we did detect enrichment of some pathways consistent with the known functions of certain factors. Upon deletion of *HPC2*, which encodes a subunit of the HIR nucleosome assembly complex involved in the regulation of histone gene transcription (**Formosa et al., 2002; Prochasson et al., 2005; Xu et al., 1992**), the term 'DNA replication-independent chromatin organization' (GO:0034724) was significantly enriched (100-fold enrichment, p -adjusted=0.021) among downregulated genes (**Figure 1D**). The GO term 'chromatin assembly factor (CAF-1) complex' (GO: 0033186) was enriched among downregulated genes only upon deletion of *CAC1*, *CAC2*, and *CAC3*, which encode CAF-1 subunits. Some

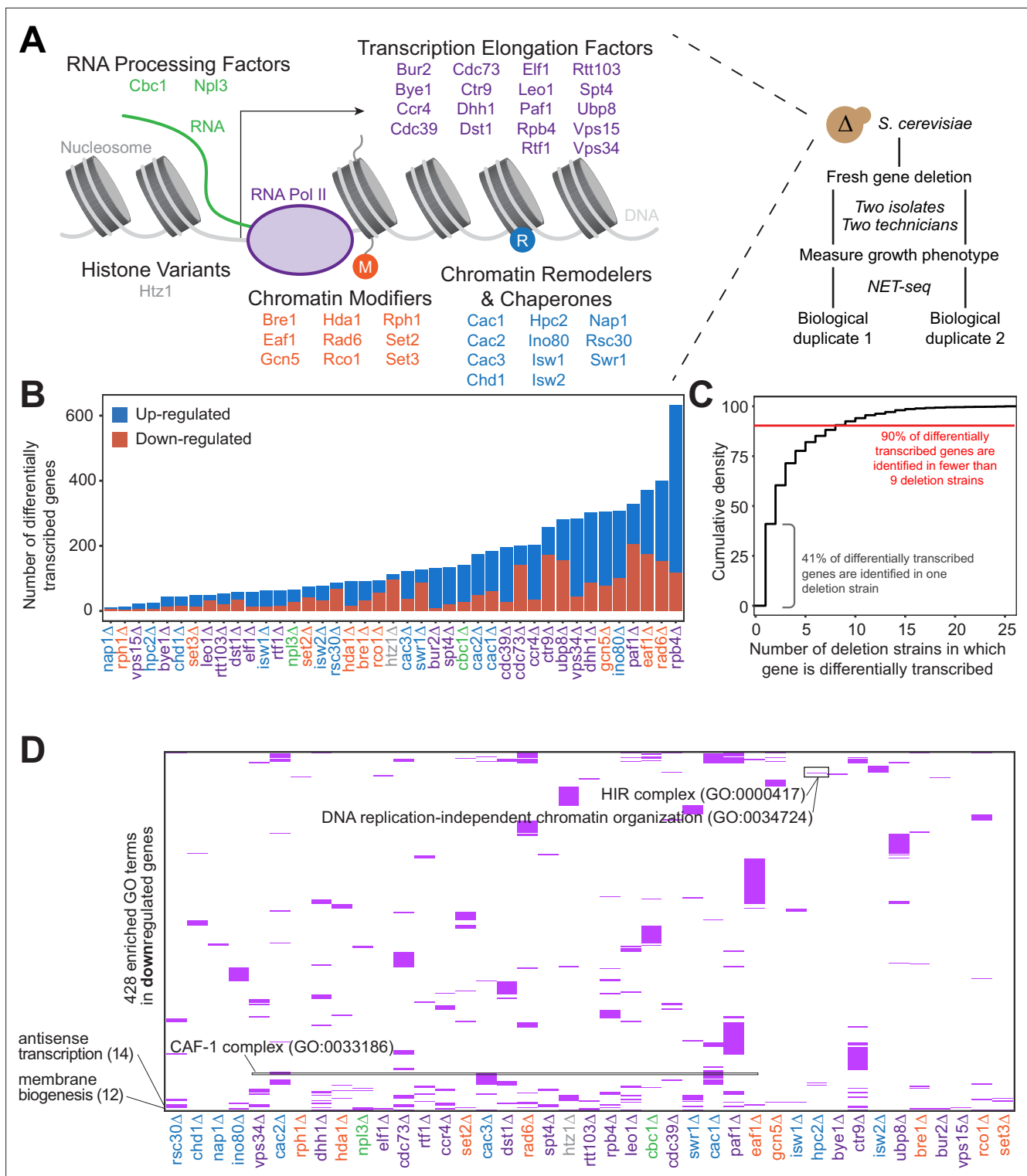


Figure 1. Gene expression is affected differently when transcription regulatory proteins are knocked out, both at the level of individual genes and gene ontology. **(A)** As polymerase II transcribes along a chromatinized template, a complex network regulates eukaryotic transcription elongation. Factors analyzed in the reverse genetic screen are listed and grouped by function: RNA processing factors (green), transcription elongation factors (purple), histone variants (gray), chromatin modifiers (orange), and chromatin remodelers and chaperones (blue). Colors of factors consistent throughout figures. Each of these factors were deleted to conduct a reverse genetic screen in *Saccharomyces cerevisiae*. For each deletion strain, a fresh gene deletion was conducted in two isolates by two technicians. After a growth phenotype was measured, native elongating transcript sequencing (NET-seq) was performed in at least two biological replicates. **(B)** A number of differentially up- (blue) and downregulated (red) genes vary widely across deletion

Figure 1 continued on next page

Figure 1 continued

strains. For differential expression analysis, all reads mapping to protein coding regions and their antisense counterparts were considered. Here, only sense genes are included in the counts. (C) Cumulative density plot illustrating that 41% of differentially expressed (DE) genes are only differentially transcribed in one strain, with 90% of DE genes differentially transcribed in nine strains or fewer. (D) A total of 420 gene ontology (GO) terms are enriched (purple) among the downregulated genes in at least one deletion strain; if a GO term is not enriched in a deletion strain's downregulated genes, the heatmap tile is white. Both axes are hierarchically clustered to group those deletion strains that share enriched ontologies. Numbers in parentheses to left of plot show the number of strains in which the GO term is enriched.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Native elongating transcript sequencing (NET-seq) screen data identifies largely different groups of genes with varying functions that are differentially expressed across deletion strains.

common functions were also revealed by GO analysis. The GO term 'Cvt complex' (GO: 0034270), a complex involved in autophagy, was enriched among upregulated genes in 17 deletion strains (Figure 1—figure supplement 1C), and the GO term 'membrane biogenesis' (GO: 0101025), was enriched among downregulated genes in 12 deletion strains (Figure 1D). These trends are consistent with the upregulation of autophagy and the downregulation of growth when transcription factors are deleted.

Antisense transcription is misregulated upon deletion of transcription regulatory factors

NET-seq is uniquely suited to detect antisense transcription (Figure 2A–B), and since we included antisense transcripts in differential expression analysis, we have a direct readout of their expression. We linked every antisense transcript to the GO term 'antisense transcription' (GO: 9999999). This GO term was significantly enriched among downregulated genes in 14 strains and among upregulated genes in 6 strains (Figure 1D, Figure 1—figure supplement 1C).

To determine the effects of removing transcriptional regulators on antisense transcription, we visualized the spread of log₂-fold changes vs wild-type of these antisense transcripts (Figure 2C). Our data revealed a continuum of median antisense transcription, with that of the wild-type strain near the middle of the range. The strains in which we observed the largest decrease in antisense transcription were those lacking factors relating to transcription elongation, such as Elf1, Rtt103, and the Pol II subunit Rpb4, suggesting an asymmetry in the impact of elongation factors on sense and antisense transcripts. The factors whose deletions led to the largest increase in the antisense transcription were those involved in the regulation of histone acetylation, including members of the Rpd3S–Set2 pathway (Set2) and the major histone H4 acetyltransferase complex NuA4 (Eaf1), emphasizing the role of acetylation/deacetylation in antisense transcription (Carrozza et al., 2005; Churchman and Weissman, 2011; Krogan et al., 2003; Murray et al., 2015; Murray and Mellor, 2016).

In many strains, changes in antisense transcription occurred in specific locations (Figure 2—figure supplement 1A). For example, increases in antisense transcription in the *dst1Δ* strain occurred primarily at the 3' end; in the *set2Δ* strain, antisense transcription increased uniformly across the gene; and in the *eaf1Δ* strain, antisense transcription increased within the gene, but not at the 3' end (Figure 2D–F). These findings imply that antisense transcription is a combination of different transcriptional activities regulated by separate sets of factors. Many of these factors had been identified as regulators of antisense transcription using northern blot analysis, microarrays, or other strategies (Carrozza et al., 2005; Li et al., 2009). In these cases, NET-seq analysis provides a higher-resolution picture that confirms and complements these earlier findings.

Antisense transcription can repress or activate sense transcription through direct (transcriptional interference) or indirect mechanisms, such as altered chromatin states (Houseley et al., 2008; Lenstra et al., 2015; Martens et al., 2004; Nevers et al., 2018; Uhler et al., 2007). However, it remains unclear whether changes in transcriptional output are generally connected to changes in antisense transcription across regulatory contexts (Murray and Mellor, 2016). We compared changes in gene transcription to changes in Pol II antisense transcription across a range of transcription regulatory landscapes. We found no correlation between antisense and sense transcriptional outputs when considering all strains together (Figure 2—figure supplement 1B). To determine whether any factor acts as a link between antisense and sense transcription, we plotted all Pearson *r* values across each strain individually (Figure 2—figure supplement 1C). Values ranged from –0.22 to 0.36, suggesting

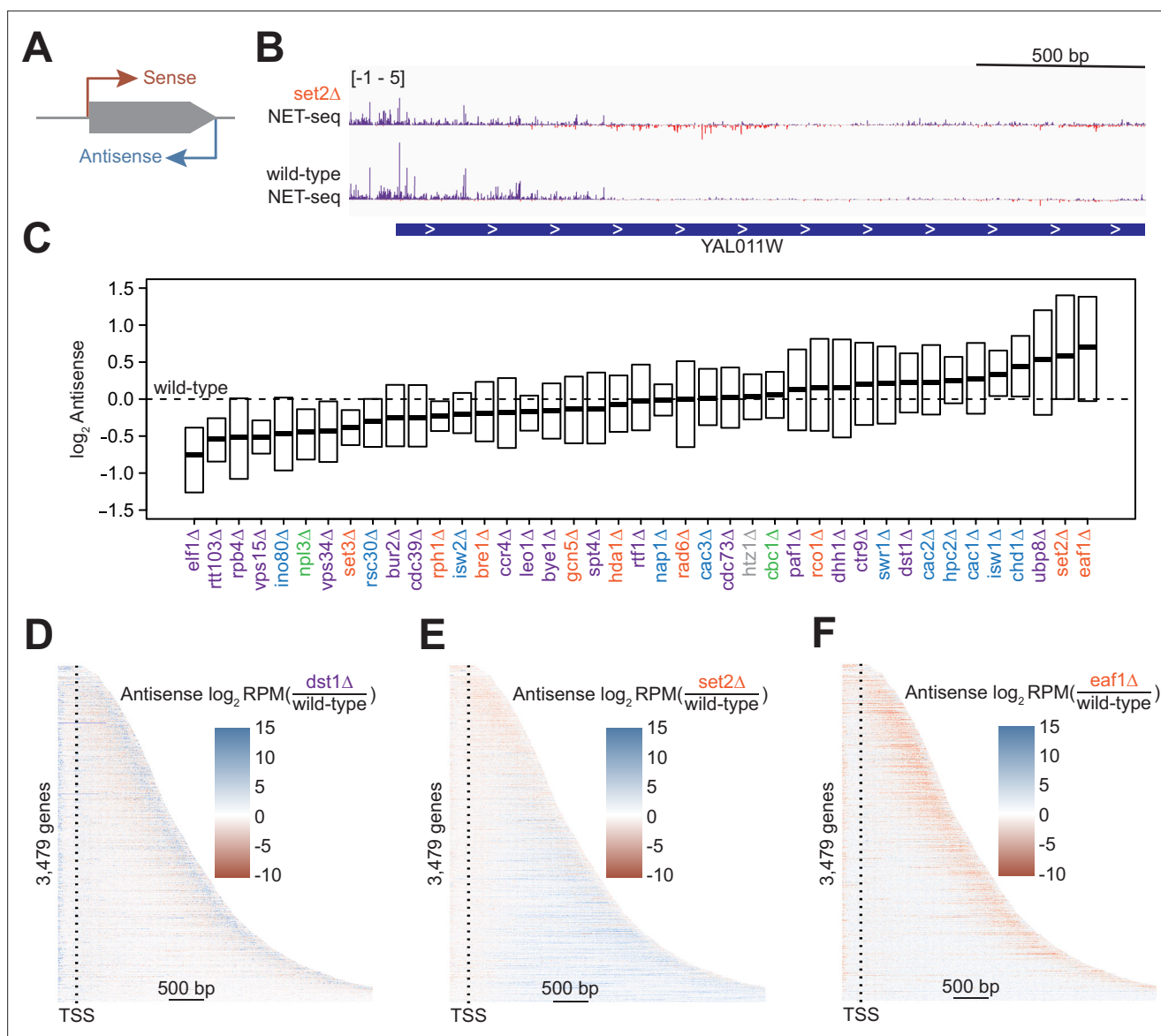


Figure 2. Antisense transcription is altered in most deletion strains. **(A)** Cartoon illustrating sense and antisense transcription of an example gene on the positive strand. **(B)** Wild-type and *set2Δ* native elongating transcript sequencing (NET-seq) data at YAL011W. Sense and antisense are displayed in purple and red, respectively. **(C)** Fold change in antisense transcription for each deletion strain compared to wild-type reveals that some strains have dramatically increased antisense transcription while others have much less than wild-type. Whiskers and outliers are omitted from visualization. **(D)** Heatmap of fold change in antisense transcription in the *dst1Δ* strain compared to wild-type reveals that most antisense transcription in the *dst1Δ* strain originates from the 3' end of genes. **(E–F)** Same as in **(D)**, for *set2Δ* and *eaf1Δ*, respectively.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Antisense transcription is largely uncorrelated with gene length and not uniformly distributed across gene bodies.

that antisense transcription levels do not generally affect sense transcription in any of the regulatory contexts that we analyzed.

