OXFORD GENETICS

GENETICS, 2026, 233(1), iyag057
https://doi.org/10.1093/genetics/iyag057
Advance Access Publication Date: 27 February 2026
Investigation

Genetic analysis of polyadenylation patterns reveals distinct classes of yeast genes and local chromatin effects on Pol II elongation

Zarmik Moqtaderi,† Joseph V. Geisberg,† Kevin Struhl *

Department Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Harvard University, Boston, MA 02115, United States

*Corresponding author: Email: kevin@hms.harvard.edu
†Equal author contribution

The poly(A) profile of a typical yeast gene comprises ~50 poly(A) sites corresponding to distinct 3′ mRNA isoforms. Poly(A) profiles are shifted upstream in strains with slow RNA polymerase II (Pol II) elongation rates resulting from Pol II mutations or from depletion of histone chaperones FACT or Spt6. Conversely, downstream-shifted poly(A) profiles occur in strains with fast Pol II elongation rates caused by Pol II mutations or by depletion of histones. Downstream-shifted poly(A) profiles also arise when components of the cleavage/polyadenylation machinery are depleted. Here we examine poly(A) profiles in 45 yeast strains with mutations that affect a wide variety of DNA- or RNA-based processes. Typically, altered poly(A) profiles occur in only a subset of genes, and some strains show both upstream and downstream poly(A) shifts depending on the gene. Hierarchical clustering of the poly(A) profiles reveals 6 classes of genes that behave similarly over the genetic conditions tested and have different RNA sequence preferences in their 3′ untranslated regions. Many chromatin-modifying or chromatin-associated proteins affect poly(A) profiles similarly to Pol II elongation rate mutants, indicating that they affect the rate of transcriptional elongation. Chromatin-modifying activities, but not cleavage/polyadenylation factors, often have non-uniform effects on 3′ mRNA isoform levels of the same gene, suggesting that chromatin differentially affects Pol II dwell time at individual sites during transcriptional elongation.

Keywords: polyadenylation; RNA polymerase II; transcriptional elongation; chromatin; poly(A) site choice

Introduction

Eukaryotic mRNA 3′ ends are generated during transcriptional elongation in a process that involves first cleavage and then polyadenylation of nascent transcripts after RNA polymerase II (Pol II) traverses the protein-coding sequence (Proudfoot et al. 2002; Tian and Manley 2013, 2017; Kumar et al. 2019). Although there are many possible cleavage/polyadenylation sites per gene, only one is used per nascent mRNA molecule. This is because transcript cleavage triggers transcriptional termination such that the availability of sites further downstream becomes moot. A gene's polyadenylation profile is thus a compendium representing a frequency distribution of polyadenylation sites used by the many individual mRNA molecules arising from that gene.

In yeast, a typical polyadenylation profile encompasses ~50 mRNA 3′ isoforms differing in the lengths of their 3′ untranslated regions (3′ UTRs) (Ozsolak et al. 2010; Moqtaderi et al. 2013; Pelechano et al. 2013). These isoforms—some highly preferred, others much less so—result from cleavage/polyadenylation at many different sites, with most occurring over a ~200 nt region termed the "end zone." Analysis of chimeric genes in which the protein-coding sequence of one gene is juxtaposed with the 3′ UTR of another demonstrates that the polyadenylation profile is determined by the 3′ UTR (Lui et al. 2022). In fact, deleting the entire coding sequence, although it usually results in lower steady-state mRNA levels, generally has little effect on the polyadenylation pattern, with relative isoform frequencies remaining unchanged (Lui et al. 2022). The location of poly(A) sites with respect to the translational termination codon varies considerably among genes, but it is partially conserved between *S. cerevisiae* and *S. pombe* homologs, suggestive of biological function (Geisberg et al. 2024b).

Yeast polyadenylation patterns undergo widespread changes under certain environmental conditions or in the presence of mutations, with overall polyadenylation profiles shifting upstream or downstream (Geisberg et al. 2020, 2024a). These changes do not involve the emergence of novel poly(A) sites but instead reflect changes in site utilization to favor poly(A) sites farther upstream or downstream. This indicates that certain sequences are more suitable for cleavage/polyadenylation, and it also suggests the biological utility of having many eligible polyadenylation sites. When profile changes are present, shifts occur steadily from one isoform to the next, even when the polyadenylation sites are immediately adjacent (Geisberg et al. 2022). The continuous upstream- and downstream-shifted poly(A) profiles differ dramatically from the altered poly(A) profiles that occur when transcriptional initiation is blocked, and changed 3′ isoform levels reflect differences in their half-lives (Geisberg et al. 2014). The continuous nature of the poly(A) shifts in the Pol II speed mutants indicates a spatial association between the cleavage/polyadenylation and Pol II

Received on 09 December 2025; accepted on 21 February 2026

elongation complexes, with cleavage/polyadenylation occurring just as the nascent transcript exits the Pol II elongation complex (Geisberg et al. 2022).

Cells undergoing the metabolic changes of the diauxic shift exhibit a marked upstream shift in the polyadenylation pattern at most genes and have reduced Pol II processivity typical of a lower elongation rate (Geisberg et al. 2020). This is notable because Pol II mutants with slow elongation rates confer continuous upstream shifts in isoform profiles, remarkably like those seen in diauxic cultures. However, mutations that reduce Pol II processivity without decreasing the elongation rate (Mason and Struhl 2005) are insufficient to confer shifts in the isoform profile, indicating that slow elongation rates cause the upstream-shifted pattern (Geisberg et al. 2020). Conversely, Pol II mutations with fast elongation speeds lead to downstream shifts in isoform profiles, though the shifts are subtler than the upstream shifts in slow Pol II strains and affect fewer genes (Geisberg et al. 2020).

Chromatin, through its effects on the Pol II elongation rate, also affects poly(A) profiles. Cells depleted of either of the histone chaperones FACT or Spt6 have reduced Pol II processivity and continuous upstream-shifted polyadenylation profiles that strikingly resemble those conferred by slow Pol II derivatives (Geisberg et al. 2024a). Thus, these factors each stimulate the normal Pol II elongation rate, and their depletion leads to upstream shifts characteristic of slow Pol II mutants. In contradistinction, depletion of either histone H3 or histone H4 leads to a dramatic and continuous downstream shift in polyadenylation profiles of most genes, indicating that nucleosomes inhibit the rate of Pol II elongation *in vivo* (Geisberg et al. 2024a). The downstream shifts upon histone depletion are distinct from those occurring in strains defective for cleavage/polyadenylation, presumably reflecting independent mechanisms. Modulation of the Pol II elongation rate by chromatin can provide an additional regulatory layer to alternative polyadenylation that is distinct from direct effects on the cleavage/polyadenylation apparatus.

Although these upstream and downstream poly(A) shifts are observed for most genes, a small minority of genes show atypical poly(A) profile shifts that occur in the opposite direction (Geisberg et al. 2020). This observation suggests that sequence differences among 3′ UTRs affect shifts in poly(A) profiles, but direct evidence is lacking. In addition, it is largely unknown whether proteins not directly linked to chromatin or cleavage/polyadenylation can influence poly(A) profiles.

Here we examine poly(A) profiles in 45 yeast strains carrying mutations in proteins involved in chromatin, transcriptional elongation, mRNA splicing, cleavage/polyadenylation, mRNA stability, and RNA binding. In general, strains differ in the proportion of upstream- vs downstream-shifted genes, the number of genes shifted, and the average magnitude of a typical poly(A) shift. We identify 6 classes of genes whose poly(A) profiles behave similarly across the strains tested and whose 3′ UTRs exhibit distinct RNA sequence preferences. Mutations in chromatin-modifying activities, but not cleavage/polyadenylation factors, often have non-uniform effects on the same-gene, 3′ mRNA isoform levels. These observations suggest that chromatin differentially affects Pol II dwell time at individual sites during transcriptional elongation.

Materials and methods

Yeast strains

For the yeast genes examined here that are essential for viability, we obtained anchor-away (AA) alleles (Haruki et al. 2008) for rapid rapamycin-induced depletion of the protein products from cell nuclei. The Dpb2-AA strain YAM2673 (Cloutier et al. 2016) was provided by Elizabeth Tran, and the Top2-AA strain HHY173 (Haruki et al. 2008) was purchased from Euroscarf. For almost all nonessential yeast genes in this study, we used CRISPR to construct strains harboring precise deletions of protein-coding sequences (Supplementary File 1). To make each strain, we first constructed a Cas9-guide molecule by ligating the Cas9 parental vector pML104 (Laughery et al. 2015) and an annealed oligonucleotide pair representing a guide sequence targeting the desired coding sequence. We co-transformed the parental yeast strain JGY2000 (Geisberg et al. 2014) with this Cas9-guide molecule and a repair molecule (either a single oligonucleotide or double-stranded molecule resulting from mutually primed synthesis of two oligonucleotides) bridging the 5′ and 3′ flanking sequences of the targeted gene. The *hbs1Δ, ski7Δ,* and *air1Δ, air2Δ* double deletion strains are the products of sequential single deletions by the same method. We constructed the *gcn5::LEU2* strain by integrative transformation of JGY2000 with the *SacI-SphI* insert from plasmid MB25 (provided by Mark Benson). The *nhp6aΔ, nhp6bΔ* double deletion strain (supplied by Dan Hall) was produced by sequential PCR-mediated deletions (Longtine et al. 1998) in parental strain BY4741 (Brachmann et al. 1998). Several of the deletion and anchor-away strains included in this work have been described previously (Jeronimo et al. 2019; Geisberg et al. 2020, 2024a; Khitun et al. 2023). Strains used and additional details on their construction are listed in Supplementary File 1.

