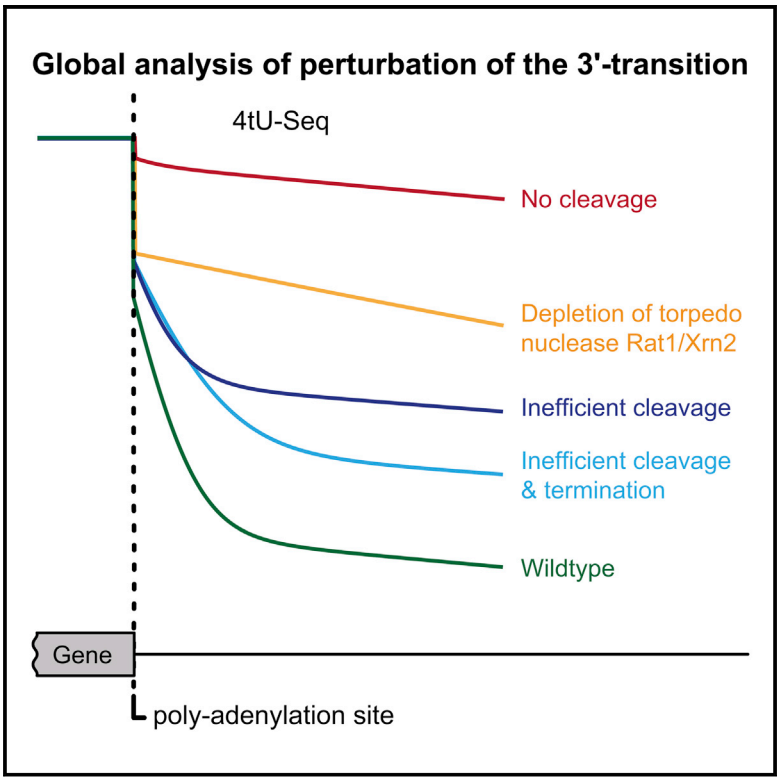
Article

Molecular Cell

Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes

Graphical Abstract



Highlights

- Pcf11 and Ysh1 are globally required for the 3′-transition at protein-coding genes
- Spt5 binds pre-mRNA near the pA site and contributes to the 3′-transition
- Rat1 is globally required for Pol II release from DNA in a narrow termination window

Authors

Carlo Baejen, Jessica Andreani, Phillipp Torkler, ..., Stephanie Esslinger, Johannes Söding, Patrick Cramer

Correspondence

johannes.soeding@mpibpc.mpg.de (J.S.), patrick.cramer@mpibpc.mpg.de (P.C.)

In Brief

RNA polymerase II undergoes a concerted transition at the 3′ end of protein coding genes. Baejen et al. present a genome-wide analysis of this 3′-transition in budding yeast. They show that this transition requires the Spt5 elongation factor and demonstrate that polymerase II release from DNA requires the Rat1 exonuclease.

Baejen et al., 2017, Molecular Cell 66, 1–12
April 6, 2017
http://dx.doi.org/10.1016/j.molcel.2017.02.009

CellPress

Please cite this article in press as: Baejen et al., Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes, Molecular Cell (2017), http://dx.doi.org/10.1016/j.molcel.2017.02.009

Molecular Cell
Article

Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes

Carlo Baejen,[1,4] Jessica Andreani,[1,4,5] Phillipp Torkler,[1] Sofia Battaglia,[1] Bjoern Schwalb,[1] Michael Lidschreiber,[3] Kerstin C. Maier,[1] Andrea Boltendahl,[1] Petra Rus,[1] Stephanie Esslinger,[2] Johannes Söding,[1,*] and Patrick Cramer[1,6,*]
[1]Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[2]Gene Center Munich and Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany
[3]Karolinska Institutet, Department of Biosciences and Nutrition, Center for Innovative Medicine and Science for Life Laboratory, Novum, Hälsovägen 7, 141 83 Huddinge, Sweden
[4]Co-first author
[5]Present address: Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette Cedex, France
[6]Lead Contact
*Correspondence: johannes.soeding@mpibpc.mpg.de (J.S.), patrick.cramer@mpibpc.mpg.de (P.C.)
http://dx.doi.org/10.1016/j.molcel.2017.02.009

SUMMARY

At the end of protein-coding genes, RNA polymerase (Pol) II undergoes a concerted transition that involves 3′-processing of the pre-mRNA and transcription termination. Here, we present a genome-wide analysis of the 3′-transition in budding yeast. We find that the 3′-transition globally requires the Pol II elongation factor Spt5 and factors involved in the recognition of the polyadenylation (pA) site and in endonucleolytic RNA cleavage. Pol II release from DNA occurs in a narrow termination window downstream of the pA site and requires the "torpedo" exonuclease Rat1 (XRN2 in human). The Rat1-interacting factor Rai1 contributes to RNA degradation downstream of the pA site. Defects in the 3′-transition can result in increased transcription at downstream genes.