Peaks of Pol II density across the gene body are altered in the absence of key transcription regulators

We found that Pol II density increases at loci critical for gene regulation, namely the transcription start sites (TSS), poly(A) sites, and splice sites (**Figure 3—figure supplement 1A–D**). At the 5' ends of genes, loss of Dst1, a homolog of the general transcription elongation factor TFIIIS, dramatically increased Pol II pausing just downstream of the TSS (**Figure 3A**). We also observed peaks in Pol II

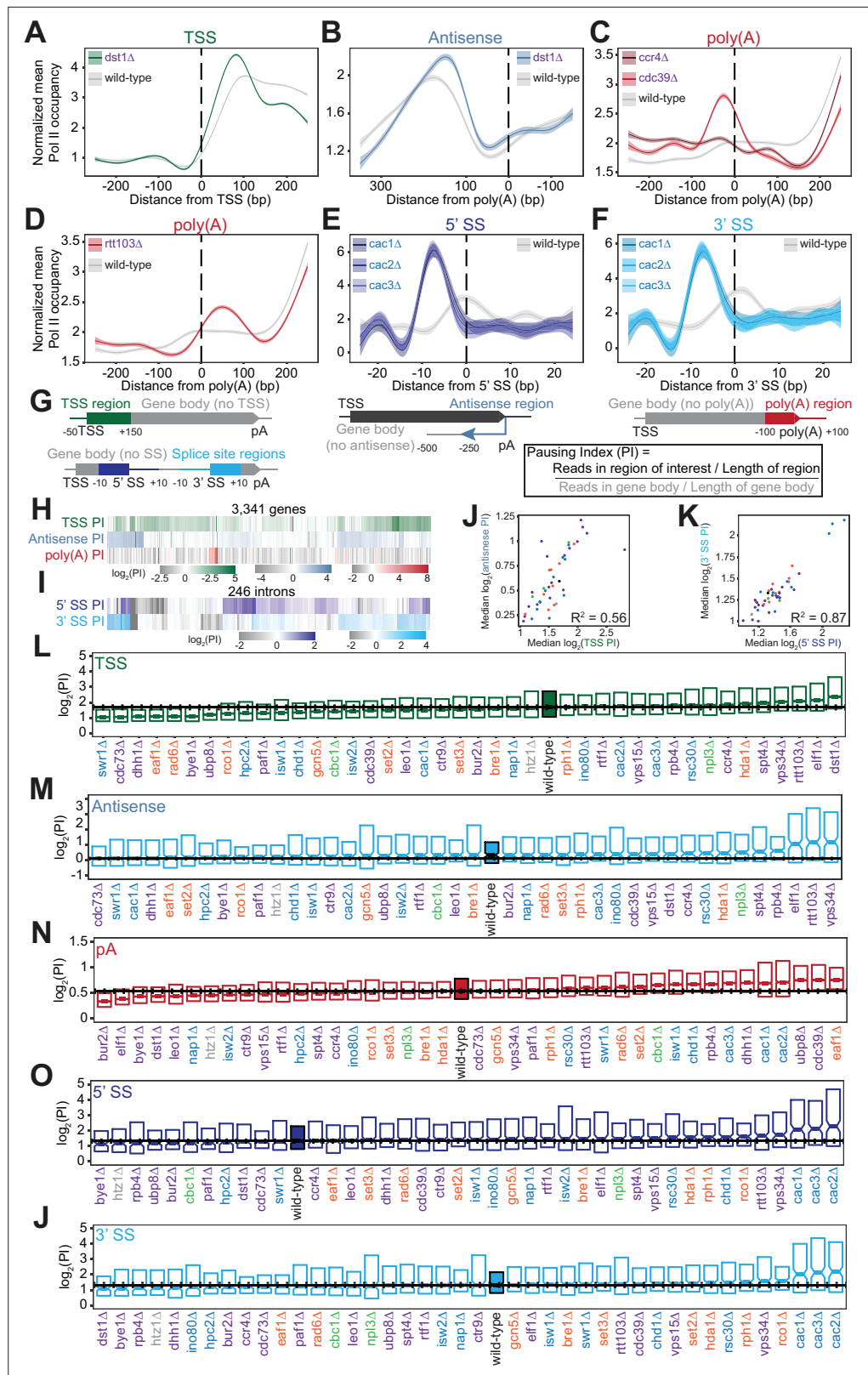


Figure 3. Polymerase II (Pol II) density is increased around transcription start sites (TSS), polyadenylation sites, and splice sites (SS). **(A)** Metagenes plot of normalized mean Pol II occupancy and the surrounding 95% confidence interval for the 500 bp surrounding the most abundant annotated TSS (*Pelechano et al., 2013*) ($n=2415$ genes). Metagenes for *dst1Δ* (green) can be compared to the Pol II density in the wild-type strain (gray). **(B)** Normalized
 Figure 3 continued on next page

Figure 3 continued

mean Pol II occupancy and the surrounding 95% confidence interval for the 600 bp surrounding the most abundant annotated poly(A) sites (**Pelechano et al., 2013**) in the antisense orientation. Metagene for *dst1Δ* (blue) can be compared to the Pol II density in the wild-type strain (gray). (C) Normalized mean Pol II occupancy and the surrounding 95% confidence interval for the 500 bp surrounding the most abundant annotated poly(A) sites (**Pelechano et al., 2013**). Metagenes for subunits of the Ccr4-NOT complex deleted (red) can be compared to the Pol II density in the wild-type strain (gray). (D) Same as (C), for *rtt103Δ*. (E–F) Normalized mean Pol II occupancy and the surrounding 95% confidence interval for the 50 bp surrounding annotated 5' and 3' splice sites (SS). Metagenes for subunits of the Caf1 complex deleted (blue) can be compared to the Pol II density in the wild-type strain (gray). (G) Cartoon and equation illustrating pausing index (PI) calculation. (H) PI for the TSS (green), polyadenylation [poly(A)] (red), and 3' antisense (blue) regions across genes. Horizontal axis is hierarchically clustered, revealing TSS, poly(A), and antisense pausing indices for genes in wild-type yeast. (I) Same as (H), for 5' and 3' SS pausing indices. (J) Scatter plot of the median pausing indices in the TSS and 3' antisense regions for all deletion strains. Relationship was quantified using Pearson correlation. (K) Same as in (J), comparing pausing the 5' and 3' SS surrounding introns. (L) Boxplot of TSS PI distributions in each deletion strain, ordered by median PI. Horizontal solid line indicates median value for wild-type yeast; dotted lines indicate the 45th and 55th percentile of wild-type PI values. (M–P) Same as (L), for 3' antisense PI, poly(A) site PI, 5' SS PI, and 3' SS PI.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Heatmaps of polymerase II (Pol II) density around RNA processing sites reveal differences in polymerase behavior across deletion strains, which can have functional consequences in specific deletion strains.

Figure supplement 2. Polymerase II density is increased around RNA processing sites to varying degrees across deletion strains.

density at the start of antisense transcripts opposite the 3' ends of genes. Interestingly, deletion of *DST1* had an effect on antisense transcription similar to its impact on sense transcription (**Figure 3B**).

At the 3' ends of genes, we observed changes in Pol II density upon loss of factors that regulate termination or polyadenylation. The screen included two subunits of the Ccr4-Not complex, which plays many roles in gene regulation including deadenylation (**Figure 3C**; **Funakoshi et al., 2007**; **Raisch et al., 2019**; **Temme et al., 2014**; **Tucker et al., 2002**; **Wahle and Winkler, 2013**; **Yamashita et al., 2005**; **Yi et al., 2018**). Deletion of the scaffolding Cdc39 subunit of the complex resulted in substantial pausing before poly(A) sites, followed by reduced Pol II density. By contrast, loss of the catalytic Ccr4 subunit decreased density only downstream, with a much less prominent upstream pause (**Figure 3C**). Loss of proteins more directly involved in transcription termination, such as Rtt103, resulted in Pol II stalling just downstream of poly(A) sites, suggesting that Pol II may slow down during recruitment of this termination factor (**Figure 3D**). In these deletion strains and others, the locations of 3'-end Pol II peaks varied, with some strains exhibiting a Pol II peak before poly(A) sites and others exhibiting a peak after (**Figure 3—figure supplement 1C**) indicating that Pol II is controlled both before and after poly(A) sites.

Pol II density increases around splice sites upon the loss of several transcription regulators. For example, pause indices increased most strongly when any of the CAF-I complex components (i.e. Cac1, Cac2, Cac3) were deleted (**Figure 3E–F**). CAF-I promotes histone deposition onto newly synthesized DNA (**Kaufman et al., 1997**), and to the best of our knowledge has not been implicated in splicing. To determine whether splicing is altered upon loss of CAF-1, we analyzed *cac2Δ* RNA-seq data (**Hewawasam et al., 2018**). We detected a modest but statistically significant increase in splicing in the *cac2Δ* strain relative to the wild-type ($p=0.02$; **Figure 3—figure supplement 1E, F**). Thus, CAF-1 decreases Pol II density at splice sites and regulates splicing, suggesting that the complex links Pol II pausing with splicing efficiency.

To quantify Pol II pausing at each site, we defined a pausing index (PI), a length-normalized metric comparing Pol II density in the region of interest to that in the rest of the gene body (**Figure 3G**). Interestingly, genes with a high PI in one location did not tend to have a high index for other locations (**Figure 3H–I**). Overall, at the per gene level, there was a poor correlation between all pausing indices in the wild-type strain (e.g. TSS PI vs poly(A) PI for each gene has $R^2=0.06$; all $R^2 \leq 0.10$, $p > 0.05$; **Figure 3—figure supplement 2A**). Even across each intron, pause indices differ at 5'- and 3'-splice sites although strong pausing occurs at 5' splice sites as often as at 3' splice sites (**Figure 3—figure**

supplement 2B). Thus, pause indices vary across each gene, from the TSS to poly(A) sites, suggesting that each region of high Pol II density is regulated in a different manner.

Across deletion strains, the median PI varied, with the wild-type indices lying near the middle of the dynamic range (**Figure 3L–P**, S4D–H). For example, the median TSS PI ranged from 1.06 in *cdc73Δ* to 2.81 in *dst1Δ*, with wild-type at 1.68 (**Figure 3L**, S3A). The levels of antisense pausing also vary substantially across the strains (**Figure 3M**).

We asked whether the same factors are implicated in regulating the different Pol II peaks. Indeed, there was a relatively strong correlation between median TSS pausing indices and antisense pausing indices across the deletion strains ($R^2=0.56$, $p<0.001$; **Figure 3J**). Of the 10 strains with the highest TSS pausing indices, 9 were also in the top 10 for median antisense pausing indices (**Figure 3L–M**). These strains tended to lack known elongation factors, such as *Dst1* and *Spt4*, indicating the role of transcription elongation factors in relieving pausing at the start of transcription. In addition, factors that modulate pausing at splice sites tended to do so at both sites overall, but not at the same intron ($R^2=0.87$, $p<0.001$; **Figure 3K**, S4B). However, we did not observe similar relationships between other pause indices (**Figure 3—figure supplement 2C**). For example, factors impacting pausing near the TSS do not have a similar impact at splice sites or at poly(A) sites, indicating that different mechanisms control Pol II pausing in different genic regions.

Pol II pausing locations are affected by deletion of transcription regulators

Along with identifying regions of elevated Pol II density, NET-seq data pinpoints precise positions that Pol II pauses within regions of high Pol II density and elsewhere. Because NET-seq is performed in bulk on a population of cells, only the sites that consistently induce pausing are observed, and we refer to these as ‘stereotypical’ pause positions. These precise sites of Pol II pausing at single nucleotides are reminiscent of in vitro RNA polymerase pausing observed at specific positions of DNA templates (*Galburt et al., 2007; Hodges et al., 2009; Kingston and Chamberlin, 1981; Mayer et al., 2017; Wang et al., 1998*). We systematically identified putative pause sites in strains with sufficient coverage as positions with read densities that deviate from the statistical fluctuations of the surrounding 200 nucleotides, modeled as a negative binomial distribution (>3 standard deviations from the mean; **Figure 4A–B**). Using an irreproducibility discovery rate (IDR) analysis, the putative pause sites are ranked and compared across replicates (*Landt et al., 2012; Li et al., 2011*). Pause sites that correspond across replicates using an IDR threshold of 1% are considered reproducible and used for downstream analyses. Approximately, one-third of the initially called pause sites is determined reproducible between two wild-type replicates using this criteria, but the majority of reproducible pause sites using various combinations of replicates overlap (**Figure 4—figure supplement 1A, B**). Stereotypical pause sites in NET-seq data represent loci where Pol II pauses in many cells and represent a fraction of the overall pausing by Pol II. The *E. coli* RNA polymerase pauses both at specific pause sites and randomly across a DNA template (*Adelman et al., 2002; Neuman et al., 2003*). Thus, Pol II is likely to similarly pause ubiquitously across gene bodies in noncanonical ways that would not lead to a detectable signal in NET-seq data. Nevertheless, the stereotypical pause sites identified here provide insight into the underlying features that induce Pol II pausing.