Cell culture

We cultured all the deletion strains in liquid YPD medium at 30 °C to an OD_{600} of ∼0.4 before collecting cells from 25 ml of culture by centrifugation and freezing them at −80 °C. We grew the anchor-away strains in the same manner but subjected them to a 1 h treatment with 1 μM rapamycin before harvesting.

RNA isolation and 3′ READS library construction

We isolated total RNA using the hot acid phenol procedure followed by Qiagen RNeasy column purification (including DNase I before column loading). We performed 3′ READS library construction in a manner identical to that previously described (Jin et al. 2015), using barcodes for multiplex sequencing, and determined individual library concentrations using an Agilent BioAnalyzer before pooling and sequencing on an Illumina NextSeq 500 instrument at the Bauer sequencing core of Harvard University.

Data analysis

We obtained demultiplexed sequences from the Bauer core and have previously described our pipeline, summarized below, for preparing/mapping the sequence reads and obtaining polyadenylation profiles (Geisberg et al. 2020). Because our library construction method targets the polyadenylated RNA population, our desired sequence reads (after adapter removal) should begin with one or more T nucleotides, representing terminal A nucleotides. A key feature of our preliminary analysis is keeping track of the number of initial Ts from each sequence read before trimming the remainder of the read to 17 nt and mapping to the Sac cer 3 build of the *Saccharomyces cerevisiae* genome (allowing 0 mismatches) using Bowtie (Langmead et al. 2009). This enables us to compare a read's recorded initial T count to the number of As downstream of its mapped position in the genome, and thus to infer whether the read represents a molecule with post-transcriptionally added As. We discarded all reads not meeting this criterion.

For each strain in this study, the numbers of initial sequence reads, mapped reads, and reads representing polyadenylated

products are shown in Supplementary Fig. 1. For each strain, we scaled the poly(A)-RNA-derived reads to 25 million per biological replicate and to 50 million reads for both combined replicates. We tabulated mapped, poly(A)-RNA-derived reads whose termini fell within 400 nt downstream of the stop codon (or annotated 3′ end, for noncoding RNAs) of each gene. The relative reads number for a given position represents isoforms ending at that position as a percentage of isoforms ending at the maximally used position in the 400 nt window. For subsequent analysis, we worked with the combined data from both replicates and limited the study to 3′ UTRs with a minimum of 500 total reads in the merged data set. For each gene, percentile coordinates are the 3′ UTR positions at which the specified percentage (10, 25, 50, 75, or 90) of isoform endpoints occurs at or upstream of the position.

Classification of genes by poly(A) profile shift

Genes were categorized by their poly(A) profile shift relative to isogenic controls as either upstream-shifted, downstream-shifted, or unshifted, essentially as described (Geisberg et al. 2020). Briefly, each gene's poly(A) profile was compared relative to either an isogenic or a pre-depletion control as follows. Genes were classified as upstream-shifted if they possessed both a negative overall net shift and if the net number of percentile coordinate positions (10th, 25th, 50th, 75th, and 90th) shifted upstream was ≥1. Genes were deemed downstream-shifted if they contained both a positive overall net shift and if the net number of percentile coordinates shifted downstream was ≥1. Genes were called as unshifted if either the overall net shift or the net number of percentile coordinates shifted equaled zero.

Correlations of biological replicates and poly(A) profiles across strains

Reproducibility of biological replicates was assessed by pairwise comparisons of reads in individual isoforms (Geisberg et al. 2020). All isoforms with ≥10 reads in both replicates (21,188 to 68,967) isoforms per strain (median 58,168) were selected, and a Pearson correlation was computed for each set of replicates (Supplementary File 1). Median poly(A) profile correlations were computed across 2,770 genes (≥500 reads per gene in all strains) for each strain:strain combination and are displayed in a heat map (Supplementary Fig. 1).

Hierarchical clustering of poly(A) profiles: overall similarity across strains

2,811 genes with ≥500 reads in 3′UTR in Anchor Away (AA), deletion (KO), and isogenic control strains were selected to assess overall poly(A) profile similarity among AA and KO strains. In order to minimize genetic background differences, poly(A) profiles for any given gene in the relevant isogenic control strains were first subtracted from the same gene's poly(A) profile in AA or KO strains. Isogenic control strains consisted of HHY168 (without rapamycin treatment) for all AA strains, BY4741 for the *nhp6BΔ nhp6BΔ* double deletion, and JGY2000 for all others. These net poly(A) profiles were then hierarchically clustered using the parameters distance = euclidean and hclust = ward.D. Dendrograms of strain similarity by poly(A) profile are plotted in Fig. 1b.

Hierarchical clustering of poly(A) profiles: similarity across gene groups

A total of 2,811 genes with ≥500 reads in the 3′UTRs of Anchor Away (AA), deletion (KO), and isogenic control strains were analyzed to identify groups of genes whose poly(A) profiles exhibit similar patterns across all strains. First, poly(A) profiles of every gene in the relevant isogenic control strain were subtracted from the same gene's poly(A) profiles in AA or KO strains. Isogenic control strains consisted of HHY168 (without rapamycin treatment) for all AA strains, BY4741 for the *nhp6BΔ nhp6BΔ* double deletion, and JGY2000 for all others. Next, these net poly(A) profiles were hierarchically clustered using the parameters distance = euclidean and hclust = ward.D. Clustering yielded 6 distinct groups according to the majority rule of indexes using the Nbclust package (Charrad et al. 2014); cutting the tree into additional groups resulted in groups with few members.

Histograms of net percentile shifts

For each of the 5 percentile coordinates (10th, 25th, 50th, 75th, and 90th) used in this study, we first computed the net shift for a given gene at each percentile by subtracting the value of each percentile coordinate in the isogenic control from the value of the same percentile coordinate in every AA or KO strain. For example, for the GCN4 gene, if the values of the 50th percentile positions in the JGY2000 and the *ref2Δ* strains are 88 and 112, respectively, then we recorded the net shift in the 50th percentile coordinate of GCN4 in the *ref2Δ* strain as +24. After tabulating all the net shifts at each of the 5 percentile coordinates for all genes in all AA and KO strains, we calculated the percentage of all genes in every AA and KO strain which exhibited shifts of <−30, −30 to −10, −10 to −2, −2 to 0, 0 to +2, +2 to +10,+10 to +30, and >+30 (Supplementary Fig. 1).

GO enrichment analysis

Enriched GO categories were identified for each of the six groups of genes using the enrichGO function of the R BioConductor package (Yu et al. 2012). Categories possessing Bonferroni-corrected *P*-values of<0.01 were selected for further analysis. For additional validation of enriched GO categories, we performed a permutation test by randomly shuffling the six group identities (while keeping group membership proportions identical) for all genes 1,000 times and performing the same GO analysis on the shuffled datasets. We then tabulated the number of times (out of 1,000) that the enriched GO categories appeared as over-represented (Bonferroni-corrected *P* value < 0.01) in any of the 1,000 shuffles. Any categories that were enriched < 50 times in the shuffled dataset ($P < 0.05$) were deemed significant.

Analysis of k-mer distribution frequencies in the 6 groups

First, we tabulated the frequencies of all 2-mers to 6-mers in the 3′ UTR sequences (defined as from +1 relative to the stop codon to the 3′ end of each poly(A) profile) of all the genes in each of the six groups. We then computed Bonferroni-corrected chi-square probabilities for the likelihood that each k-mer was unevenly distributed among the groups. We then selected all k-mers possessing Bonferroni-corrected *P* values < 0.05 and performed a permutation test on each one as further validation. We shuffled the individual group identities 1,000 times for all the genes while keeping group membership proportions identical. For each shuffled instance, we computed Bonferroni-corrected *P* values for all 2- to 6-mers. We then tabulated the number of instances in which each of the k-mers previously deemed unevenly distributed (i.e. those with Bonferroni-corrected *P* values < 0.05) were also unevenly distributed in the shuffled datasets using the same significance cutoff. 18 k-mers possessed a permutation *P* value of <0.05 (i.e. found unevenly distributed <50 times out of 1,000 shuffles; Chi Square test) and were designated as unevenly distributed among the groups. Pearson residuals, which provide a visual

measure of k-mer frequency deviation from the expected, were computed for each k-mer using the formula $r_i = (O_i—E_i)/SQRT(E_i)$, where O_i and E_i are the number of observed and expected times, respectively, that a k-mer is found in a group i.

Slope and RMSE analysis

We limited our analysis to genes with ≥4 identically positioned major isoforms (isoforms with ≥5% of the reads of the maximally expressed isoform of that gene) in both the AA/KO and isogenic control strains (2,189–2,537 genes/per strain; median = 2,462 genes; >100,000 data points in total). We then computed the slope, Pearson R, and RMSE for each gene using least squares regression. Slopes were then multiplied by 100 and binned (Fig. 6a); each bin contains 1,456 to 44,657 slopes (median = 7,290). Within each bin, we then computed the medians of the RMSE for AA/KO vs isogenic control (blue line on graph in Fig. 6a) and for biological replicates (yellow line). Differences in slope and RMSE distributions of AA/KO strains vs biological controls were assayed via the Kolmogorov-Smirnov test (both $P < 10^{-300}$).