INTRODUCTION

During transcription termination, RNA polymerase discontinues elongation of the RNA product and releases the DNA template. Bacterial and archaeal RNA polymerases terminate when they transcribe certain DNA sequences (Ray-Soni et al., 2016). In contrast, the eukaryotic RNA polymerase II (Pol II) requires many accessory proteins to release DNA and mRNA at the end of protein-coding genes. Pol II termination is tightly linked to the processing of the nascent transcript 3′ end (Connelly and Manley, 1988; Whitelaw and Proudfoot, 1986; Birse et al., 1998). Such coupling between pre-mRNA 3′-processing and transcription termination ensures complete formation of stable mRNA products that are competent for nuclear export and translation in the cytoplasm. We refer to coupled 3′-processing and termination as "3′-transition".

A large body of biochemical and genetic work has elucidated how the Pol II machinery accomplishes the 3′-transition in the budding yeast *S. cerevisiae*. When Pol II transcribes over the polyadenylation (pA) site, the nascent RNA displays the pA sequence and the protein machinery for pre-mRNA cleavage and polyadenylation is recruited (Keller and Minvielle-Sebastia, 1997; Mischo and Proudfoot, 2013; Porrua and Libri, 2015; Manley and Takagaki, 1996). This machinery includes the multiprotein complexes CPF (cleavage and polyadenylation factor) and CF (cleavage factor, consisting of CFIA and CFIB), which are arranged in a 5′-3′ direction (Baejen et al., 2014). Several subunits in CPF and CF additionally bind the Pol II C-terminal repeat domain (CTD). In particular, the CFIA subunit Pcf11 contains a CTD-interacting domain that preferentially binds the Ser2-phosphorylated CTD (Komarnitsky et al., 2000; Mayer et al., 2012a; Meinhart and Cramer, 2004). Efficient recruitment of CFIA also requires the C-terminal region (CTR) of the Pol II-associated general elongation factor Spt5 (Mayer et al., 2012b).

After assembly of the 3′-processing machinery, the endonucleolytic CPF subunit Ysh1/Brr5 cleaves RNA (Mandel et al., 2006; Zhelkovsky et al., 2006). The resulting RNA 3′ end is polyadenylated to complete the mRNA transcript. The new 5′ end of the nascent RNA is unprotected and prone to degradation by exonucleases. Two decades ago, a model was proposed that an exonuclease degrades nascent RNA to chase the transcribing polymerase and promote its release from DNA and RNA (Connelly and Manley, 1988; Proudfoot, 1989). Evidence for this model was later obtained for the yeast (Kim et al., 2004) and human (West et al., 2004) systems, with the 5′-3′ RNA exonuclease Rat1/XRN2 (yeast/human) being responsible for termination. Based on changes in polymerase subunit Rpb3 occupancy on chromosome III, it was concluded that Rat1 probably has a function in transcription termination at multiple loci throughout the yeast genome (Kim et al., 2004). This termination mechanism is now commonly referred to as "torpedo" termination, because the exonuclease acts like a torpedo to dismantle the Pol II elongation complex (Luo and Bentley, 2004).

The torpedo nuclease Rat1 forms a complex with Rai1 and Rtt103 (Dengl and Cramer, 2009; Kim et al., 2004). Rai1 stimulates Rat1 nuclease activity (Dengl and Cramer, 2009; Jiao et al., 2010) and has pyrophosphatase activity that is used for

removal of incomplete cap structures that can form downstream of the transcription start site. Rtt103 is a CTD-binding protein (Kim et al., 2004) and shows peak occupancy downstream of the pA site, near the occupancy peaks for the cleavage factor IA subunits Pcf11 and Rna15 (Lidschreiber et al., 2013; Mayer et al., 2012a). These peaks coincide with a decrease in CTD phosphorylation at residue Tyr1, which counteracts association of these factors and Pol II termination (Mayer et al., 2012a; Schreieck et al., 2014).

Pol II termination is highly context-dependent and requires 3′-processing factors (Mischo and Proudfoot, 2013). Pol II termination also requires the presence of an upstream pA site (Minvielle-Sebastia et al., 1998). Downstream of the pA site, the stability of the Pol II elongation complex is reduced (Grosso et al., 2012; Mason and Struhl, 2005), and Pol II often pauses (Gromak et al., 2006; Hyman and Moore, 1993; Kazerouninia et al., 2010; Larson et al., 2011). Such pausing is likely a prerequisite for termination, as known for bacterial RNA polymerase (Ray-Soni et al., 2016). Transcription termination sites (TTSs) were recently mapped in human cells and found to coincide with sites of Pol II pausing (Schwalb et al., 2016). Slow Pol II movement downstream of the pA site may also be required for efficient torpedo action, because termination shifts upstream or downstream when human Pol II elongation speed is decreased or increased, respectively (Fong et al., 2015), with the use of mutations characterized in yeast (Malagon et al., 2006; Hazelbaker et al., 2013). These results suggest that the nature of the transcribing Pol II complex is changed at the pA site, and this makes the enzyme less processive and prone to termination.