In NET-seq analysis and other 3′ end mapping approaches, mispriming events during reverse transcription (RT) can occur when the RT primer anneals internally within the nascent RNA rather than with the oligo ligated to the 3′ end (*Gajos et al., 2021; Mayer et al., 2015*). RT mispriming is far more likely to occur on nascent RNA derived from large genomes as there are many more sequences that could be recognized by the RT primer. Such events can be identified computationally and removed as the reads lack a unique molecular identifier sequence and align proximal to sites complementary to the RT primer. To reduce their occurrence in the first place a nested NET-seq library strategy has been employed to lessen mispriming in human NET-seq analysis (*Gajos et al., 2021*). In yeast, we found that the nested NET-seq library approach does not change the number of pauses identified (**Figure 4—figure supplement 1C**) nor does it decrease the fraction of pause sites with adapter-like sequence downstream, which is expected at sites of mispriming (**Figure 4—figure supplement 1D**). We similarly found that the number of pauses identified with and without removing reads with identical molecular barcodes (‘PCR duplicates’) shows virtually the same number of pause sites (**Figure 4—figure supplement 1E**). Before identifying the locations of pause sites, we computationally removed all reads that

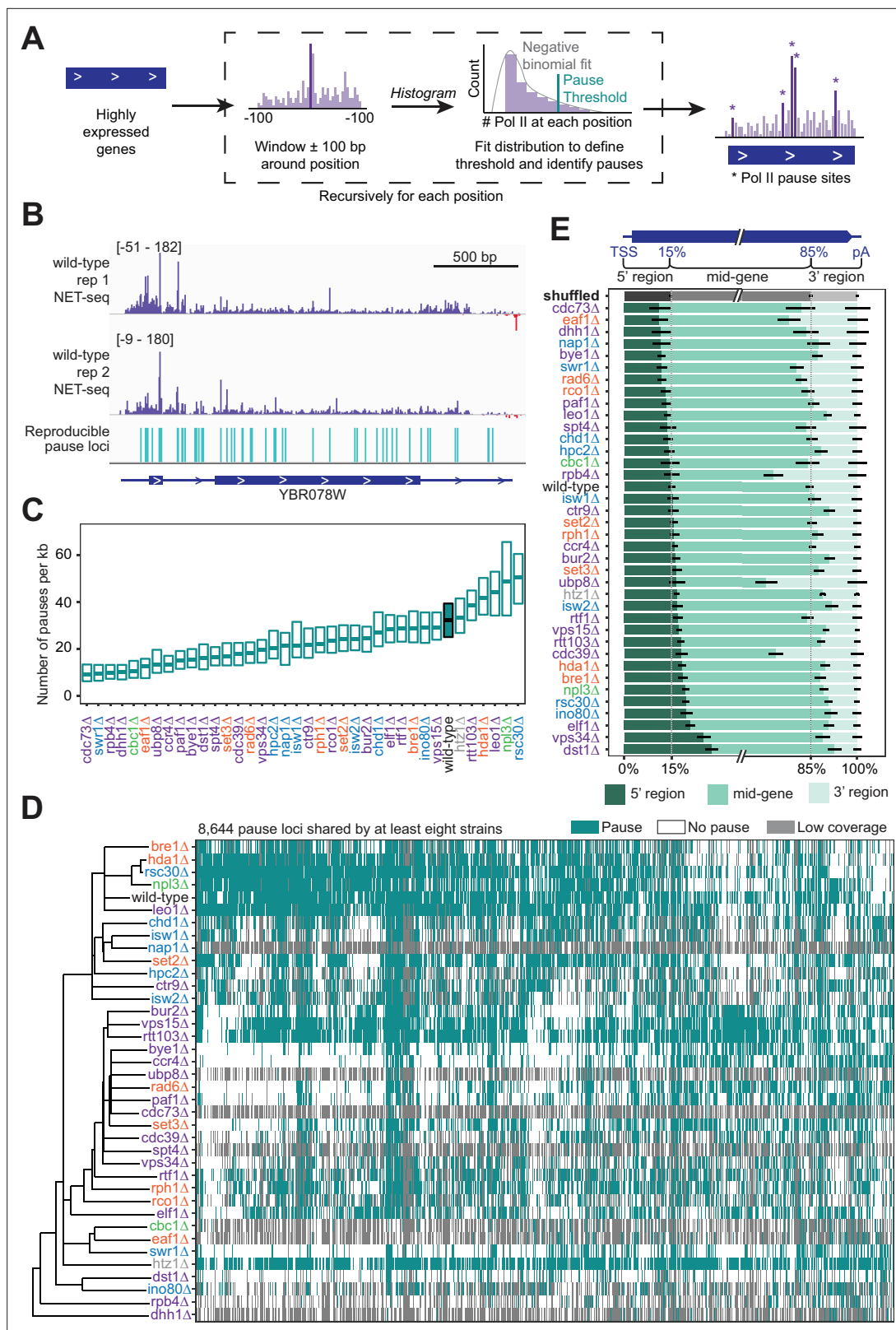


Figure 4. Trends in polymerase II (Pol II) pausing behavior at single-nucleotide resolution across deletion strains. **(A)** Cartoon illustrating algorithm for robust and reproducible Pol II pause detection. **(B)** Example of Pol II density on the positive (purple) and negative (red) strands, as measured by native elongating transcript sequencing (NET-seq) in two wild-type replicates. Pauses that meet the 1% irreproducibility discovery rate (IDR) reproducibility threshold are shown as blue vertical lines. **(C)** Boxplot of the distribution of Pol II pause densities, the number of pauses per kilobase examined, in *Figure 4 continued on next page*

Figure 4 continued

each deletion strain, ordered by median pausing density. Whiskers and outliers were removed for visualization. (D) Hierarchically clustered heatmap of 8644 Pol II pause loci across the genome reveals locations of pauses shared by multiple deletion strains. Heatmap is colored based on if that locus was identified as a pause (teal), not a pause (white), or if there was not sufficient coverage to determine pause status (gray). Analyses conducted only on deletion strains with biological replicates and only at loci at which there was enough coverage to determine the absence of a Pol II pause in at least one deletion strain. (E) The percent of Pol II pause loci located in the 5' gene region, mid-gene, and 3' gene region varies across deletion strains. The 5' gene region was identified for each well-expressed gene as extending from the transcription start site to the 15th percentile of the gene length. Similarly, the 3' gene region was defined as the last 15th percentile of the gene length, with the mid-gene region spanning in between. The control (gray) was created by scrambling all identified pauses across all deletion strains within the genes they were identified in. Rows are ordered by the percent of pauses found in the 5' region. Bars represent the 95% confidence intervals across all expressed genes.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Polymerase II (Pol II) pausing behavior at single-nucleotide resolution across deletion strains reveals that pausing is balanced and dynamic in wild-type.

are due to RT mispriming, but to avoid possible distortions that occur during deduplication, we did not remove putative PCR duplicates (Fu et al., 2018; Parekh et al., 2016).

We calculated the pause site density, or the number of sites per kilobase, for genes that had sufficient coverage. The density varied widely across deletion strains (Figure 4C), which cannot be explained by differences in sequencing depth across deletion strains ($R^2=0.003$, $p=0.743$; Figure 4—figure supplement 1F). In the wild-type strain, we found Pol II pause sites every 33 bp on average. Some of the deletion strains exhibited more pausing overall at stereotypical pause sites; for example, upon loss of Rsc30, a subunit of the RSC chromatin remodeling complex, 33% of all NET-seq reads mapping to highly expressed genes constituted pause sites, versus only 21% in the wild-type (Figure 4—figure supplement 1G). Thus, the RSC complex obscures Pol II pause sites, which is likely related to its role in diminishing the nucleosomal barrier to Pol II elongation (Carey et al., 2006). Perhaps unexpectedly, loss of canonical transcription elongation factors, such as Spt4 and Dst1, resulted in a lower pause site density relative to the wild-type (Figure 4C). However, pause site density describes only one feature of Pol II elongation. The density includes only the stereotypical locations at which Pol II typically pauses in many cells, so it is not a measure of the absolute frequency of Pol II pausing. In addition, the density is not related to the Pol II catalysis rate. Thus, these transcription elongation factors may facilitate other aspects of transcription elongation or they may act locally to influence Pol II during specific points of regulation, consistent with their impact on peaks of Pol II density only near TSS (Figure 3L).

The pause loci for each strain included many that were not observed in wild-type yeast (Figure 4D). Indeed, when the sets of pause loci are used to cluster deletion strains by principal component analysis, the wild-type strain stands away from most strains (Figure 4—figure supplement 1H). However, some deletion strains shared many pause sites with those observed under in the wild-type: 81% of pause sites identified in wild-type yeast were also identified in the *htz1Δ* strain, consistent with its confined role at the +1 nucleosome (Bagchi et al., 2020; Zhang et al., 2005).

We wondered whether loss of related factors would lead to the same sets of pause sites. We first identified all pause sites observed in at least eight strains and used the presence or absence of these pauses in each strain to perform hierarchical clustering (Figure 4D). *dst1Δ* pause sites clustered far away from those in wild-type cells, consistent with the backtracking role of Dst1 that leads to downstream-shifted pause sites (Churchman and Weissman, 2011; Noe Gonzalez et al., 2021). H2B ubiquitination increases the nucleosomal barrier to Pol II (Chen et al., 2019), so alterations to histone ubiquitination might lead to new pause sites. Interestingly, pause sites after the loss of Rad6, Ubp8, Paf1, and Cdc73 all cluster together. Rad6 and Ubp8 ubiquitinate and deubiquitinate H2B, respectively (Amerik et al., 2000; Jentsch et al., 1987). Paf1 and Cdc73, members of the Paf1 complex, are responsible for recruiting Rad6 to chromatin (Kim and Roeder, 2009). The clustering of these factors indicates a role for H2B ubiquitination in determining the locations of many pause sites. Finally, we figured that differences in nucleosome positioning may lead to differential pause sites usage, so we inspected how pause sites change after the loss of different chromatin remodelers. Interestingly, we observed that loss of ISWI and CHD chromatin remodelers, *lsw1*, *lsw2*, and *Chd1*, leads to pause sites that cluster together (Figure 4D). For example, most of the pause sites observed in *lsw1Δ* (76%) were also observed in *chd1Δ*, consistent with their joint roles in maintaining chromatin structure (Ocampo et al., 2016; Smolle et al., 2012). In contrast, loss of INO80, SWR1, and SWI/SNF family remodelers,

Ino80, Rsc30, and Swr1, all leads to distinct sets of pause sites consistent with their separate roles in chromatin remodeling (**Figure 4D**; **Singh and Mueller-Planitz, 2021**).

Pol II pause sites in the wild-type strain were distributed evenly throughout gene bodies (**Figure 4E**). By contrast, deletion strains exhibited a range between twofold decreased and twofold increased Pol II pause sites in the 3' regions of genes, with slightly less variability at the 5' regions of genes relative to a scrambled control or wild-type pausing (**Figure 4E**). The enrichment of pause sites at 5' end and 3' regions generally corresponds with our PI results (**Figure 3H, L and N**). For example, deletion of *DST1* approximately doubled pause loci in the 5' regions at the expense of pausing in 3' regions. This localized effect exemplifies how overall pause density (see **Figure 4C**) of a gene could be decreased in a deletion strain lacking a canonical elongation factor. However, in general, changes in 5' vs 3' pause sites in deletion strains were not correlated (**Figure 4E**). We find substantially more pause sites at the 3' regions of genes in *rpb4Δ*. Rpb4 is a Pol II subunit that dissociates with the complex at the ends of genes (**Mosley et al., 2013**) and is responsible for sustained transcription elongation through the 3' ends of genes (**Runner et al., 2008**). Thus, Rpb4 prevents Pol II from pausing at the 3' regions of genes that may protect from premature termination before the canonical 3' cleavage site is transcribed. Similarly, more 3' pause sites are found in the *ubp8Δ* strain, consistent with the global increase in this strain of H2B ubiquitination, a mark that increases the nucleosomal barrier to Pol II and is coincident with Pol II pausing at transcription termination sites (**Bonnet et al., 2014**; **Chen et al., 2019**; **Harlen et al., 2016**). Together, these data show how the chromatin landscape and transcriptional regulatory network of the cell dictate stereotypical sites of Pol II pausing that in turn controls where and for how long Pol II pauses during elongation.

Chromatin and DNA features can accurately classify Pol II pausing locations in deletions strains

Given the number of reproducible pause sites we identified, we next investigated whether we could determine which genomic features, if any, were responsible for the stereotypical pause sites. In vitro studies have shown that Pol II pausing has many causes, including specific DNA sequences, nucleosomes, and histone modifications (**Bintu et al., 2012**; **Herbert et al., 2006**; **Hodges et al., 2009**; **Kassavetis and Chamberlin, 1981**; **Kireeva et al., 2005**; **Kireeva and Kashlev, 2009**; **Shaevitz et al., 2003**). In vivo, the dominant factors globally associated with Pol II pause sites remain unclear, although sequence elements, transcription factors, nucleosomes, and CTD modifications have all been connected to Pol II pausing (**Alexander et al., 2010**; **Churchman and Weissman, 2011**; **Gajos et al., 2021**; **Nechaev et al., 2010**; **Noe Gonzalez et al., 2021**; **Nojima et al., 2018**; **Shukla et al., 2011**). Recently, DNA sequence and shape were shown to be important contributors to pause site locations in human cells (**Gajos et al., 2021**). We first asked whether specific DNA sequences were connected with Pol II pausing loci. Previous studies reported that Pol II has a strong bias toward pausing at adenine (**Churchman and Weissman, 2011**), which we also observed here. More specifically, we observed a 3.4-fold enrichment of real Pol II pause sites at TAT trinucleotide sequences relative to shuffled control sites in the same well-expressed genes (**Figure 5A**). The shape of the DNA itself, as predicted from sequence, also appears to inform the location and propensity for Pol II to stall: DNA low helix twist values were more common under real pause loci than in the shuffled control (**Figure 5B**). These observations were consistent, as the AT dinucleotide step has a low average twist angle of 32.1° (**Ussery, 2002**).

Beyond the trinucleotides, significantly enriched sequence motifs were also associated with Pol II pause sites in most deletion strains (**Supplementary file 4**), including three motifs related to pauses in the wild-type strain (**Figure 5C**). Notably, not all motifs are shared across strains, and upon deletion of some factors, new motifs were associated with Pol II pause sites. 13 of the 26 identified sequence motifs with high relative entropies significantly matched known transcription factor binding site motifs (**Figure 5D**). Thus, it is likely that Pol II pause sites can partially, but not fully, be explained by DNA sequence and/or proteins binding to DNA.