For analysis of RMSE by strain, we calculated both the median and the standard deviation of all RMSE values for each strain (2,189–2,537 genes/strain; median number of genes/strain = 2,463) for both the AA/KO vs control and biological replicate datasets (Fig. 6b). In the slope vs standard deviation of the slope analysis (Fig. 6c), we required each gene to (1) have ≥500 combined reads in all strains, (2) possess ≥4 identically positioned major isoforms in all AA/KO vs control as well as in all biological replicate comparisons and (3) contain slope measurements for >50% of all AA and KO strains (i.e. have slope values for 23 or more strains; >90% of genes had slope entries in 39 or more strains). For the 2,436 genes that passed these criteria, we computed each gene's median slope as well as the standard deviation of its slope across all the conditions.

Results

Poly(A) profiles in yeast mutant strains

Using the 3′READS procedure (Jin et al. 2015) and analytical methods (Geisberg et al. 2020, 2024a) identical to those described previously, we examined poly(A) profiles in 45 yeast strains with mutations in a wide variety of DNA- or RNA-based processes (Fig. 1a and Supplementary File 1). The mutant alleles are either gene deletions or anchor-away (AA) alleles (Haruki et al. 2008) that deplete a protein essential for cell viability. Two biological replicates of each strain exhibit high reproducibility on overall poly(A) profiles (Supplementary File 1). We combined the reads in biological replicates and limited most of our analyses to the 2,811 genes that are well expressed (≥500 reads) in all strains. Genes that are less well expressed do not have enough reads to obtain reliable poly(A) profiles.

Hierarchical clustering of strains by overall poly(A) profiles

To examine the relationships among the 45 yeast strains, we performed hierarchical clustering (Fig. 1b) and pairwise Pearson correlations (Supplementary Fig. 1a) on poly(A) profiles of 2,811 genes. As expected, strains with mutations in functionally related proteins tend to cluster together, although not always. Multiple strains lacking chromatin-modifying activities cluster together (see below) as do multiple strains defective for cleavage/polyadenylation. Chromatin-modifying activities and cleavage/polyadenylation factors are independently correlated with the relevant Pol II speed mutants on in-common shifted genes (Supplementary Fig. 1b and c), confirming the link between these activities and Pol II elongation. Polyadenylation patterns of strains deleted for the Rai1 and Rtt103 components of the Rai1-Rtt103-Rat1 decapping/endonuclease complex are most closely related to each other. The poly(A) profile of the strain deleted for the non-canonical poly(A) polymerase Trf4/Pap2 is closest to that of a double knockout strain lacking both the Trf4/Pap2-stimulating paralogs Air1 and Air2. These results emphasize the value of using overall poly(A) profiles as a detailed phenotypic assay.

In some cases, loss of proteins that appear to be functionally unrelated nevertheless show unexpectedly similar poly(A) profiles. The strain harboring a knockdown allele of Dbp2, the DEAD-box ATP-dependent RNA helicase involved in mRNA transport (Aydin et al. 2024), most closely resembles the strain deleted for Nhp6A and Nhp6B, chromatin-binding proteins that affect nucleosome dynamics (Dowell et al. 2010). Similar poly(A) profiles are observed in strains deleted for Sto1, the large subunit of the mRNA cap binding complex that affects translational initiation (Garre et al. 2012), and Rpb4, a Pol II subunit that plays a role in mRNA decay and 3′ processing (Lotan et al. 2005). Perhaps global similarity of poly(A) profiles can reveal heretofore unknown functional connections between these pairs of proteins.

Variation in upstream- and downstream-shifted genes in mutant strains

For every gene, we compared the poly(A) profiles of each mutant strain with its wild-type counterpart and determined the overall poly(A) shift (in nucleotides, either upstream or downstream). The results are presented in two ways. First, for each mutant strain, we counted the number of genes whose poly(A) profiles are significantly shifted (with respect to biological replicate controls; $P < 0.05$) upstream or downstream from the wild-type control strain and then calculated the upstream:downstream ratio (Fig. 2). This approach measures the overall poly(A) shift of each mutant strain (i.e. predominantly upstream- or downstream-shifted) on all well-expressed genes in that strain, but it does not account for differences in the average magnitudes of shifts among strains. In the second approach, we limited our analysis to genes with >500 reads in all strains and significantly (with respect to the biological replicate controls; $P < 0.01$) upstream- or downstream-shifted poly(A) profiles. This cutoff for statistical significance reliably identifies genes with highly significant poly(A) shifts at the expense of genes with smaller magnitude shifts; hence, it is less ideal for observing overall patterns. We also computed average magnitudes of upstream and downstream poly(A) shifts for all strains (Fig. 3). Supplementary Fig. 2 depicts genome-wide distributions of poly(A) shift magnitudes at five different percentile coordinates, where each percentile coordinate is defined by the percentage of reads that lie at or upstream of its position.

Mutant strains exhibit a wide range in poly(A) profile shifts

As analyzed by both approaches, poly(A) profiles in most strains differ from the wild-type control strains, albeit to various extents (Figs. 2 and 3). In strains deficient for Spt16, Spt6, Gcn5, Snf2, Sto1, and Rpb4, many genes have strongly upstream-shifted poly(A) profiles. Conversely, in strains deficient for factors involved in cleavage/polyadenylation (Pcf11, Ref2, Npl3, Syc1), components of TRAMP, a complex with poly(A) RNA polymerase activity (Trf4/Pap2, Air1, Air2); or proteins involved in transcriptional termination (Rai1, Rtt103); many genes show downstream-shifted profiles and few show upstream-shifted profiles. Interestingly, in nearly all cases showing overall upstream- or downstream-

a

Category	Number	Strains
Pol II elongation	11	*paf1Δ*, *spt4Δ*, *dst1Δ*, *hpr1Δ*, *sub1Δ*, *bye1Δ*, *elf1Δ*, *vps15Δ*, *bre1Δ*, *isw1Δ*, *vps34Δ*
Pol II modification	2	*ssn3Δ*, *rtr1Δ*
Chromatin remodeling	11	*chd1Δ*, *pfd1Δ*, *asf1Δ*, *snf2Δ*, *set1Δ*, *set2Δ*, *eaf3Δ*, *gcn5Δ*::LEU2, *nhp6AΔ nhp6BΔ*, Spt6-AA, Spt16-AA
Pre-mRNA splicing	2	*bud31Δ*, *isy1Δ*
mRNA cleavage/polyadenylation	4	*ref2Δ*, *npl3Δ*, *syc1Δ*, Pcf11-AA
mRNA turnover	9	*rrp6Δ*, *hbs1Δ ski7Δ*, *dom34Δ*, *ccr4Δ*, *dhh1Δ*, *rai1Δ*, *xrn1Δ*, *rtt103Δ*, Dbp2-AA
mRNA processing	5	*rpb4Δ*, *trf5Δ*, *pap2Δ*, *air1Δ air2Δ*, *sto1Δ*
DNA unwinding	1	Top2-AA

b

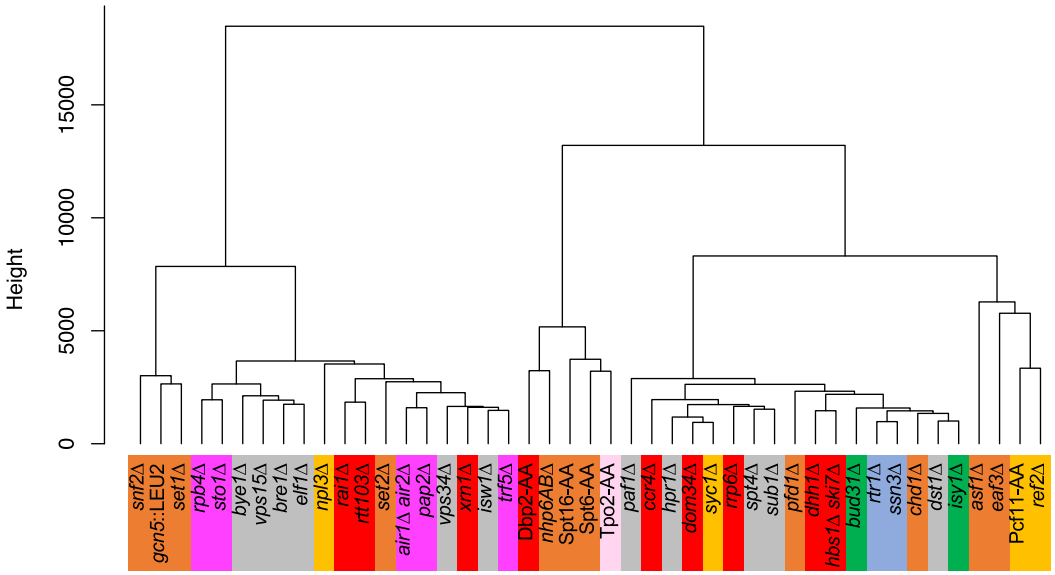

Fig. 1. Hierarchical clustering of 45 mutant strains. a) Mutant strains with complete deletion (Δ) and anchor-away (AA) alleles in the indicated categories. b) Clustering dendrogram.

shifted poly(A) profiles, there is a minority class of genes that behave in the opposite manner. Finally, strains deficient for some factors (Set1, Nhp6A/B, and Eaf3) have the unusual property that some genes show strong upstream poly(A) shifts and a roughly similar number exhibit strong downstream shifts.

The poly(A) shifts in many mutant strains, though statistically significant, are qualitatively modest, suggesting that deletion or cytoplasmic sequestration of the corresponding proteins has limited, if any, direct effects on cleavage/polyadenylation. In general, these mutant strains have poly(A) profiles in which many genes show weak downstream shifts and few genes show small upstream shifts (Fig. 3). This frequently observed poly(A) phenotype is not linked to the growth rate of the mutant strains. Moreover, this phenotype is not observed when wild-type cells are grown under non-optimal conditions such as galactose as a carbon source, osmotic stress, or minimal medium (Fig. 3). This suggests that the mild upstream shifts in many mutant strains arise from physiological imbalance, not merely slow growth.

Nucleosome remodeling and histone-modifying activities affect the rate of transcriptional elongation

Based on their effects on poly(A) profiles, the histone chaperones FACT or Spt6 increase the rate of transcriptional elongation, whereas histones H3 or H4 inhibit the elongation rate (Geisberg et al. 2024a). We extended this analysis by examining poly(A) profiles in strains lacking other chromatin-modifying activities (Fig. 3). Loss of Gcn5 histone acetylase or Swi/Snf nucleosome remodeling activity (Snf2) causes upstream-shifted poly(A) profiles similar to those caused by Pol II elongation mutants. These upstream-shifted poly(A) profiles are consistent with previous observations suggesting that Swi/Snf (Schwabish and Struhl 2007) and Gcn5 (Govind et al. 2007) are important for evicting histones during transcriptional elongation. Elimination of Set1 histone methylase results in a roughly equal proportion of upstream- and downstream-shifted poly(A) profiles. While many of these shifts are modest in magnitude, some genes show a strong upstream or downstream shift. Thus, enzymatic activities linked to

Strain	Upstream (percent)	Downstream (percent)	Non-shifted (percent)	Median percent of ORF reads	Total genes	Upstream to downstream ratio	Probability of ratio
air1Δ air2Δ	5	52	44	1.6	4371	0.09	$< 10^{-100}$
asf1Δ	11	20	69	1.5	4498	0.57	3.5×10^{-25}
bre1Δ	16	22	62	2.5	4344	0.74	5.4×10^{-10}
bud31Δ	11	41	48	2.3	4416	0.27	3.5×10^{-25}
bye1Δ	6	43	51	2.1	4366	0.15	$< 10^{-100}$
ccr4Δ	6	40	54	1.1	4409	0.16	$< 10^{-100}$
chd1Δ	9	37	55	1.6	4484	0.24	$< 10^{-100}$
Dbp2-AA	11	29	60	0.9	4330	0.36	9.6×10^{-88}
dhh1Δ	16	21	64	1.8	4432	0.77	6.2×10^{-8}
dom34Δ	4	33	64	1.9	4498	0.12	$< 10^{-100}$
dst1Δ	6	46	48	2.9	4488	0.14	$< 10^{-100}$
eaf3Δ	19	27	54	1.1	3834	0.71	1.2×10^{-12}
elf1Δ	13	23	64	2.6	4358	0.58	7.7×10^{-27}
gcn5::LEU2	40	24	36	1.3	3802	1.69	3.1×10^{-37}
hbs1Δ ski7Δ	31	7	62	5.3	4515	4.18	$< 10^{-100}$
hpr1Δ	8	39	53	1.1	4472	0.21	$< 10^{-100}$
isw1Δ	15	23	62	3.8	4467	0.68	1.0×10^{-15}
isy1Δ	5	46	50	2.8	4509	0.10	3.1×10^{-37}
nhp6AΔ nhp6BΔ	31	38	31	1.5	3816	0.80	1.3×10^{-8}
npl3Δ	4	53	44	3.7	4520	0.07	$< 10^{-100}$
paf1Δ	34	19	47	1.9	4399	1.82	2.1×10^{-45}
pap2Δ	6	37	56	1.9	4356	0.17	$< 10^{-100}$
Pcf11-AA	4	47	49	0.4	4322	0.09	$< 10^{-100}$
pfd1Δ	17	24	59	1.9	4462	0.72	2.7×10^{-12}
rai1Δ	7	48	45	2.0	4325	0.14	$< 10^{-100}$
ref2Δ	3	71	25	2.3	4435	0.05	$< 10^{-100}$
rpb4Δ	30	19	51	0.9	4271	1.56	4.7×10^{-24}
rrp6Δ	4	56	40	3.3	4396	0.08	$< 10^{-100}$
rtr1Δ	4	32	65	2.0	4498	0.12	$< 10^{-100}$
rtt103Δ	10	39	51	2.3	4335	0.26	$< 10^{-100}$
set1Δ	23	26	50	1.3	4432	0.88	1×10^{-3}
set2Δ	11	42	46	1.3	3430	0.26	$< 10^{-100}$
snf2Δ	35	24	41	0.9	3922	1.48	1.4×10^{-20}
Spt16-AA	47	5	49	0.6	4320	9.67	$< 10^{-100}$
spt4Δ	8	42	50	3.5	4526	0.20	$< 10^{-100}$
Spt6-AA	39	8	53	0.8	3958	4.62	$< 10^{-100}$
ssn3Δ	3	36	61	3.4	4518	0.08	$< 10^{-100}$
sto1Δ	29	17	54	1.7	4372	1.72	2.5×10^{-32}
sub1Δ	5	18	76	2.7	4486	0.30	3.6×10^{-72}
syc1Δ	4	42	54	2.1	4542	0.11	$< 10^{-100}$
Top2-AA	11	8	81	0.8	4407	1.39	1.7×10^{-6}
trf5Δ	15	47	38	2.3	4356	0.32	$< 10^{-100}$
vps15Δ	16	24	60	2.7	4273	0.64	2.2×10^{-19}
vps34Δ	14	27	59	1.5	4320	0.52	1.4×10^{-41}
xrn1Δ	17	23	61	3.0	4345	0.74	4.9×10^{-10}

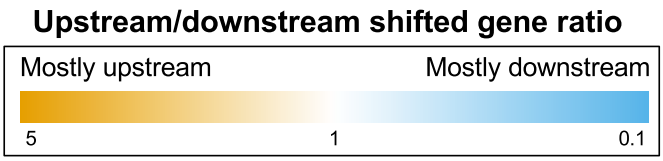

Fig. 2. Properties of poly(A) profiles in the 45 mutant strains. The following properties are indicated for each strain: the percent of genes with upstream-, downstream-, or non-shifted profiles; the median percentage of poly(A) sites in the coding region (ORF); the number of genes with >500 reads, the arbitrary number necessary to obtain reliable poly(A) profiles; the upstream:downstream ratio of genes with poly(A) shifts; the probability that this upstream:downstream ratio is significant with respect to control experiments involving biological replicates.

transcriptional activity and the generation of active chromatin are also important for transcriptional elongation.

In contrast, loss of Set2 histone methylase or (to a lesser extent) the Eaf3 subunit of the small Rpd3 histone deacetylase complex causes a downstream poly(A) shift (Fig. 3). Set2 and Eaf3 are mechanistically connected via the interaction of the Eaf3 chromodomain with methylated H3-K36 nucleosomes (Carrozza et al. 2005; Joshi and Struhl 2005; Keogh et al. 2005). Loss of either