In addition to the structure of the DNA itself, chromatin features, such as nucleosome positions and histone modifications, are also connected to Pol II pausing behavior. To search broadly for genomic features underlying sites of Pol II pausing, we evaluated 51 features (**Supplementary file 5**), including nucleotide sequence, DNA shape, position of pauses within a gene, histone modifications, and Pol II CTD phosphorylation marks. 35 out of 42 exhibited a statistically significant difference between

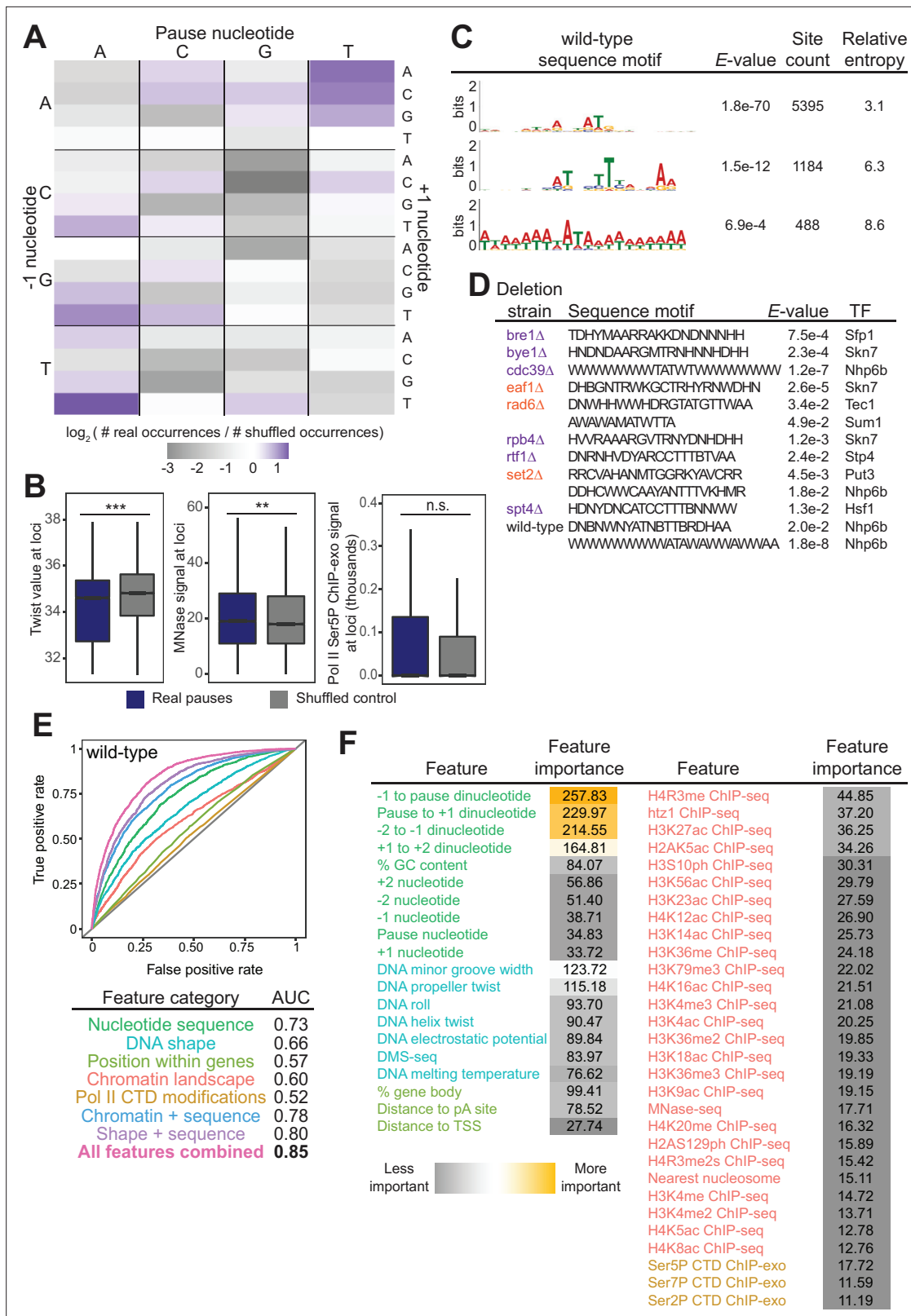


Figure 5. Chromatin and DNA features explain the location of some polymerase II (Pol II) pauses in wild-type yeast. **(A)** Heatmap illustrating the relative frequency of each trinucleotide sequence surrounding real and shuffled control pauses centered on Pol II pauses identified in wild-type. **(B,** left) Comparison in the distribution of values for twist values underlying Pol II pauses in wild-type yeast ($n=13,994$) compared to a shuffled control, in which the same number of pauses is shuffled, maintaining the same number of pauses within each well-expressed gene. Differences between the real
Figure 5 continued on next page

Figure 5 continued

and shuffled distributions were determined as significantly significant by a Student's t-test with Bonferroni correction for multiple hypotheses. p-values are reported in **Supplementary file 5**. (* adjusted p-value ≤ 0.05 ; ** adjusted p-value ≤ 0.01 ; *** adjusted p-value ≤ 0.001). Also shown for MNase-seq signal (center) and Ser5P CTD ChIP-exo signal (right). (C) Table showing the three significant motifs identified under Pol II pauses in the wild-type strain. All analyses were performed using the MEME suite of tools. Significant motifs were those with an E-value greater than 0.05. Pause sites were scrambled within well-expressed genes to be used as a negative control and to calculate enrichment of motifs. (D) Table with all sequence motifs underlying pauses across deletion strains that are significantly similar to known transcription factor binding motifs. Only the top match, as assessed by E-value, is reported. (E) Receiver operating characteristic curve from a random forest classifier that measures the predictive value of chromatin and DNA features on Pol II pauses in wild-type yeast (10,495 training and 3499 training loci). (F) Table of all features used in random forest classifier for pause loci classification and the importance of each feature. Feature importance is calculated as the mean decrease in accuracy upon removing that feature from the model.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Chromatin and DNA features explain the location of some polymerase II (Pol II) pauses in wild-type yeast using a random forest classifier.

real wild-type pause sites and shuffled controls (the remaining 9 of the 51 are sequence features that cannot be compared on a numeric scale) (**Figure 5—figure supplement 1A, Supplementary file 6**). For example, the MNase-seq signal around pause loci and the distance to the nearest nucleosome differed significantly between real and shuffled pause sites (**Figure 5B, S6A**), consistent with observations of pauses at nucleosomes (**Churchman and Weissman, 2011**). Interestingly, Ser2, Ser5, and Ser7 phosphorylation of the Pol II CTD did not differ relative to random positions, indicating that connections between Pol II phosphorylation and pausing at intron-exon boundaries are specific to pausing at those loci (**Alexander et al., 2010**). Among the features that differed significantly was DNA melting temperature, which was previously shown to influence Pol II stalling (**Nechaev et al., 2010**).

To determine which features may underlie where Pol II pauses, we created a random forest classifier to discriminate between real and shuffled control Pol II pause sites based on the surrounding chromatin and DNA features. A random forest classifier using all 51 features performed well (AUC = 0.85, **Figure 5E**) relative to a random model (AUC = 0.5) at classifying Pol II pauses in wild-type yeast. Which features contribute the most to the random forest classifier can help shape models for the molecular underpinnings of stereotypical Pol II pausing. The most critical features for accurate identification of Pol II pause sites were DNA sequence surrounding the pause locus and topology features of the DNA at that locus (**Figure 5F**). A reliance on DNA sequence and DNA shape for determining pause sites was also observed in human NET-seq data despite a different DNA motif (**Gajos et al., 2021**). Together, these analyses showed that DNA sequence and shape contribute strongly to Pol II pause locations, but their effects are enhanced by many other features.

To ask whether features underlying Pol II pausing vary in different regulatory and chromatin landscapes, we built random forest models for each deletion strain. Across all deletion strains, an AUC of at least 0.78 was attained. These AUC values were only partially correlated with the total number of pauses detected in each deletion strain ($R^2=0.37$, $p=0.000064$; **Figure 6—figure supplement 1A**). Although nucleotide sequence and DNA shape were the most important features for classifying Pol II pause loci in the wild-type and many deletion strains, models for a subset of strains (including *cdc39Δ*, *dst1Δ*, *ubp8Δ*) revealed that wild-type chromatin modifications were more powerful for Pol II classification (**Figure 6A, S7B-E**). We next performed a transfer of learning analysis to ask how each model would perform when classifying pauses in other strains. When trained on Pol II pause sites identified in wild-type yeast, the AUC when testing on pauses across all other strains ranged from 0.53 (*cbc1Δ*) to 0.82 (*vps15Δ*), revealing the differences across the strains (**Figure 6B**). We previously observed that loss of Dst1 leads to ~75% of pause sites to shift downstream (**Churchman and Weissman, 2011**). Thus, training a model on *dst1Δ* pause sites should not do well to classify pauses in another strain. Indeed, a model trained on *dst1Δ* pause sites performed well in classifying *dst1Δ* pause sites (AUC = 0.83); however, it performed the worst of all models in classifying pause sites in other deletion strains, obtaining a median AUC of 0.63 across them. These models indicate that the nucleotide sequence, DNA topology, position within a gene, and chromatin landscape all play roles in determining the location of Pol II pauses during transcription elongation.

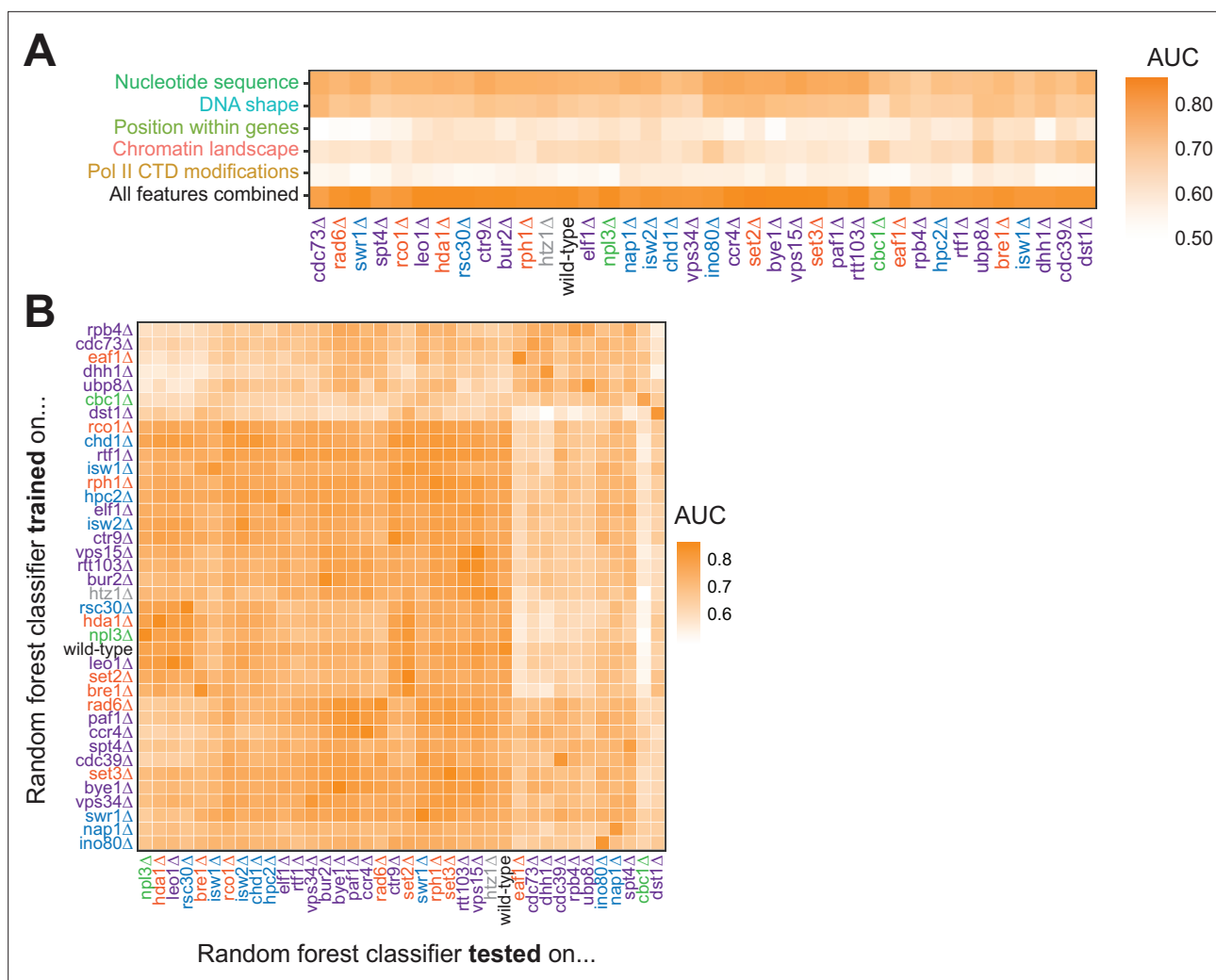


Figure 6. Random forest classifiers identify polymerase II (Pol II) pause loci across deletion strains, with different feature importance values across deletion strains. **(A)** Heatmap illustrating the mean AUC for the random forest classifier when trained (75% of loci) and tested (25% of loci) on each deletion strain. Deletion strains are hierarchically clustered along the x-axis. **(B)** Heatmap showing the AUC values from random forest classifiers trained on all pauses from one deletion strain (y-axis) and tested on those unique pauses observed in another deletion strain (x-axis). Both axes are hierarchically clustered to reveal similarities in AUC values across deletion strains. Tiles when the same training and testing strain are indicated are colored according to the AUC for that deletion strain when 75% of pauses in that deletion strain are used for training and the remaining 25% are used for testing as reported in **(A)**.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Random forest classifiers can predict polymerase II pause loci across deletion strains, with different feature importance values across deletion strains.