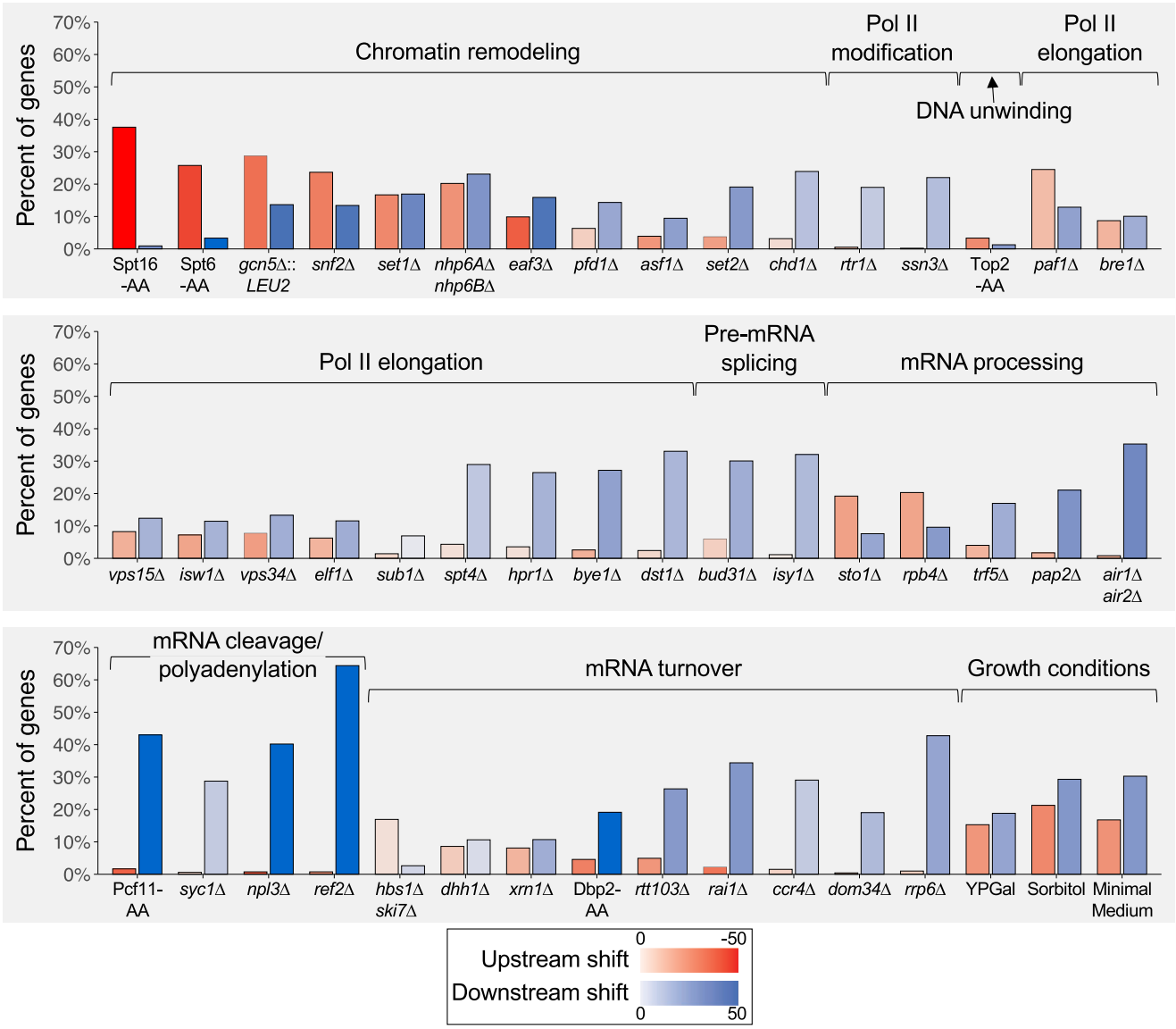

Fig. 3. Nature and magnitude of poly(A) shifts in 45 mutant strains. For each mutant strain, the percent of genes with poly(A) shifts that are upstream (red) or downstream (blue) profiles is shown. Median magnitudes of the poly(A) shifts are indicated by the darkness of the red or blue color.

protein results in increased histone acetylation in coding regions (Reid et al. 2004; Carrozza et al. 2005; Joshi and Struhl 2005; Keogh et al. 2005), the opposite of what occurs in the strain lacking Gcn5 histone acetylase, which shows an upstream poly(A) shift. Thus, histone-modifying activities can positively or negatively affect the rate of transcriptional elongation and thereby affect poly(A) profiles.

Hierarchical clustering of poly(A) profiles reveals 6 classes of genes

Each gene has a composite poly(A) shift profile consisting of upstream and downstream shifts for all 45 genetic conditions tested. We performed hierarchical clustering of these composite poly(A) profiles and identified 6 classes of genes with 125 to 789 members (Fig. 4 and Supplementary File 2). Genes belonging to a given class have similar median poly(A) shift profiles across the genetic conditions tested (Fig. 4b and Supplementary Fig. 3). For example, genes in group 1 exhibit the greatest variation (either upstream or downstream) in median poly(A) shift among strains, and this variation is readily observed for all percentiles (Supplementary Fig. 3). By contrast, group 5 genes exhibit median poly(A) shifts that are nearly unchanged across all percentiles irrespective of strain. Other groups exhibit varying levels of shift across percentiles, with strains harboring null or knock-down alleles in the chromatin regulatory proteins Spt6 and Spt16, as well as the cleavage/polyadenylation factors Pcf11, Ref2, and Npl3 exhibiting the greatest shifts within each group.

Median poly(A) profiles of gene groups vary in physical properties

S. cerevisiae genes can be separated into groups that differ by median poly(A) profile locations; these groups are evolutionarily conserved in *S. pombe* (Geisberg et al. 2024b). These groups of genes also exhibit variation in overall expression and are enriched for different functional categories (Geisberg et al. 2024b). Similarly, median poly(A) profiles for the 6 groups of genes identified here also vary significantly in location within the 3′UTR (Fig. 4c; $P < 0.001$ by PERMANOVA analysis) and exhibit differences in the

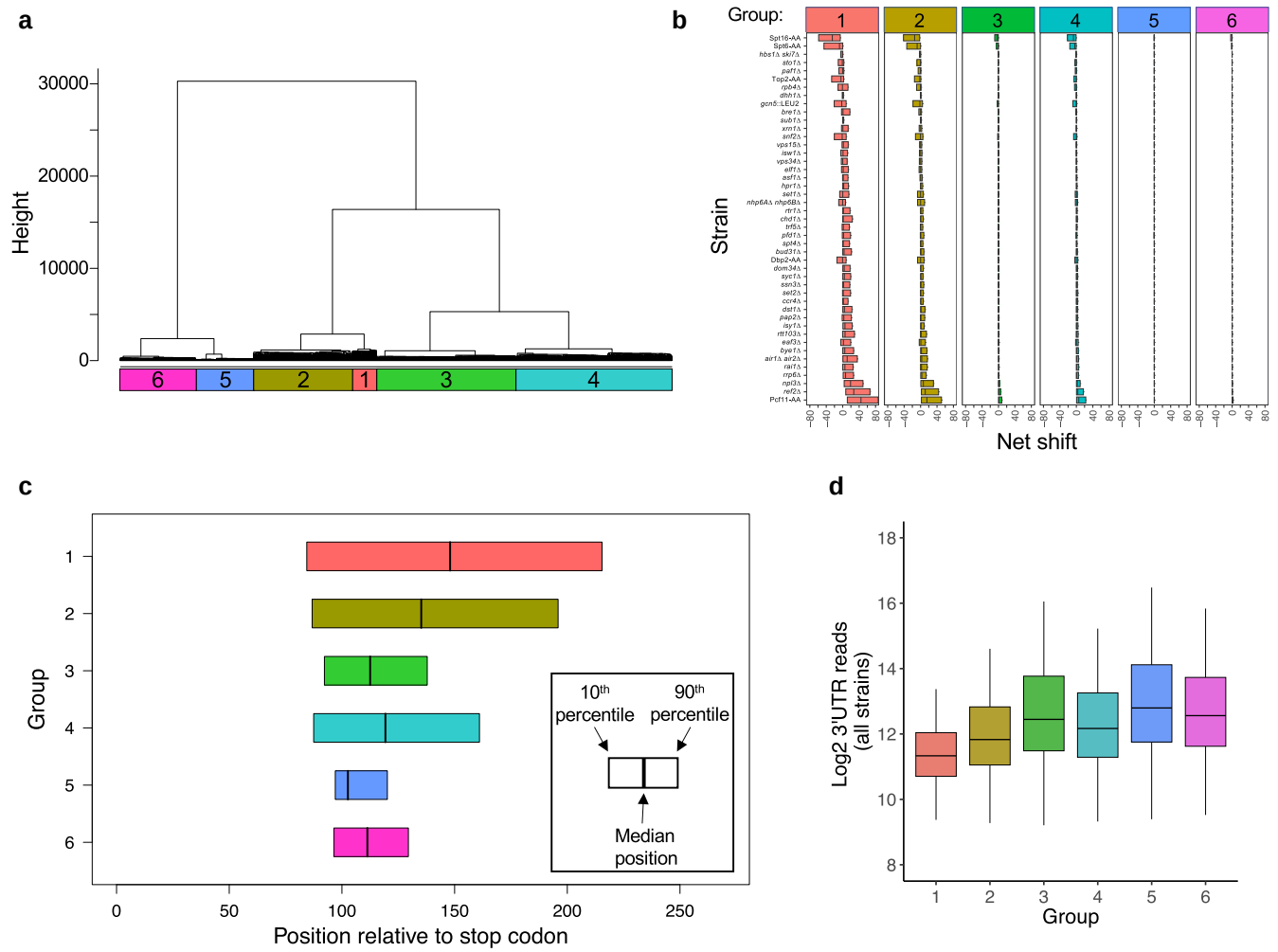

Fig. 4. Six classes of genes based on their similarity of poly(A) profiles across the set of mutant strains. a) Hierarchical clustering dendrogram. b) Magnitudes of the median poly(A) shifts of the six classes of genes in all 45 mutant strains. c) Location of poly(A) sites relative to the stop codon for the six classes of genes. Boxes indicate the locations of the 10th and 90th percentile coordinates, and vertical lines indicate the median position. d) Log_2 expression levels of the six classes of genes in the overall combined dataset.

median steady-state expression levels in all strains (Fig. 4d and Supplementary Fig. 4). Group 1, which contains genes that exhibit the greatest upstream and downstream poly(A) shifts, is the least transcribed group, whereas genes belonging to groups 5 and 6, which show the least poly(A) pattern variation, are the most highly expressed. Other parameters, including ORF length, strand, and type of or distance to the neighboring upstream or downstream features, all appear to be unlinked to group identity.

Interestingly, GO analysis of the 6 groups reveals 26 over-represented categories, most of which (> 20) are unique to a single group (Supplementary Fig. 5). Because our gene cohort is enriched for genes that are highly expressed across all strains, we performed a permutation analysis on the GO category list by shuffling group labels 1,000 times (while maintaining the same group sizes) and counting the number of times each category appears as significant during the 1,000 permutations. Using a somewhat stringent *P*-value cutoff of <0.05, which likely underestimates significance, group 3 is enriched in genes involved in nuclear export (GO:0051168), ribosome large subunit biogenesis (GO:0042273), periribosome large subunit precursor (GO:0030687), and ribosome biogenesis (GO:0042254). Group 4 is enriched in the peptidase complex category (GO:1905368), while group 1 genes are overrepresented in genes with ATP-dependent activity (GO:0140657).

Distinct sequence frequencies within the 3′ UTRs of the 6 gene classes

The simplest explanation for why the 6 classes of genes have similar poly(A) profiles across mutant strains is that they have distinct RNA sequence frequencies within their 3′ UTRs. To investigate this idea, we examined the frequency of all possible 2–6 nt sequences (k-mers) in 3′UTR subregions (+1 relative to the stop codon to the 3′-most major isoform) for the different groups (Fig. 5). We initially identified 18 k-mers whose frequency distributions were significantly skewed among the 6 groups as compared to shuffled controls (Bonferroni-corrected $P < 0.05$). Some of the k-mers are related by sequence composition (AG-rich, AT-rich, GG-rich, TGTA-containing k-mers) and exhibit similar patterns of over-representation (large blue circles) or under-representation (large red circles) among the 6 groups. Interestingly, two pentamers, ATTCA and CGGAC, are predominantly enriched in 3′UTR sequences of group 6 and group 1 genes, respectively.

Fig. 5. Sequence preferences in the 3′ UTR for the six classes of genes. For each listed k-mer, positive (blue) or negative (red) Pearson residual enrichment is indicated by color intensity, and significance is indicated by the size of the circle. The shuffled P value compares the actual data with data generated after shuffling genes among groups.

Chromatin-modifying, but not cleavage/polyadenylation, factors differentially affect poly(A) site utilization within individual genes

Upstream and downstream poly(A) shifts are measured by calculating the ratio of same-gene 3′ isoform levels in mutant vs wild-type strains (Geisberg et al. 2022). Transcriptome-level analysis of slow and fast Pol II derivatives indicates that upstream and downstream poly(A) shifts can be expressed as a linear slope, thereby indicating that they occur steadily from one isoform to the next across nucleotide distances (Geisberg et al. 2022). The simplest model to explain this observation is that the Pol II speed derivatives cause a uniform change in the dwell time of Pol II as it travels through the 3′ UTR. However, this model is based on overall poly(A) profiles and hence does not address the nature of poly(A) shifts of individual genes in various mutant strains.

If poly(A) shifts on an individual gene occur in a uniform manner during transcriptional elongation, the mutant:wild-type ratios of individual 3′ mRNA isoform levels across the gene should generate a linear slope corresponding to the magnitude of the shift (Fig. 6a). Depending on the gene/mutant combination, the slope can be positive, zero, or negative. Thus, for 2,811 genes in 45 mutant strains, we addressed how well the mutant:wild-type ratios fit a linear slope. As a control for experimental error in 3′ mRNA isoform level measurements, we performed the same analysis in biological replicates of each of the 45 mutant strains; by definition, control slopes should be zero. We defined the degree of fit to a linear slope as the median root-mean-square error (RMSE) as a function of slope. Uniform poly(A) shifts should have low RMSE values comparable to those of the biological replicates (Fig. 6a), whereas non-uniform shifts should have significantly higher RMSE values compared to controls (Fig. 6b).

Overall, poly(A) shifts in mutant strains show considerably higher median RMSE from a linear slope than the biological replicate controls ($P < 10^{-300}$, Mann-Whitney U and Kolmogorov-Smirnov tests), with the median amount of deviation increasing at higher slopes (Fig. 6c). Interestingly, the median amount of RMSE differs greatly among mutant strains (Fig. 6d; compare x-axis placement of blue circles representing mutant strains to orange circles representing biological replicate controls). The highest level of median RMSE occurs in mutant strains involving chromatin-modifying and topoisomerase activities (Spt16, Spt6, Top2, Snf2, Gcn5). This observation suggests that chromatin-based effects on Pol II elongation rate differentially affect cleavage/polyadenylation at different sites within the same gene. In addition, the strain depleted for Dbp2, an RNA helicase involved in transport, also shows a high median RMSE from a linear slope. In contrast, median RMSEs from a linear slope in mutant strains defective in cleavage/polyadenylation components (Pcf11, Npl3, Ref2) are comparable to those observed in the biological replicate controls. This latter observation suggests that the level of cleavage/polyadenylation occurs in a uniform manner as Pol II travels through the 3′ UTR, presumably reflecting the sequence requirements for cleavage/polyadenylation at individual sites.

Uniformity of poly(A) shifts can differ among genes and strains

The RMSEs for individual mutant conditions described above are based on the median values of 2,811 genes and hence do not address differences among genes. To address gene-specific differences, we first calculated the median standard deviation of RMSEs for all genes in each of the mutant conditions (Fig. 6d; compare y-axis placement of blue and orange circles). Many, but not

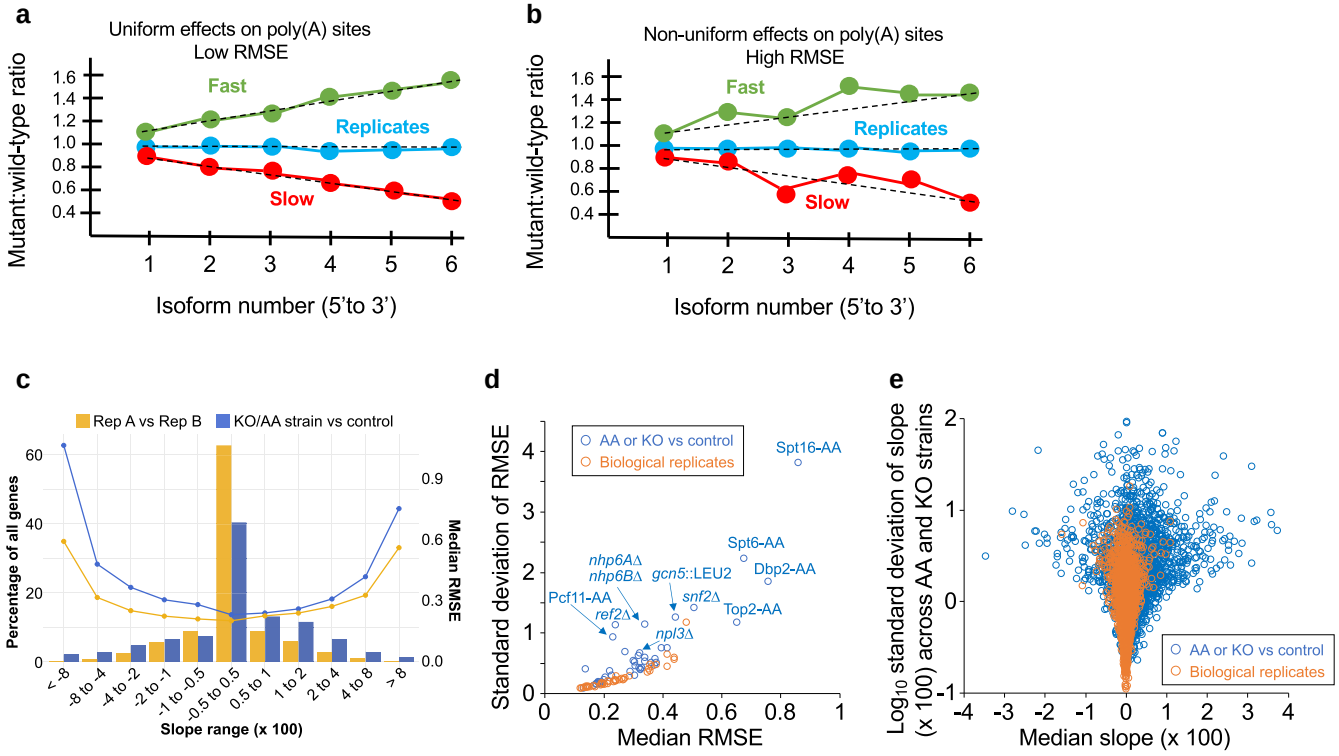

Fig. 6. Uniformity of poly(A) shifts differs among strains and genes. a) Theoretical data for poly(A) profiles resulting from uniform Pol II elongation rates. Mutant:wild-type ratios of levels of 3′ isoforms 1–6 (defined by 5′ to 3′ location) of an individual gene in strains with fast Pol II elongation (green), slow elongation (red), or biological replicates (blue). The slopes are indicated by lines, and the degree of fit of data points to the theoretical slope line (black dashed line) is defined as the route-mean-square error (RMSE). RMSE values in biological replicates reflect experimental error, and uniform effects of fast or slow elongation have RMSEs values similar to those of biological replicates. b) Theoretical data for poly(A) profiles resulting from non-uniform Pol II elongation rates. Same as panel a) except that data points show greater deviation from the slope line and hence have higher RMSE values. c) Percentage of genes (boxes) and median root-mean-square error (RMSE lines) as a function of slope (×100) for biological replicates (yellow) and mutant:wild-type ratio. d) Relationship between uniformity of the poly(A) shift (standard deviation of RMSE) as a function of median RMSE in pairwise comparisons of mutant and wild-type control strains (blue) and in biological replicates (orange). Mutant strains with high standard deviations of RMSE and/or median RMSE are indicated. e) Relationship between standard deviations of slopes and median slopes for 2,811 genes across all strains in pairwise comparisons of mutant and wild-type control strains (blue) and biological replicates (orange).

all, mutant conditions show higher median standard deviations than the biological replicate controls, with Spt16, Spt6, and Dbp2 showing particularly high standard deviations. Thus, in some mutant strains, the uniformity of poly(A) shifts differs among genes.