Discussion

Advances in high-throughput sequencing of nascent RNA have revealed that, in many eukaryotes, the vast majority of the genome is transcribed (*Hangauer et al., 2013; Struhl, 2007*). Nevertheless, this broad transcriptional activity is one of the most highly regulated processes within the cell. Multiple levels of regulation are orchestrated by DNA sequence, transcription factors, RNA processing factors, and chromatin modulators. Here, we used NET-seq to study 41 factors with connections to transcription elongation and discovered the remarkable tunability of transcription elongation. For all of the transcriptional phenotypes analyzed, the wild-type strain fell in the middle of the dynamic range observed across the deletion strains, revealing the intricate balance of transcriptional activity.

The 41 factors chosen for this study were previously annotated to regulate transcription elongation. However, loss of each factor had a unique impact on gene expression, suggesting that genes are

differentially sensitized to perturbations of the transcription regulatory network. Levels of antisense transcription in the deletion strains vary across a broad dynamic range, revealing that antisense transcription is finely tuned by many factors. Interestingly, loss of 20 factors decreased antisense transcription in cells (**Figure 2C**), indicating that it is possible to suppress antisense transcription further than what is observed in wild-type. Conversely, loss of 14 factors increased antisense transcription. Together, these results imply that wild-type antisense transcription is balanced by the influence of many factors and, in turn, can be precisely controlled. The possibility of tight control of antisense transcription indicates that regulatory mechanisms can exist where antisense transcription impacts sense transcription, consistent with the mechanisms described thus far (*Hongay et al., 2006; Houseley et al., 2008; Lenstra et al., 2015; Martens et al., 2004; Uhler et al., 2007*). Although, we did not observe a general correspondence between sense transcription and antisense transcription in this study.

Peaks of Pol II density were detected near TSSs, poly(A) sites, and both 5' and 3' splice sites. Interestingly, factors that impacted pausing at the 5' ends of genes were not the same as those that impacted pausing at 3' ends or at SS. Clearly, different mechanisms regulate Pol II pausing at different points during elongation. However, pausing around the TSS and pausing during antisense transcription were controlled by a similar set of factors that are highly enriched for established transcription elongation factors, such as *SPT4* and *DST1*. These findings suggest that there is a checkpoint early in transcription, in the sense and antisense directions.

Unexpectedly, we found that loss of the CAF-1 complex leads to pronounced Pol II peaks at 5' and 3' splice sites (**Figure 3E and F**). The CAF-1 complex is characterized as a chromatin assembly factor that promotes nucleosome assembly on newly synthesized DNA, sets the size of nucleosome depleted regions, and suppresses divergent transcription (*Fennessy and Owen-Hughes, 2016; Kaufman et al., 1997; Marquardt et al., 2014*). In addition, our findings connect the complex to splicing. It is tempting to speculate that loss of the CAF-1 complex leads to poorly deposited nucleosomes near SS, which alters Pol II pausing and co-transcriptional splicing.

Within the regions of elevated Pol II density (e.g. TSSs and SS) and across gene bodies are discrete pauses at single nucleotides that represent locations where Pol II has a higher propensity to pause. This set of positions varies substantially across the deletion strains (**Figure 4D**), indicating that there are a large number of possible pause sites, but the presence of regulatory factors modulates the pausing landscape such that they are not utilized. Our machine learning models of pause site preferences found that DNA sequence and shape are the most influential, followed by the chromatin landscape. We propose that the DNA template presents a varying energy landscape to the elongating Pol II through sequence variation and that nucleosome positions alter the landscape by lowering or enhancing pausing energetics and the associated chance of Pol II pausing. We also found that some transcription factor binding sites are enriched near pause sites, indicating a possible role for DNA binding proteins in Pol II pausing. A future analysis of the role of transcription factors, RNA binding proteins, and RNA structure in pausing would be an interesting avenue of investigation.

This work reveals the complex regulation of transcription elongation by a network of factors. In addition, it serves as a resource of NET-seq data to explore more specific hypothesis-driven research questions relating to individual factors and an open-source code base with which to analyze these data. Many of the transcription elongation regulators studied here are conserved in all domains of life, as are many of the transcriptional phenotypes we examined, including antisense transcription and Pol II pausing. These insights into transcription regulation in *S. cerevisiae* will serve as a foundation for learning more about transcription in multicellular eukaryotes.

Materials and methods

Yeast mutant generation

To create deletion mutants of the 41 factors analyzed, the parent strain YSC001 (BY4741 *rpb3::rpb3-3xFLAG NAT*) (*Churchman and Weissman, 2011*) was transformed with PCR products of the *HIS3* gene flanked by 40 bp of homology upstream and downstream of the start and stop codons for the gene of interest. Standard lithium acetate transformations were used.

NET-seq library generation

Cultures for NET-seq were prepared as described in *Churchman and Weissman, 2012*. Briefly, overnight cultures from single yeast colonies grown in Yeast Extract–Peptone–Dextrose (YPD) medium were diluted to $OD_{600}=0.05$ in 1 L of YPD medium and grown at 30°C shaking at 200 rpm until reaching an $OD_{600}=0.6$ –0.8. Cultures were then filtered over 0.45-mm pore size nitrocellulose filters (Whatman). Yeast was scraped off the filter with a spatula pre-chilled in liquid nitrogen and plunged directly into liquid nitrogen as described in *Churchman and Weissman, 2012*. Mixer mill pulverization was performed using the conditions described above for six cycles. NET-seq growth conditions, immunoprecipitations, and isolation of nascent RNA and library construction were carried out as described in *Churchman and Weissman, 2012*. A random hexamer sequence was added to the linker to improve ligation efficiency and allow for the removal of any library biases generated from the RT step as described in *Mayer et al., 2015*. After library construction, the size distribution of the library was determined by using a 2100 Bioanalyzer (Agilent), and library concentrations were determined by Qubit 2.0 fluorometer (Invitrogen). 3' end sequencing of all samples was carried out on an Illumina NextSeq 500 with a read length of 75 bp. For analysis of *cac1Δ*, *cac2Δ*, and *cac3Δ*, raw Fastq files were obtained from *Marquardt et al., 2014* and re-aligned using the parameters described below.

Processing and alignment of NET-seq data

The adapter sequence (ATCTCGTATGCCGTCTTCTGCTTG) was removed using cutadapt with the following parameters: `-O 3 m 1 --length-tag 'length=.'` Raw fastq files were filtered using PRINSEQ (<http://prinseq.sourceforge.net/>) with the following parameters: `-no_qual_header -min_len 7 min_qual_mean 20 -trim_right 1 -trim_ns_right 1 -trim_qual_right 20 -trim_qual_type mean -trim_qual_window 5 -trim_qual_step 1`. Random hexamer linker sequences (the first six nucleotides at the 5' end of the read) were removed using custom Python scripts, but remained associated with the read. Reads were then aligned to the SacCer3 genome obtained from the *Saccharomyces* Genome Database using the TopHat2 aligner (*Kim et al., 2013*) with the following parameters: `--read-mismatches 3 --read-gap-length 2 --read-edit-dist 3 --min-anchor-length 8 --splice-mismatches 1 --min-intron-length 50 --max-intron-length 1200 --max-insertion-length 3 --max-deletion-length 3 --num-threads --max-multihits 100 --library-type fr-firststrand --segment-mismatches 3 --no-coverage-search --segment-length 20 --min-coverage-intron 50 --max-coverage-intron 100000 --min-segment-intron 50 --max-segment-intron 500000 --b2-sensitive`. To avoid any biases toward favoring annotated regions, the alignment was performed without providing a transcriptome. RT mispriming events were identified and removed where molecular barcode sequences correspond exactly to the genomic sequence adjacent to the aligned read. With NET-seq, the 5' end of the sequencing, which corresponds to the 3' end of the nascent RNA fragment, is recorded with a custom Python script using the HTSeq package (*Anders et al., 2015*). NET-seq data were normalized by million mapped reads. Replicate correlations were performed comparing RPKM of each gene in each replicate; replicates were considered highly correlated with a Pearson correlation of $R^2 \geq 0.75$. Raw NET-seq data of highly correlated replicates were merged, and then re-normalized by million mapped reads. For analysis of *rco1Δ*, raw Fastq files were obtained from *Churchman and Weissman, 2011* and re-aligned using the parameters described, except without removal of hexamer sequences.

Differential gene transcription and gene ontology enrichment analysis

Differential transcription analysis between deletion strains (two replicates each) and wild-type strains (four replicates) was performed using DESeq2 (*Love et al., 2014*) for all sense transcription units annotated in *Xu et al., 2009*. To account for antisense transcription, matching antisense transcription units were added to the annotation, as long as they did not overlap with a known sense gene. These added antisense transcription units were ignored in reporting the number of differentially expressed genes (*Figure 1B and C; Figure 1—figure supplement 1B*). Genes were considered differentially transcribed if they had an adjusted p-value < 0.05 and an absolute \log_2 -fold change > 1.0 .

GO term enrichment analysis was performed with The Ontologizer (<http://ontologizer.de/intro/>) (*Bauer et al., 2008; Grossmann et al., 2007; Ashburner et al., 2000; The Gene Ontology Consortium, 2019*) using the parent-child analysis method. The GO term 'antisense transcription' (GO: 9999999) was added to the go.obo file, and this new GO term was associated with all antisense

transcription units described above by modifying the file `sdg.gaf`. Fold enrichment and adjusted p-value for each GO by deletion strain pair are reported in **Supplementary file 3**.

Antisense transcription

For analysis of antisense transcription, the coordinates of protein-coding transcription units from **Xu et al., 2009** were reversed and annotated as 'antiXXXX', where 'XXXX' is the name of the gene encoded on the sense strand. Those that overlapped known sense transcription units were removed. This expanded annotation file was used to produce read count tables for DESeq2. To generate antisense heatmaps, the \log_2 RPKM of NET-seq reads was used. Analysis at coding genes ranged from 250 bp upstream of the TSS to 4000 bp downstream of the coding TSS. To allow comparison between mutant and wild-type samples, a pseudocount of 1 was added to every position in all samples before calculating the \log_2 RPKM. Differential heatmaps were calculated by taking the \log_2 ratio of mutant/wild-type RPKM at each position.

Pausing index calculation

Pausing indices were calculated as the length-normalized Pol II density in the region of interest (−50 bp to +150 bp around TSS, ±100 bp around poly(A) sites, and ±10 bp around 5' and 3' splice sites) divided by the length-normalized Pol II density in the remainder of the gene, as illustrated in **Figure 3G**.

Metagene analysis

Only protein-coding, non-overlapping genes were included in the metagene analysis. The regions analyzed were −100 to +600 bp surrounding the most abundant TSS, −500 to +200 bp surrounding poly(A) sites, as identified in **Pelechano et al., 2013**, and ±25 bp surrounding annotated 3' and 5' splice sites. NET-seq signal across each region was normalized, and the Loess smoothed mean (span = 0.01) and 95% confidence interval are plotted for NET-seq generated from each deletion strain across each region of interest.

Splicing index calculation

Cac2Δ and wild-type RNA-seq data were retrieved from **Hewawasam et al., 2018** under the GEO accession number GSE98397. Splicing index calculations were determined for each gene by counting the number of reads that span exon junctions by at least three nucleotides and measuring the number of spliced reads divided by unspliced reads; splicing index = $2 * \text{spliced reads} / (5' \text{ SS unspliced} + 3' \text{ SS unspliced reads})$ as in **Drexler et al., 2020**.

Extracting pause positions

Pauses were identified in previously annotated transcription units (**Xu et al., 2009**) of well-expressed genes (average of >2 reads per base-pair in two replicates). Pauses were defined as having reads higher than three standard deviations above the mean of the surrounding 200 nucleotides which do not contain pauses. Mean and standard deviation were calculated from a negative binomial distribution fit to the region of interest. Pauses were required to have at least two reads regardless of the gene's sequencing coverage. Our analysis algorithm for identifying pause sites uses the IDR analysis, which is the standard for analyzing ENCODE ChiP-seq data (**Li et al., 2011; Landt et al., 2012**). Here, many pause sites are identified in each replicate and ranked. The peaks in each biological replicate are compared, starting with the strongest. When the ranks of the peaks stop corresponding, a transition point is identified and the lower ranked peaks are marked as irreproducible. The methodology does not require an arbitrary cutoff, and all pause sites are considered in the comparison between replicates, reducing false negatives. Pauses were considered reproducible and used in downstream analyses when the IDR is <1% between two replicates. To calculate the IDR of each pause, \log_{10} of pause strength (number of reads in pause) for each replicate was used as a proxy for pause score. IDR was calculated using the `est.IDR` function of the `idr` R package ($\mu = 3$, $\sigma = 1$, $\rho = 0.9$, $p = 0.5$) (**Li et al., 2011**). Reproducible pauses were visualized using the IGV genome browser (**Robinson et al., 2011**). Because the *cac1Δ*, *cac2Δ*, and *cac3Δ* strains were constructed by a different lab (**Marquardt et al., 2014**), these strains were excluded from these analyses. Additionally *gcn5Δ* was excluded because of low sequencing coverage resulting in only 15 genes passing the coverage threshold.

Pol II pausing location and strength

Pause density was calculated as the ratio of total number of pauses to the total length of the genome considered when extracting pause positions (combined length of all well-expressed genes in both replicates of each deletion strain). To identify deletions that induced similar pausing patterns, 8644 pauses were found to be shared in at least eight strains and in regions sufficiently covered in multiple deletion strains. Shared pauses were visualized with a heatmap, clustered on both axes using the eisenCluster correlation clustering method in the hybridHclust R package (*Chipman and Tibshirani, 2006*), which takes into account missing data (where there was not enough coverage to confidently identify pausing in a particular deletion strain). Similarity in pause loci was also visualized as a scatter plot of the first two principal components. When calculating distribution of pauses across the gene body, all genes in which pauses were identified were normalized in length; the 5' gene region was defined as the first 15% of each gene, the mid-gene region was defined as extending from the 15th percentile of gene length to the 85th percentile, and the 3' gene region was defined as starting at 85% of gene length and extending to the annotated poly(A) site. The scrambled control for the pausing location analysis was created by randomly scrambling all identified pauses in all deletion strains across the gene in which they were discovered.