To address gene-specific differences more directly, we calculated medians and standard deviations of the slopes for 2,436 genes in each of the 45 conditions and in all biological replicates. Overall, individual genes show somewhat higher median slopes and standard deviations of the slope than in biological replicates (Fig. 6e, compare blue and orange circles). A significant number of genes possess high slope standard deviation (i.e. blue circles >1 on the y-axis) and are part of a cohort whose magnitude of poly(A) shift is highly strain-dependent. High median values and/or high standard deviations suggest that poly(A) shifts do not occur in a uniform manner as Pol II traverses the 3′ UTR.

Discussion

Many chromatin-modifying activities affect the Pol II elongation rate: Implication for disease

Pol II derivatives that decrease or increase the elongation rate respectively cause upstream and downstream poly(A) shifts (Geisberg et al. 2020, 2022). A decreased Pol II elongation rate is also associated with upstream poly(A) shifts due to environmental conditions (diauxic shift) or mutations in histone chaperones (FACT and Spt6) (Geisberg et al. 2020, 2024a). Although not directly tested *in vivo*, it is presumed from nucleosome inhibition of Pol II elongation *in vitro* that the downstream poly(A) shifts observed upon histone depletion reflect increased Pol II elongation rate. As it is difficult to imagine how chromatin could directly affect the cleavage/polyadenylation reaction that occurs on RNA, these observations strongly suggest that poly(A) profiles represent a transcriptome-level assay of Pol II elongation *in vivo*.

Here we extend our previous analysis of poly(A) profiles to show that histone acetylase Gcn5, nucleosome remodeler Swi/Snf, and chromatin modulator Paf1 are important for transcriptional elongation *in vivo*. Mutations in any of these components cause an overall upstream-shifted poly(A) profile that strongly resembles that of strains containing slow Pol II mutants or depleted of histone chaperones. In contrast, the loss of Set2 histone methylase or the Chd1 chromatin remodeler causes increased histone acetylation within coding regions and a downstream poly(A) shift. Thus, many chromatin-modifying activities affect the rate of transcriptional elongation.

Mutations in a wide variety of chromatin-modifying activities are associated with human diseases such as cancer. As these chromatin-modifying activities affect transcriptional activation and/or repression of many genes, it is generally presumed that

the disease phenotypes are associated with changes in mRNA levels. However, in addition to such effects on gene expression levels, our results indicate that many chromatin-modifying activities, via their effects on Pol II elongation, alter poly(A) profiles on a transcriptome scale. Such disease-causing mutations in chromatin-modifying activities might also affect phenotypes by changing the constellation of 3′ mRNA isoforms. For example, upstream poly(A) shifts favor shorter 3′ mRNA isoforms that are associated with the oncogenic state, and such shorter 3′ mRNA isoforms are more likely to lack target sites for microRNAs that inhibit expression levels. In a similar vein, chromatin-modifying activities can also affect mRNA isoform distributions via effects on splicing (Bhattacharya et al. 2021; Francette and Arndt 2024; Redin et al. 2024). By analogy with cleavage/polyadenylation, it seems likely that chromatin affects alternative mRNA splicing indirectly via the Pol II elongation rate.

Gene- and factor-specific differences in poly(A) shifts

Although overall upstream or downstream poly(A) shifts occur in most mutant strains, these poly(A) shifts do not occur in all genes, and a minority class of genes displays poly(A) shifts in the opposite direction of the majority class. Using hierarchical clustering of poly(A) profiles, we identified 6 classes of genes that behave similarly over the 45 mutant strains. Variation in 3′ UTR sequence preferences among the 6 gene classes likely underlies their similar poly(A) profiles under diverse genetic conditions. In addition to gene-specific differences in an individual mutant strain, poly(A) shifts of individual genes differ among the mutant strains. Strains showing an overall upstream poly(A) shift differ in which genes are affected, particularly if the factors affect different biochemical activities involved in transcriptional elongation. Strains showing an overall downstream poly(A) shift also differ with respect to the affected genes, with the notable exception of mutations directly affecting cleavage/polyadenylation. Strains with mutations in cleavage/polyadenylation factors Pcf11, Ref2, and Npl3 exhibit much greater overlap of downstream-shifted genes than any combination of chromatin factors, along with virtually no upstream-shifted genes. These observations suggest that much of the gene- and factor-specific effects on poly(A) profiles are due to the process of transcriptional elongation, not cleavage/polyadenylation itself.

The relationship between Pol II elongation and cleavage/polyadenylation

As elongating Pol II traverses the gene, there is a kinetic competition at each nucleotide between cleavage/polyadenylation, which generates 3′ mRNA isoforms and triggers transcriptional termination, and continued Pol II elongation. The general model for how the Pol II elongation rate affects the poly(A) profile invokes altered kinetic competition linked to the dwell time of Pol II at individual nucleotides as it traverses the gene. When Pol II elongates slowly, dwell time at individual nucleotides is longer, thereby allowing more efficient cleavage/polyadenylation. As cleavage/polyadenylation only occurs once per elongating Pol II molecule, increased cleavage/polyadenylation due to increased Pol II dwell time results in an upstream poly(A) shift. Conversely, a fast elongation rate results in reduced Pol II dwell time at individual nucleotides and hence reduced cleavage/polyadenylation and a downstream-shifted poly(A) profile. This kinetic competition also explains why a defect in the cleavage/polyadenylation machinery causes a downstream-shifted poly(A) profile.

Although a simple kinetic competition model explains overall poly(A) shifts, it does not account for gene-specific differences, including the genes in which the shift occurs in the opposite direction from the majority class. Such gene-specific differences strongly suggest that the Pol II elongation rate is not uniform and is affected by the DNA sequence. Indeed, poly(A) profiles of individual genes in mutant strains defective in chromatin-modifying activities show significant RMSE deviation from a linear slope. Non-uniformity of the elongation rate across a gene is further supported by the observation that Pol II speed mutants have altered Pol II pausing patterns (Khitun et al. 2023). These Pol II mutations are located in the trigger loop, which affects entry to and stability of the nucleotide precursor at the active site (Wang et al. 2006; Vassylyev et al. 2007; Mazumder et al. 2020). Unlike Pol II elongation, the cleavage/polyadenylation reaction *per se* occurs at a uniform rate, because strains defective in cleavage/polyadenylation behave similarly to biological replicates with respect to RMSE deviation. Although the kinetic component of cleavage/polyadenylation is uniform throughout the gene, the levels of cleavage/polyadenylation at different sites within an individual gene depend on the RNA sequence.

In wild-type strains, Pol II pauses at favored positions indicative of increased dwell time at those positions (Couvillion et al. 2022; Khitun et al. 2023). The location of Pol II pauses influences the choice of poly(A) sites and hence the poly(A) profile because there is a stereochemical relationship between the location of elongating Pol II and cleavage/polyadenylation (Geisberg et al. 2022). Mutations in many chromatin-modifying factors alter the location of Pol II pauses in coding regions, and presumably in 3′ UTRs (Couvillion et al. 2022), indicative of nucleotide-level changes in elongation rate. Thus, opposite poly(A) shifts in the minority class of genes likely reflect local changes in Pol II elongation rate/dwell time that go in the opposite direction from the overall change. For example, in strains showing a general upstream poly(A) shift due to increased Pol II dwell time, DNA sequences linked to a specific poly(A) site might cause decreased dwell time, leading to a downstream shift. Local changes in Pol II elongation rate are likely to arise from effects on nucleosome occupancy, positioning, stability, or dynamics.

Local changes in Pol II dwell time do not easily explain the small minority of genes showing atypical upstream shifts in strains defective in cleavage/polyadenylation. One possibility is that sequences in some 3′ UTRs affect the connection between elongating Pol II and the cleavage/polyadenylation machinery, leading to increased Pol II dwell time and hence upstream shifts. Alternatively, a defect in cleavage/polyadenylation should inhibit transcriptional termination of antisense transcripts, resulting in readthrough transcription that might inhibit Pol II elongation of the gene.