Pol II pause loci sequence motifs

All analyses related to sequence motifs underlying pause loci were conducted using the MEME suite of tools (*Bailey et al., 2009; Bailey and Elkan, 1994*). The sequence ± 10 bp around each identified, reproducible pause (as well as the matched scrambled control) was extracted and used to run the MEME tool using parameters to find 0–1 motif per sequence, motifs 6–21 bp in length, and up to 10 motifs with an *E*-value significance threshold of 0.05 (*Bailey and Elkan, 1994*). These significant motifs were compared to known transcription factor binding site motifs in the YEASTRACT_20130918 database (*Teixeira et al., 2014*) using the TOMTOM tool (*Gupta et al., 2007*) using default parameters, calling all hits as significant with an *E*-value greater than 0.1. TOMTOM searches were only performed on those motifs with a relative entropy greater than five and only the top match is reported.

Random forest classifier for Pol II pausing loci

The predictive value of chromatin and DNA features for identifying Pol II pause loci was determined using a random forest model with the randomForest R package (*Breiman, 2001*). All reproducible Pol II pause loci were included in these analyses, as were an equal number of shuffled control loci. The shuffled control loci were selected to maintain the same number of real and control loci in each gene, controlling for effects of differential gene expression. In total, 51 chromatin and DNA features were compiled for all pause loci (*Supplementary file 5; Chiu et al., 2016; Oberbeckmann et al., 2019; Pelechano et al., 2013; Turner and Mathews, 2010; Umeyama and Ito, 2018; Vinayachandran et al., 2018; Weiner et al., 2015*). Before applying the random forest classifier, we examined the distribution of values for each numeric feature (not discrete sequence) for real Pol II pauses compared to the scrambled control loci; statistical significance in the difference between these distributions was calculated with a Student's *t*-test, correcting for multiple hypothesis testing with the Bonferroni correction. From the random forest classifier, feature importance scores were generated using a random forest classifier with 75% training and 25% testing sets; for wild-type yeast, this is 10,495 training and 3499 training loci. Due to the low number of reproducible pauses identified in the *gcn5Δ* deletion strain, it was excluded from these analyses.

Reported feature importance values are the mean decreases of accuracy over all out-of-bag cross-validated predictions, when a given feature is permuted after training, but before prediction. Optimized parameters were selected for random forest classifiers trained using all features (*Figure 5—figure supplement 1B*): *n*cat = 4, *m*try = 20, *n*trees = 2500. ROC curve and AUC measurements were determined from binary prediction probabilities and calculated using the ROCR R package (*Sing et al., 2005*). Prediction accuracy was determined by measuring the difference between the model's predictions on a held-out test set and measured variables. The baseline score was determined using a 'null' parameter that has the same value for every training and testing pair; thus, baseline represents the prediction accuracy with no additional information added to the model. To assess the transferability of a random forest classifier trained on Pol II pause loci in one strain, a model was trained on

100% of real and shuffled control Pol II loci from one deletion strain and then tested on all those pause loci in a second deletion strain, which was not included in the training set.

Code availability

All scripts and data analyses are available at https://github.com/churchmanlab/Yeast_NETseq_Screen; **Couvillion and Churchman Lab, 2022**. All plots were created in R using ggplot2 (**R Development Core Team, 2013**; **Wickham, 2016**).

Acknowledgements

We thank S Issac and C Patil for constructive feedback on the manuscript. This work was supported by National Institutes of Health grant R01-HG007173 (LSC) and a Ruth L Kirschstein National Research Service Award F31 HG010570 (KCL).

Additional information

Competing interests

L Stirling Churchman: Reviewing editor, *eLife*. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
National Institutes of Health	R01-HG007173	L Stirling Churchman
National Institutes of Health	F31 HG010570	Kate C Lachance

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Mary Couvillion, Formal analysis, Software, Validation, Visualization, Writing - review and editing; Kevin M Harlen, Data curation, Formal analysis, Investigation, Methodology, Supervision; Kate C Lachance, Formal analysis, Software, Visualization, Writing - original draft; Kristine L Trotta, Erin Smith, Christian Brion, Brendan M Smalec, Data curation; L Stirling Churchman, Conceptualization, Supervision, Writing - review and editing

Author ORCIDs

Kristine L Trotta  <http://orcid.org/0000-0002-8166-7696>
L Stirling Churchman  <http://orcid.org/0000-0003-3888-2574>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.78944.sa1>

Author response <https://doi.org/10.7554/eLife.78944.sa2>

Additional files

Supplementary files

- Supplementary file 1. Pairwise correlation between all replicates included in reverse genetic screen.
- Supplementary file 2. Differential transcription of each gene across all deletion strains. Lists every gene differentially transcribed, both sense and antisense strands, as determined using DESeq2 (**Love et al., 2014**), for every deletion strain included in screen. Significance was determined to be those genes with an adjusted $-value \leq 0.05$ and an absolute $\log_2(\text{fold change})$ in expression compared to wild-type ≥ 1 . For each significantly differentially transcribed gene, the $\log_2(\text{fold})$

change) and adjusted p-value is reported.

- Supplementary file 3. Differentially transcribed genes are enriched for GO terms. This table lists all GO terms that were significantly enriched in at least one deletion strain. For each GO term, if it was found to be significant in a given deletion strain, the fold enrichment and adjusted p-value (in parentheses) are listed. This table is separated into three sheets: those GO terms derived from either significantly up- or downregulated genes (purple), only significantly downregulated genes (red), and only significantly upregulated genes (blue).
- Supplementary file 4. Significant motifs underlying pauses across deletion strains with transcription factor binding site matches.
- Supplementary file 5. Sources of chromatin features used in random forest classifier.
- Supplementary file 6. Results of t-test between distributions of feature values comparing real and shuffled control pauses. For each numeric chromatin feature, the t-value, p-value, and indication of significance is given resulting from a Student's t-test comparing the distribution of values surrounding real and shuffled control pauses. Table corresponds to boxplots illustrating distributions for all numeric chromatin features (**Figure 5—figure supplement 1**). Significance indicators are applied after a Bonferroni correction for multiple hypotheses (* <0.05 , ** <0.01 , *** <0.001).
- MDAR checklist

Data availability

The accession number for the Illumina sequencing reported in this paper is Gene Expression Omnibus (GEO): GSE159603.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Couvillion MT, Lachance KC, Harlen KM, Trotta KL, Smith E, Churchman LS	2021	Dynamics of transcription elongation are finely-tuned by dozens of regulatory factors	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159603	NCBI Gene Expression Omnibus, GSE159603

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Hewawasam GS, Dhatchinamoorthy K, Mattingly M, Seidel C	2017	Chromatin assembly factor-1 (CAF-1) chaperone regulates Cse4 deposition at active promoter regions in budding yeast	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98397	NCBI Gene Expression Omnibus, GSE98397

References

- Adelman K, La Porta A, Santangelo TJ, Lis JT, Roberts JW, Wang MD. 2002. Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *PNAS* **99**:13538–13543. DOI: <https://doi.org/10.1073/pnas.212358999>, PMID: 12370445
- Alexander RD, Innocente SA, Barrass JD, Beggs JD. 2010. Splicing-dependent RNA polymerase pausing in yeast. *Molecular Cell* **40**:582–593. DOI: <https://doi.org/10.1016/j.molcel.2010.11.005>, PMID: 21095588
- Amerik AY, Li SJ, Hochstrasser M. 2000. Analysis of the deubiquitinating enzymes of the yeast *Saccharomyces cerevisiae*. *Biological Chemistry* **381**:981–992. DOI: <https://doi.org/10.1515/BC.2000.121>, PMID: 11076031
- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* **11**:R106. DOI: <https://doi.org/10.1186/gb-2010-11-10-r106>, PMID: 20979621
- Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)* **31**:166–169. DOI: <https://doi.org/10.1093/bioinformatics/btu638>, PMID: 25260700
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290**:457–465. DOI: <https://doi.org/10.1038/290457a0>, PMID: 7219534
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**:25–29. DOI: <https://doi.org/10.1038/75556>

- Bagchi DN**, Battenhouse AM, Park D, Iyer VR. 2020. The histone variant H2A.Z in yeast is almost exclusively incorporated into the +1 nucleosome in the direction of transcription. *Nucleic Acids Research* **48**:157–170. DOI: <https://doi.org/10.1093/nar/gkz1075>, PMID: 31722407
- Bailey TL**, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Proceedings. International Conference on Intelligent Systems for Molecular Biology. .
- Bailey TL**, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**:W202–W208. DOI: <https://doi.org/10.1093/nar/gkp335>, PMID: 19458158
- Bauer S**, Grossmann S, Vingron M, Robinson PN. 2008. Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics (Oxford, England)* **24**:1650–1651. DOI: <https://doi.org/10.1093/bioinformatics/btn250>, PMID: 18511468
- Bentley DL**, Groudine M. 1986. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. *Nature* **321**:702–706. DOI: <https://doi.org/10.1038/321702a0>, PMID: 3520340
- Bentley DL**. 2014. Coupling mRNA processing with transcription in time and space. *Nature Reviews. Genetics* **15**:163–175. DOI: <https://doi.org/10.1038/nrg3662>, PMID: 24514444
- Bertone P**, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science (New York, N.Y.)* **306**:2242–2246. DOI: <https://doi.org/10.1126/science.1103388>, PMID: 15539566
- Bibb MJ**, Van Etten RA, Wright CT, Walberg MW, Clayton DA. 1981. Sequence and gene organization of mouse mitochondrial DNA. *Cell* **26**:167–180. DOI: [https://doi.org/10.1016/0092-8674\(81\)90300-7](https://doi.org/10.1016/0092-8674(81)90300-7), PMID: 7332926
- Bintu L**, Ishibashi T, Dangkulwanich M, Wu YY, Lubkowska L, Kashlev M, Bustamante C. 2012. Nucleosomal elements that control the topography of the barrier to transcription. *Cell* **151**:738–749. DOI: <https://doi.org/10.1016/j.cell.2012.10.009>, PMID: 23141536
- Bonnet J**, Wang CY, Baptista T, Vincent SD, Hsiao WC, Stierle M, Kao CF, Tora L, Devys D. 2014. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes & Development* **28**:1999–2012. DOI: <https://doi.org/10.1101/gad.250225.114>, PMID: 25228644
- Breiman L**, Random Forests. 2001. *Machine Learning* **45**:5–32. DOI: <https://doi.org/10.1023/A:1010933404324>
- Camblong J**, Iglesias N, Fickentscher C, Dieppois G, Stutz F. 2007. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* **131**:706–717. DOI: <https://doi.org/10.1016/j.cell.2007.09.014>, PMID: 18022365
- Carey M**, Li B, Workman JL. 2006. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Molecular Cell* **24**:481–487. DOI: <https://doi.org/10.1016/j.molcel.2006.09.012>, PMID: 17081996
- Carrozza MJ**, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP, Workman JL. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**:581–592. DOI: <https://doi.org/10.1016/j.cell.2005.10.023>, PMID: 16286007
- Chen Z**, Gabizon R, Brown AI, Lee A, Song A, Díaz-Celis C, Kaplan CD, Koslover EF, Yao T, Bustamante C. 2019. High-resolution and high-accuracy topographic and transcriptional maps of the nucleosome barrier. *eLife* **8**:e48281. DOI: <https://doi.org/10.7554/eLife.48281>, PMID: 31364986
- Cheng J**, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science (New York, N.Y.)* **308**:1149–1154. DOI: <https://doi.org/10.1126/science.1108625>, PMID: 15790807
- Chipman H**, Tibshirani R. 2006. Hybrid hierarchical clustering with applications to microarray data. *Biostatistics (Oxford, England)* **7**:286–301. DOI: <https://doi.org/10.1093/biostatistics/kxj007>, PMID: 16301308
- Chiu TP**, Comoglio F, Zhou T, Yang L, Paro R, Rohs R. 2016. DNashapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics (Oxford, England)* **32**:1211–1213. DOI: <https://doi.org/10.1093/bioinformatics/btv735>, PMID: 26668005
- Churchman LS**, Weissman JS. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**:368–373. DOI: <https://doi.org/10.1038/nature09652>, PMID: 21248844
- Churchman LS**, Weissman JS. 2012. Native elongating transcript sequencing (NET-seq). *Current Protocols in Molecular Biology Chapter 4*:s98. DOI: <https://doi.org/10.1002/0471142727.mb0414s98>, PMID: 22470065
- Core LJ**, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)* **322**:1845–1848. DOI: <https://doi.org/10.1126/science.1162228>, PMID: 19056941
- Couvillion M**, Churchman Lab. 2022. Yeast_NETseq_Screen. 33d4041. GitHub. https://github.com/churchmanlab/Yeast_NETseq_Screen
- Dahlberg JE**, Blattner FR. 1973. In vitro transcription products of lambda DNA: nucleotide sequences and regulatory sites. *Virus Res.*
- David L**, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM. 2006. A high-resolution map of transcription in the yeast genome. *PNAS* **103**:5320–5325. DOI: <https://doi.org/10.1073/pnas.0601091103>, PMID: 16569694
- Drexler HL**, Choquet K, Churchman LS. 2020. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Molecular Cell* **77**:985–998. DOI: <https://doi.org/10.1016/j.molcel.2019.11.017>, PMID: 31839405

- Eick D**, Bornkamm GW. 1986. Transcriptional arrest within the first exon is a fast control mechanism in c-myc gene expression. *Nucleic Acids Research* **14**:8331–8346. DOI: <https://doi.org/10.1093/nar/14.21.8331>, PMID: [3537956](https://pubmed.ncbi.nlm.nih.gov/3537956/)
- Fennessy RT**, Owen-Hughes T. 2016. Establishment of a promoter-based chromatin architecture on recently replicated DNA can accommodate variable inter-nucleosome spacing. *Nucleic Acids Research* **44**:7189–7203. DOI: <https://doi.org/10.1093/nar/gkw331>, PMID: [27106059](https://pubmed.ncbi.nlm.nih.gov/27106059/)
- Ferrari F**, Plachetka A, Alekseyenko AA, Jung YL, Oszolak F, Kharchenko PV, Park PJ, Kuroda MI. 2013. “Jump start and gain” model for dosage compensation in *Drosophila* based on direct sequencing of nascent transcripts. *Cell Reports* **5**:629–636. DOI: <https://doi.org/10.1016/j.celrep.2013.09.037>, PMID: [24183666](https://pubmed.ncbi.nlm.nih.gov/24183666/)
- Formosa T**, Ruone S, Adams MD, Olsen AE, Eriksson P, Yu Y, Rhoades AR, Kaufman PD, Stillman DJ. 2002. Defects in SPT16 or POB3 (yFACT) in *Saccharomyces cerevisiae* cause dependence on the Hir/Hpc pathway: polymerase passage may degrade chromatin structure. *Genetics* **162**:1557–1571. DOI: <https://doi.org/10.1093/genetics/162.4.1557>, PMID: [12524332](https://pubmed.ncbi.nlm.nih.gov/12524332/)
- Fu Y**, Wu PH, Beane T, Zamore PD, Weng Z. 2018. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics* **19**:531. DOI: <https://doi.org/10.1186/s12864-018-4933-1>, PMID: [30001700](https://pubmed.ncbi.nlm.nih.gov/30001700/)
- Funakoshi Y**, Doi Y, Hosoda N, Uchida N, Osawa M, Shimada I, Tsujimoto M, Suzuki T, Katada T, Hoshino S. 2007. Mechanism of mRNA deadenylation: evidence for a molecular interplay between translation termination factor eRF3 and mRNA deadenylases. *Genes & Development* **21**:3135–3148. DOI: <https://doi.org/10.1101/gad.1597707>, PMID: [18056425](https://pubmed.ncbi.nlm.nih.gov/18056425/)
- Gajos M**, Jasnovidova O, van Bömmel A, Freier S, Vingron M, Mayer A. 2021. Conserved DNA sequence features underlie pervasive RNA polymerase pausing. *Nucleic Acids Research* **49**:4402–4420. DOI: <https://doi.org/10.1093/nar/gkab208>, PMID: [33788942](https://pubmed.ncbi.nlm.nih.gov/33788942/)
- Galburt EA**, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, Kashlev M, Bustamante C. 2007. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**:820–823. DOI: <https://doi.org/10.1038/nature05701>, PMID: [17361130](https://pubmed.ncbi.nlm.nih.gov/17361130/)
- Gariglio P**, Bellard M, Chambon P. 1981. Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Research* **9**:2589–2598. DOI: <https://doi.org/10.1093/nar/9.11.2589>, PMID: [6269056](https://pubmed.ncbi.nlm.nih.gov/6269056/)
- Gilbert W**, Maizels N, Maxam A. 1974. Sequences of controlling regions of the lactose operon. *Cold Spring Harbor Symposia on Quantitative Biology* **38**:845–855. DOI: <https://doi.org/10.1101/sq.1974.038.01.087>, PMID: [4598642](https://pubmed.ncbi.nlm.nih.gov/4598642/)
- Gilmour DS**, Lis JT. 1986. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Molecular and Cellular Biology* **6**:3984–3989. DOI: <https://doi.org/10.1128/mcb.6.11.3984-3989.1986>, PMID: [3099167](https://pubmed.ncbi.nlm.nih.gov/3099167/)
- Grossmann S**, Bauer S, Robinson PN, Vingron M. 2007. Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. *Bioinformatics (Oxford, England)* **23**:3024–3031. DOI: <https://doi.org/10.1093/bioinformatics/btm440>, PMID: [17848398](https://pubmed.ncbi.nlm.nih.gov/17848398/)
- Gupta S**, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome Biology* **8**:R24. DOI: <https://doi.org/10.1186/gb-2007-8-2-r24>, PMID: [17324271](https://pubmed.ncbi.nlm.nih.gov/17324271/)
- Hangauer MJ**, Vaughn IW, McManus MT. 2013. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLOS Genetics* **9**:e1003569. DOI: <https://doi.org/10.1371/journal.pgen.1003569>, PMID: [23818866](https://pubmed.ncbi.nlm.nih.gov/23818866/)
- Harlen KM**, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Reports* **15**:2147–2158. DOI: <https://doi.org/10.1016/j.celrep.2016.05.010>, PMID: [27239037](https://pubmed.ncbi.nlm.nih.gov/27239037/)
- Herbert KM**, La Porta A, Wong BJ, Mooney RA, Neuman KC, Landick R, Block SM. 2006. Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell* **125**:1083–1094. DOI: <https://doi.org/10.1016/j.cell.2006.04.032>, PMID: [16777599](https://pubmed.ncbi.nlm.nih.gov/16777599/)
- Herbert KM**, Greenleaf WJ, Block SM. 2008. Single-molecule studies of RNA polymerase: motoring along. *Annual Review of Biochemistry* **77**:149–176. DOI: <https://doi.org/10.1146/annurev.biochem.77.073106.100741>, PMID: [18410247](https://pubmed.ncbi.nlm.nih.gov/18410247/)
- Herzel L**, Ottoz DSM, Alpert T, Neugebauer KM. 2017. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nature Reviews. Molecular Cell Biology* **18**:637–650. DOI: <https://doi.org/10.1038/nrm.2017.63>, PMID: [28792005](https://pubmed.ncbi.nlm.nih.gov/28792005/)
- Hewawasam GS**, Dhatchinamoorthy K, Mattingly M, Seidel C, Gerton JL. 2018. Chromatin assembly factor-1 (CAF-1) chaperone regulates Cse4 deposition into chromatin in budding yeast. *Nucleic Acids Research* **46**:4831. DOI: <https://doi.org/10.1093/nar/gky405>, PMID: [29733388](https://pubmed.ncbi.nlm.nih.gov/29733388/)
- Hodges C**, Bintu L, Lubkowska L, Kashlev M, Bustamante C. 2009. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science (New York, N.Y.)* **325**:626–628. DOI: <https://doi.org/10.1126/science.1172926>, PMID: [19644123](https://pubmed.ncbi.nlm.nih.gov/19644123/)
- Holmes RK**, Tuck AC, Zhu C, Dunn-Davies HR, Kudla G, Clauder-Munster S, Granneman S, Steinmetz LM, Guthrie C, Tollervey D. 2015. Loss of the Yeast SR Protein Npl3 Alters Gene Expression Due to Transcription Readthrough. *PLOS Genetics* **11**:e1005735. DOI: <https://doi.org/10.1371/journal.pgen.1005735>, PMID: [26694144](https://pubmed.ncbi.nlm.nih.gov/26694144/)
- Hongay CF**, Grisafi PL, Galitski T, Fink GR. 2006. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127**:735–745. DOI: <https://doi.org/10.1016/j.cell.2006.09.038>, PMID: [17110333](https://pubmed.ncbi.nlm.nih.gov/17110333/)

- Houseley J**, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. 2008. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Molecular Cell* **32**:685–695. DOI: <https://doi.org/10.1016/j.molcel.2008.09.027>, PMID: 19061643
- Jentsch S**, McGrath JP, Varshavsky A. 1987. The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature* **329**:131–134. DOI: <https://doi.org/10.1038/329131a0>, PMID: 3306404
- Jin Y**, Eser U, Struhl K, Churchman LS. 2017. The Ground State and Evolution of Promoter Region Directionality. *Cell* **170**:889–898. DOI: <https://doi.org/10.1016/j.cell.2017.07.006>, PMID: 28803729
- Kapranov P**, Cheng J, Dike S, Nix DA, Dutttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, et al. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, N.Y.)* **316**:1484–1488. DOI: <https://doi.org/10.1126/science.1138341>, PMID: 17510325
- Kassavetis GA**, Chamberlin MJ. 1981. Pausing and termination of transcription within the early region of bacteriophage T7 DNA in vitro. *The Journal of Biological Chemistry* **256**:2777–2786 PMID: 7009597.,
- Kaufman PD**, Kobayashi R, Stillman B. 1997. Ultraviolet radiation sensitivity and reduction of telomeric silencing in *Saccharomyces cerevisiae* cells lacking chromatin assembly factor-I. *Genes & Development* **11**:345–357. DOI: <https://doi.org/10.1101/gad.11.3.345>, PMID: 9030687
- Kim J**, Roeder RG. 2009. Direct Bre1-Paf1 complex interactions and RING finger-independent Bre1-Rad6 interactions mediate histone H2B ubiquitylation in yeast. *The Journal of Biological Chemistry* **284**:20582–20592. DOI: <https://doi.org/10.1074/jbc.M109.017442>, PMID: 19531475
- Kim D**, Perteu G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**:R36. DOI: <https://doi.org/10.1186/gb-2013-14-4-r36>, PMID: 23618408
- Kindgren P**, Ivanov M, Marquardt S. 2020. Native elongation transcript sequencing reveals temperature dependent dynamics of nascent RNAPII transcription in Arabidopsis. *Nucleic Acids Research* **48**:2332–2347. DOI: <https://doi.org/10.1093/nar/gkz1189>, PMID: 31863587
- Kingston RE**, Chamberlin MJ. 1981. Pausing and attenuation of in vitro transcription in the *rrnB* operon of *E. coli*. *Cell* **27**:523–531. DOI: [https://doi.org/10.1016/0092-8674\(81\)90394-9](https://doi.org/10.1016/0092-8674(81)90394-9), PMID: 6086107
- Kireeva ML**, Hancock B, Cremona GH, Walter W, Studitsky VM, Kashlev M. 2005. Nature of the nucleosomal barrier to RNA polymerase II. *Molecular Cell* **18**:97–108. DOI: <https://doi.org/10.1016/j.molcel.2005.02.027>, PMID: 15808512
- Kireeva ML**, Kashlev M. 2009. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *PNAS* **106**:8900–8905. DOI: <https://doi.org/10.1073/pnas.0900407106>, PMID: 19416863
- Krogan NJ**, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C, Shilatifard A, Buratowski S, Greenblatt J. 2003. Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Molecular and Cellular Biology* **23**:4207–4218. DOI: <https://doi.org/10.1128/MCB.23.12.4207-4218.2003>, PMID: 12773564
- Krumm A**, Meulia T, Brunvand M, Groudine M. 1992. The block to transcriptional elongation within the human *c-myc* gene is determined in the promoter-proximal region. *Genes & Development* **6**:2201–2213. DOI: <https://doi.org/10.1101/gad.6.11.2201>, PMID: 1427080
- Kwak H**, Fuda NJ, Core LJ, Lis JT. 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science (New York, N.Y.)* **339**:950–953. DOI: <https://doi.org/10.1126/science.1229386>, PMID: 23430654
- Landt SG**, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research* **22**:1813–1831. DOI: <https://doi.org/10.1101/gr.136184.111>, PMID: 22955991
- Larson MH**, Mooney RA, Peters JM, Windgassen T, Nayak D, Gross CA, Block SM, Greenleaf WJ, Landick R, Weissman JS. 2014. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science (New York, N.Y.)* **344**:1042–1047. DOI: <https://doi.org/10.1126/science.1251871>, PMID: 24789973
- Lee F**, Squires CL, Squires C, Yanofsky C. 1976. Termination of transcription in vitro in the *Escherichia coli* tryptophan operon leader region. *Journal of Molecular Biology* **103**:383–393. DOI: [https://doi.org/10.1016/0022-2836\(76\)90318-1](https://doi.org/10.1016/0022-2836(76)90318-1), PMID: 781272
- Lenstra TL**, Coulon A, Chow CC, Larson DR. 2015. Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Molecular Cell* **60**:597–610. DOI: <https://doi.org/10.1016/j.molcel.2015.09.028>, PMID: 26549684
- Li B**, Jackson J, Simon MD, Fleharty B, Gogol M, Seidel C, Workman JL, Shilatifard A. 2009. Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *The Journal of Biological Chemistry* **284**:7970–7976. DOI: <https://doi.org/10.1074/jbc.M808220200>, PMID: 19155214
- Li Q**, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics* **5**:1752–1779. DOI: <https://doi.org/10.1214/11-AOAS466>
- Love MI**, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. DOI: <https://doi.org/10.1186/s13059-014-0550-8>, PMID: 25516281
- Maizels NM**. 1973. The nucleotide sequence of the lactose messenger ribonucleic acid transcribed from the UV5 promoter mutant of *Escherichia coli*. *PNAS* **70**:3585–3589. DOI: <https://doi.org/10.1073/pnas.70.12.3585>, PMID: 4587256