The gene-specific effects on poly(A) profiles in mutant strains must ultimately be due to sequence differences among 3′ UTRs. In this regard, the 3′UTRs of the 6 classes of genes have different nucleotide preferences. It is likely that these different nucleotide preferences have different effects on chromatin structure and transcriptional elongation that underlie why each class has similar poly(A) profiles among the mutant strains.

Data availability

3′ READS data sets not published before this work are available in the Gene Expression Omnibus (GEO) under accession number GSE312577. We have previously published some of the 3′ READS data sets included in the present analysis (Geisberg et al. 2020,

2024a; Khitun et al. 2023). The GEO accession numbers for these are as follows: GSE262747 for HHY168, JZY21 (*dst1Δ*), JZY23 (*paf1Δ*), JZY24 (*ref2Δ*), JZY63 (*npl3Δ*), YJ012 (Pcf11-AA), YFR1478 (Spt16-AA), and YFR1480 (Spt6-AA); GSE151196 for JZY27 (*spt4Δ*) and JZY33 (*hpr1Δ*); and GSE234406 for JZY65 (*rpb4Δ*). The code underlying the analyses is presented in Supplementary File 3.

Supplemental material available at GENETICS online.

Acknowledgments

We thank Catherine Maddox for superb technical assistance, Andrew Li for helping with the construction of the *ssn3*Δ and *rtr1*Δ deletion strains, and Mark Benson, Dan Hall, Yi Jin, François Robert, and Elizabeth Tran for strains and plasmids.

Funding

This work was supported by a grant to K.S. from the National Institutes of Health (GM 131801).

Conflicts of interest

None declared.

Literature cited

Aydin E *et al.* 2024. DEAD-box ATPase Dbp2 is the key enzyme in an mRNP assembly checkpoint at the 3′-end of genes and involved in the recycling of cleavage factors. Nat Commun. 15:6829. https://doi.org/10.1038/s41467-024-51035-z.

Bhattacharya S *et al.* 2021. The methyltransferase SETD2 couples transcription and splicing by engaging mRNA processing factors through its SHI domain. Nat Commun. 12:1443. https://doi.org/10.1038/s41467-021-21663-w.

Brachmann CB *et al.* 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast. 14:115–132. https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2.

Carrozza MJ *et al.* 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell. 123:581–592. https://doi.org/10.1016/j.cell.2005.10.023.

Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. Nbclust: an R package for determining the relevant number of clusters in a dada set. J Stat Softw. 61:1–36. https://doi.org/10.18637/jss.v061.i06.

Cloutier SC *et al.* 2016. Regulated formation of lncRNA-DNA hybrids enables faster transcriptional induction and environmental adaptation. Mol Cell. 61:393–404. https://doi.org/10.1016/j.molcel.2015.12.024.

Couvillion M *et al.* 2022. Transcription elongation is finely tuned by dozens of regulatory factors. Elife. 11:e78944. https://doi.org/10.7554/eLife.78944.

Dowell NL, Sperling AS, Mason MJ, Johnson RC. 2010. Chromatin-dependent binding of the *S. cerevisiae* HMGB protein Nhp6A affects nucleosome dynamics and transcription. Genes Dev. 24:2031–2042. https://doi.org/10.1101/gad.1948910.

Francette AM, Arndt KM. 2024. Multiple direct and indirect roles of the Paf1 complex in transcription elongation, splicing, and histone modifications. Cell Rep. 43:114730. https://doi.org/10.1016/j.celrep.2024.114730.

Garre E *et al.* 2012. Yeast mRNA cap-binding protein Cbc1/Sto1 is necessary for the rapid reprogramming of translation after hyperosmotic shock. Mol Biol Cell. 23:137–150. https://doi.org/10.1091/mbc.e11-05-0419.

Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. 2014. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell. 156:812–824. https://doi.org/10.1016/j.cell.2013.12.026.

Geisberg JV *et al.* 2022. Nucleotide-level linkage of transcriptional elongation and polyadenylation. eLife. 11:e83153. https://doi.org/10.7554/eLife.83153.

Geisberg JV, Moqtaderi Z, Struhl K. 2020. The transcriptional elongation rate regulates alternative polyadenylation in yeast. eLife. 9:e59810. https://doi.org/10.7554/eLife.59810.

Geisberg JV, Moqtaderi Z, Struhl K. 2024a. Chromatin regulates alternative polyadenylation via the RNA polymerase II elongation rate. Proc Natl Acad Sci U S A. 121:e2405827121. https://doi.org/10.1073/pnas.2405827121.

Geisberg JV, Moqtaderi Z, Struhl K. 2024b. Location of polyadenylation sites within 3′ untranslated regions is linked to biological function in yeast. Genetics. 228:iyae163. https://doi.org/10.1093/genetics/iyae163.

Govind CK, Zhang F, Qiu H, Hofmeyer K, Hinnebusch AG. 2007. Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. Mol Cell. 25:31–42. https://doi.org/10.1016/j.molcel.2006.11.020.

Haruki H, Nishikawa J, Laemmli UK. 2008. The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell. 31:925–932. https://doi.org/10.1016/j.molcel.2008.07.020.

Jeronimo C, Poitras C, Robert F. 2019. Histone recycling by FACT and Spt6 during transcription prevents the scrambling of histone modifications. Cell Rep. 28:1206–1218.e8. https://doi.org/10.1016/j.celrep.2019.06.097.

Jin Y *et al.* 2015. Mapping 3′ mRNA isoforms on a genomic scale. Curr Protoc Mol Biol. 110:4.23.21–24.23.17. https://doi.org/10.1002/0471142727.mb0423s110.

Joshi AA, Struhl K. 2005. Interaction of the Eaf3 chromodomain with methylated histone H3-K36 mediates preferential histone deacetylation at mRNA coding regions. Mol Cell. 20:971–978. https://doi.org/10.1016/j.molcel.2005.11.021.

Keogh MC *et al.* 2005. Cotranscriptional Set2 methylation of histone h3 lysine 36 recruits a repressive rpd3 complex. Cell. 123:593–605. https://doi.org/10.1016/j.cell.2005.10.025.

Khitun A *et al.* 2023. Elongation rate of RNA polymerase II affects pausing patterns across 3′ UTRs. J Biol Chem. 299:105289. https://doi.org/10.1016/j.jbc.2023.105289.

Kumar A, Clerici M, Muckenfuss LM, Passmore LA, Jinek M. 2019. Mechanistic insights into mRNA 3′-end processing. Curr Opin Struct Biol. 59:143–150. https://doi.org/10.1016/j.sbi.2019.08.001.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Laughery MF *et al.* 2015. New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. Yeast. 32:711–720. https://doi.org/10.1002/yea.3098.

Longtine MS *et al.* 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. Yeast. 14:953–961. https://doi.org/10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U.

Lotan R *et al.* 2005. The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. Genes Dev. 19:3004–3016. https://doi.org/10.1101/gad.353205.

Lui KH, Geisberg JV, Moqtaderi Z, Struhl K. 2022. 3′ untranslated regions are modular entities that determine polyadenylation profiles. Mol Cell Biol. 42:e0024422. https://doi.org/10.1128/mcb.00244-22.

Mason PB, Struhl K. 2005. Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo*. Mol Cell. 17:831–840. https://doi.org/10.1016/j.molcel.2005.02.017.

Mazumder A, Lin M, Kapanidis AN, Ebright RH. 2020. Closing and opening of the RNA polymerase trigger loop. Proc Natl Acad Sci U S A. 117:15642–15649. https://doi.org/10.1073/pnas.1920427117.

Moqtaderi Z, Geisberg JV, Jin Y, Fan X, Struhl K. 2013. Species-specific factors mediate extensive heterogeneity of mRNA 3′ ends in yeasts. Proc Natl Acad Sci U S A. 110:11073–11078. https://doi.org/10.1073/pnas.1309384110.

Ozsolak F *et al.* 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell. 143:1018–1029. https://doi.org/10.1016/j.cell.2010.11.020.

Pelechano V, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. Nature. 497:127–131. https://doi.org/10.1038/nature12121.

Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. Cell. 108:501–512. https://doi.org/10.1016/S0092-8674(02)00617-7.

Redin E *et al.* 2024. SMARCA4 controls state plasticity in small cell lung cancer through regulation of neuroendocrine transcription factors and REST splicing. J Hematol Oncol. 17:58. https://doi.org/10.1186/s13045-024-01572-3.

Reid JL, Moqtaderi Z, Struhl K. 2004. Eaf3 regulates the global pattern of histone acetylation in *Saccharomyces cerevisiae*. Mol Cell Biol. 24:757–764. https://doi.org/10.1128/MCB.24.2.757-764.2004.

Schwabish MA, Struhl K. 2007. The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. Mol Cell Biol. 27:6987–6995. https://doi.org/10.1128/MCB.00717-07.

Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem Sci. 38:312–320. https://doi.org/10.1016/j.tibs.2013.03.005.

Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol. 18:18–30. https://doi.org/10.1038/nrm.2016.116.

Vassylyev DG *et al.* 2007. Structural basis for substrate loading in bacterial RNA polymerase. Nature. 448:163–168. https://doi.org/10.1038/nature05931.

Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. 2006. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell. 127:941–954. https://doi.org/10.1016/j.cell.2006.11.023.

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 16:284–287. https://doi.org/10.1089/omi.2011.0118.

Editor: E. Tran