- Marquardt S**, Escalante-Chong R, Pho N, Wang J, Churchman LS, Springer M, Buratowski S. 2014. A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* **157**:1712–1723. DOI: <https://doi.org/10.1016/j.cell.2014.04.036>, PMID: 24949978
- Martens JA**, Laprade L, Winston F. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**:571–574. DOI: <https://doi.org/10.1038/nature02538>, PMID: 15175754
- Mayer A**, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, Sandstrom R, Stamatoyannopoulos JA, Churchman LS. 2015. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**:541–554. DOI: <https://doi.org/10.1016/j.cell.2015.03.010>, PMID: 25910208
- Mayer A**, Landry HM, Churchman LS. 2017. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Current Opinion in Cell Biology* **46**:72–80. DOI: <https://doi.org/10.1016/j.ceb.2017.03.002>, PMID: 28363125
- Mercer TR**, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddleloh JA, Mattick JS, Rinn JL. 2011. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature Biotechnology* **30**:99–104. DOI: <https://doi.org/10.1038/nbt.2024>, PMID: 22081020
- Mi H**, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* **47**:D419–D426. DOI: <https://doi.org/10.1093/nar/gky1038>, PMID: 30407594
- Mosley AL**, Hunter GO, Sardi ME, Smolle M, Workman JL, Florens L, Washburn MP. 2013. Quantitative proteomics demonstrates that the RNA polymerase II subunits Rpb4 and Rpb7 dissociate during transcriptional elongation. *Molecular & Cellular Proteomics* **12**:1530–1538. DOI: <https://doi.org/10.1074/mcp.M112.024034>, PMID: 23418395
- Murray SC**, Haenni S, Howe FS, Fischl H, Chocian K, Nair A, Mellor J. 2015. Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Research* **43**:7823–7837. DOI: <https://doi.org/10.1093/nar/gkv666>, PMID: 26130720
- Murray SC**, Mellor J. 2016. Using both strands: The fundamental nature of antisense transcription. *Bioarchitecture* **6**:12–21. DOI: <https://doi.org/10.1080/19490992.2015.1130779>, PMID: 26760777
- Nagalakshmi U**, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (New York, N.Y.)* **320**:1344–1349. DOI: <https://doi.org/10.1126/science.1158441>, PMID: 18451266
- Nechaev S**, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science (New York, N.Y.)* **327**:335–338. DOI: <https://doi.org/10.1126/science.1181421>, PMID: 20007866
- Nepveu A**, Marcu KB. 1986. Intragenic pausing and anti-sense transcription within the murine c-myc locus. *The EMBO Journal* **5**:2859–2865. DOI: <https://doi.org/10.1002/j.1460-2075.1986.tb04580.x>, PMID: 3024965
- Neuman KC**, Abbondanzieri EA, Landick R, Gelles J, Block SM. 2003. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**:437–447. DOI: [https://doi.org/10.1016/s0092-8674\(03\)00845-6](https://doi.org/10.1016/s0092-8674(03)00845-6), PMID: 14622598
- Nevers A**, Doyen A, Malabat C, Néron B, Kergrohen T, Jacquier A, Badis G. 2018. Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. *Nucleic Acids Research* **46**:6009–6025. DOI: <https://doi.org/10.1093/nar/gky342>, PMID: 29788449
- Noe Gonzalez M**, Blears D, Svejstrup JQ. 2021. Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nature Reviews. Molecular Cell Biology* **22**:3–21. DOI: <https://doi.org/10.1038/s41580-020-00308-8>, PMID: 33208928
- Nojima T**, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**:526–540. DOI: <https://doi.org/10.1016/j.cell.2015.03.027>, PMID: 25910207
- Nojima T**, Rebelo K, Gomes T, Grosso AR, Proudfoot NJ, Carmo-Fonseca M. 2018. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Molecular Cell* **72**:369–379. DOI: <https://doi.org/10.1016/j.molcel.2018.09.004>, PMID: 30340024
- Oberbeckmann E**, Wolff M, Krietenstein N, Heron M, Ellins JL, Schmid A, Krebs S, Blum H, Gerland U, Korber P. 2019. Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Research* **29**:1996–2009. DOI: <https://doi.org/10.1101/gr.253419.119>, PMID: 31694866
- Ocampo J**, Chereji RV, Eriksson PR, Clark DJ. 2016. The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing in vivo. *Nucleic Acids Research* **44**:4625–4635. DOI: <https://doi.org/10.1093/nar/gkw068>, PMID: 26861626
- Parekh S**, Ziegenhain C, Vieth B, Enard W, Hellmann I. 2016. The impact of amplification on differential expression analyses by RNA-seq. *Scientific Reports* **6**:25533. DOI: <https://doi.org/10.1038/srep25533>, PMID: 27156886
- Pelechano V**, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**:127–131. DOI: <https://doi.org/10.1038/nature12121>, PMID: 23615609
- Prochasson P**, Florens L, Swanson SK, Washburn MP, Workman JL. 2005. The HIR corepressor complex binds to nucleosomes generating a distinct protein/DNA complex resistant to remodeling by SWI/SNF. *Genes & Development* **19**:2534–2539. DOI: <https://doi.org/10.1101/gad.1341105>, PMID: 16264190
- Proudfoot NJ**, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. *Cell* **108**:501–512. DOI: [https://doi.org/10.1016/s0092-8674\(02\)00617-7](https://doi.org/10.1016/s0092-8674(02)00617-7), PMID: 11909521
- R Development Core Team**. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria. R Foundation for Statistical Computing. <http://www.r-project.org>

- Raisch T**, Chang CT, Levdansky Y, Muthukumar S, Raunser S, Valkov E. 2019. Reconstitution of recombinant human CCR4-NOT reveals molecular insights into regulated deadenylation. *Nature Communications* **10**:3173. DOI: <https://doi.org/10.1038/s41467-019-11094-z>, PMID: 31320642
- Rando OJ**, Winston F. 2012. Chromatin and transcription in yeast. *Genetics* **190**:351–387. DOI: <https://doi.org/10.1534/genetics.111.132266>, PMID: 22345607
- Robinson JT**, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**:24–26. DOI: <https://doi.org/10.1038/nbt.1754>, PMID: 21221095
- Rougvie AE**, Lis JT. 1988. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**:795–804. DOI: [https://doi.org/10.1016/S0092-8674\(88\)91087-2](https://doi.org/10.1016/S0092-8674(88)91087-2), PMID: 3136931
- Runner VM**, Podolny V, Buratowski S. 2008. The Rpb4 Subunit of RNA Polymerase II Contributes to Cotranscriptional Recruitment of 3' Processing Factors. *Molecular and Cellular Biology* **28**:1883–1891. DOI: <https://doi.org/10.1128/MCB.01714-07>, PMID: 18195044
- Schwalb B**, Michel M, Zacher B, Frühauf K, Demel C, Tresch A, Gagneur J, Cramer P. 2016. TT-seq maps the human transient transcriptome. *Science* **352**:1225–1228. DOI: <https://doi.org/10.1126/science.aad9841>, PMID: 27257258
- Shaevitz JW**, Abbondanzieri EA, Landick R, Block SM. 2003. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426**:684–687. DOI: <https://doi.org/10.1038/nature02191>, PMID: 14634670
- Shandilya J**, Roberts SGE. 2012. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochimica et Biophysica Acta* **1819**:391–400. DOI: <https://doi.org/10.1016/j.bbaggm.2012.01.010>, PMID: 22306664
- Shukla S**, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**:74–79. DOI: <https://doi.org/10.1038/nature10442>, PMID: 21964334
- Sing T**, Sander O, Beerenwinkel N, Lengauer T. 2005. ROCr: visualizing classifier performance in R. *Bioinformatics* **21**:3940–3941. DOI: <https://doi.org/10.1093/bioinformatics/bti623>, PMID: 16096348
- Singh AK**, Mueller-Planitz F. 2021. Nucleosome Positioning and Spacing: From Mechanism to Function. *Journal of Molecular Biology* **433**:166847. DOI: <https://doi.org/10.1016/j.jmb.2021.166847>, PMID: 33539878
- Smolle M**, Venkatesh S, Gogol MM, Li H, Zhang Y, Florens L, Washburn MP, Workman JL. 2012. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nature Structural & Molecular Biology* **19**:884–892. DOI: <https://doi.org/10.1038/nsmb.2312>, PMID: 22922743
- Smolle M**, Workman JL. 2013. Transcription-associated histone modifications and cryptic transcription. *Biochimica et Biophysica Acta* **1829**:84–97. DOI: <https://doi.org/10.1016/j.bbaggm.2012.08.008>, PMID: 22982198
- Spencer CA**, Groudine M. 1990. Transcription elongation and eukaryotic gene regulation. *Oncogene* **5**:777–785. PMID: 2193290.,
- Spiegelman WG**, Reichardt LF, Yaniv M, Heinemann SF, Kaiser AD, Eisen H. 1972. Bidirectional transcription and the regulation of Phage lambda repressor synthesis. *PNAS* **69**:3156–3160. DOI: <https://doi.org/10.1073/pnas.69.11.3156>, PMID: 4508309
- Steinmetz EJ**, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA. 2006. Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Molecular Cell* **24**:735–746. DOI: <https://doi.org/10.1016/j.molcel.2006.10.023>, PMID: 17157256
- Strobl LJ**, Eick D. 1992. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. *The EMBO Journal* **11**:3307–3314. DOI: <https://doi.org/10.1002/j.1460-2075.1992.tb05409.x>, PMID: 1505520
- Struhl K**. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nature Structural & Molecular Biology* **14**:103–105. DOI: <https://doi.org/10.1038/nsmb0207-103>, PMID: 17277804
- Svejstrup JQ**. 2004. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochimica et Biophysica Acta* **1677**:64–73. DOI: <https://doi.org/10.1016/j.bbaexp.2003.10.012>, PMID: 15020047
- Teixeira MC**, Monteiro PT, Guerreiro JF, Gonçalves JP, Mira NP, dos Santos SC, Cabrito TR, Palma M, Costa C, Francisco AP, Madeira SC, Oliveira AL, Freitas AT, Sá-Correia I. 2014. The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **42**:D161–D166. DOI: <https://doi.org/10.1093/nar/gkt1015>, PMID: 24170807
- Temme C**, Simonelig M, Wahle E. 2014. Deadenylation of mRNA by the CCR4-NOT complex in *Drosophila*: molecular and developmental aspects. *Frontiers in Genetics* **5**:143. DOI: <https://doi.org/10.3389/fgene.2014.00143>, PMID: 24904643
- The Gene Ontology Consortium**. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* **47**:D330–D338. DOI: <https://doi.org/10.1093/nar/gky1055>, PMID: 30395331
- Tucker M**, Staples RR, Valencia-Sanchez MA, Muhlrad D, Parker R. 2002. Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. *The EMBO Journal* **21**:1427–1436. DOI: <https://doi.org/10.1093/emboj/21.6.1427>, PMID: 11889048
- Turner DH**, Mathews DH. 2010. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research* **38**:D280–D282. DOI: <https://doi.org/10.1093/nar/gkp892>, PMID: 19880381
- Uhler JP**, Hertel C, Svejstrup JQ. 2007. A role for noncoding transcription in activation of the yeast PHO5 gene. *PNAS* **104**:8011–8016. DOI: <https://doi.org/10.1073/pnas.0702431104>, PMID: 17470801

- Umeyama T, Ito T.** 2018. DMS-seq for In Vivo Genome-Wide Mapping of Protein-DNA Interactions and Nucleosome Centers. *Current Protocols in Molecular Biology* **123**:e60. DOI: <https://doi.org/10.1002/cpmb.60>, PMID: 29927065
- Ussery DW.** 2002. DNA Structure: A-, B- and Z-DNA Helix Families. *Encyclopedia of Life Sciences* **1**:e003122. DOI: <https://doi.org/10.1038/npg.els.0003122>
- Vinayachandran V, Reja R, Rossi MJ, Park B, Rieber L, Mittal C, Mahony S, Pugh BF.** 2018. Widespread and precise reprogramming of yeast protein-genome interactions in response to heat shock. *Genome Research* **1**:117. DOI: <https://doi.org/10.1101/gr.226761.117>, PMID: 29444801
- Vvedenskaya IO, Vahedian-Movahed H, Bird JG, Knoblauch JG, Goldman SR, Zhang Y, Ebright RH, Nickels BE.** 2014. Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science (New York, N.Y.)* **344**:1285–1289. DOI: <https://doi.org/10.1126/science.1253458>, PMID: 24926020
- Wahle E, Winkler GS.** 2013. RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. *Biochimica et Biophysica Acta* **1829**:561–570. DOI: <https://doi.org/10.1016/j.bbagr.2013.01.003>, PMID: 23337855
- Wang MD, Schnitzer MJ, Yin H, Landick R, Gelles J, Block SM.** 1998. Force and velocity measured for single molecules of RNA polymerase. *Science (New York, N.Y.)* **282**:902–907. DOI: <https://doi.org/10.1126/science.282.5390.902>, PMID: 9794753
- Weber CM, Ramachandran S, Henikoff S.** 2014. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Molecular Cell* **53**:819–830. DOI: <https://doi.org/10.1016/j.molcel.2014.02.014>, PMID: 24606920
- Weiner A, Hsieh THS, Appleboim A, Chen HV, Rahat A, Amit I, Rando OJ, Friedman N.** 2015. High-resolution chromatin dynamics during a yeast stress response. *Molecular Cell* **58**:371–386. DOI: <https://doi.org/10.1016/j.molcel.2015.02.002>, PMID: 25801168
- Wickham H.** 2016. Ggplot2: Elegant Graphics for Data Analysis. Cham: Springer. DOI: <https://doi.org/10.1007/978-3-319-24277-4>
- Xu H, Kim UJ, Schuster T, Grunstein M.** 1992. Identification of a new set of cell cycle-regulatory genes that regulate S-phase transcription of histone genes in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **12**:5249–5259. DOI: <https://doi.org/10.1128/mcb.12.11.5249-5259.1992>, PMID: 1406694
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM.** 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**:1033–1037. DOI: <https://doi.org/10.1038/nature07728>, PMID: 19169243
- Xu Z, Wei W, Gagneur J, Clauder-Münster S, Smolik M, Huber W, Steinmetz LM.** 2011. Antisense expression increases gene expression variability and locus interdependency. *Molecular Systems Biology* **7**:468. DOI: <https://doi.org/10.1038/msb.2011.1>, PMID: 21326235
- Yamashita A, Chang TC, Yamashita Y, Zhu W, Zhong Z, Chen CYA, Shyu AB.** 2005. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nature Structural & Molecular Biology* **12**:1054–1063. DOI: <https://doi.org/10.1038/nsmb1016>, PMID: 16284618
- Yi H, Park J, Ha M, Lim J, Chang H, Kim VN.** 2018. PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Molecular Cell* **70**:1081–1088. DOI: <https://doi.org/10.1016/j.molcel.2018.05.009>, PMID: 29932901
- Zhang H, Roberts DN, Cairns BR.** 2005. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**:219–231. DOI: <https://doi.org/10.1016/j.cell.2005.08.036>, PMID: 16239141