Several mechanisms may contribute to a slow down of Pol II downstream of the pA site. First, the pA site sequence may lead to a change in Pol II conformation. Second, elongation factors conferring processivity to Pol II may be lost upon pA site passage or may change their conformation. Third, factors may be recruited that destabilize transcribing Pol II, such as Pcf11 (Zhang and Gilmour, 2006). Pcf11 binds the flap loop on the wall of Pol II (Pearson and Moore, 2014) and cooperates with Rat1 during termination (Luo et al., 2006). Consistent with the context-dependent nature of termination, the Rat1 complex does not terminate *Escherichia coli* RNA polymerase (Park et al., 2015) and is unable to efficiently release Pol II from a reconstituted elongation complex in vitro (Dengl and Cramer, 2009), unless transcription is started from a promoter (Pearson and Moore, 2013) or paused by nucleotide misincorporation (Park et al., 2015), further consistent with the context-dependent nature of Rat1 action.

Despite all this work, we still do not know where in the yeast genome termination occurs and whether Pol II termination at protein-coding genes is generally terminated by the torpedo mechanism. To address these and other questions, we used various functional genomics techniques, as in our previous study of the global mechanism of Pol II termination at non-coding genes via the Nrd1 pathway (Schulz et al., 2013). Together with modeling, this allowed us to distinguish Pol II termination from RNA degradation downstream of the pA site and to provide evidence that Rat1 is generally required for Pol II termination at the end of protein-coding genes.

RESULTS

Transcription Termination Window in Yeast

In yeast, transcription termination has to occur shortly after the polymerase passes the pA site because the yeast genome is densely packed with transcription units. Non-coding transcription units are found within 500 base pairs (bp) downstream of most protein-coding genes. We selected for protein-coding genes that do not contain any transcribed units on the same strand within 500 bp downstream of the pA site. This led to a set of 2,501 transcribed protein-coding genes (ORF-Ts) out of a total of 4,928 units previously identified by TIF-seq (Pelechano et al., 2013) (Figure S1). These non-overlapping 2,501 ORF-Ts form a representative selection of genes with unaltered overall expression level and length distribution (Figure S1D). Consistent with a high gene density, it was initially found that Pol II is released over a range of around 200 nt downstream of the pA site at yeast genes and thus termination occurs close to the pA site in yeast (Birse et al., 1998; Dye and Proudfoot, 1999). However, the sites for Pol II termination in the yeast genome are unknown.

We therefore identified TTSs in the region following the pA site where Pol II release occurs (termination window, TW). To this end, we used sequencing data from newly synthesized RNA obtained after metabolic labeling of wild-type yeast with 4-thiouracil (4tU) (STAR Methods). In contrast to RNA-sequencing (seq) data, 4tU-seq data revealed short-lived RNA downstream of the pA site (Figure 1A). For TTS determination, we selected all ORF-Ts and detected TTSs by finding sites where coverage drops by fitting a piecewise constant curve to the coverage profiles (STAR Methods). For 903 selected genes with high 4tU-seq signal, this analysis revealed putative TTSs downstream of the last annotated pA site (Pelechano et al., 2014) as the sites where RNA coverage drops by at least 20% (Figure 1A and STAR Methods). When we considered one TTS per gene that showed the strongest coverage drop, 80% of genes contained the TTS within 200 bp downstream of the pA site (Figure 1B). We identified 1,177 TTSs with a median distance of 163 bp to the last annotated pA site (Figure 1D). More than 90% of the TTSs were located within 500 bp downstream of the pA site (Figure 1D). To confirm the TW length on the larger, previously selected set of 2,501 genes, we observed that the 4tU-seq signal dropped to baseline within ~180 bp downstream of the most frequent pA site (Figure S1B), confirming that termination occurs just downstream of the pA site in yeast.

Analysis of the sequences around the TW confirmed a previously noted A-U sequence bias around the pA site (Baejen et al., 2014; Martin et al., 2012) and further showed that the non-coding strand (DNA template strand) in the TW is pyrimidine-rich compared to gene bodies and is particularly rich in thymine residues (Figure 1C). As a consequence, the predicted stability of the DNA-RNA hybrid in the TW is lower than in gene bodies (Figure 1C). Low hybrid stability destabilizes the Pol II elongation complex (Sidorenkov et al., 1998) and renders it prone to termination. Consistent with this model, lower hybrid stability correlated with shorter TWs (Figure S1C).

Taken together, our analysis showed that TTSs for protein-coding genes in yeast are generally located near the pA site

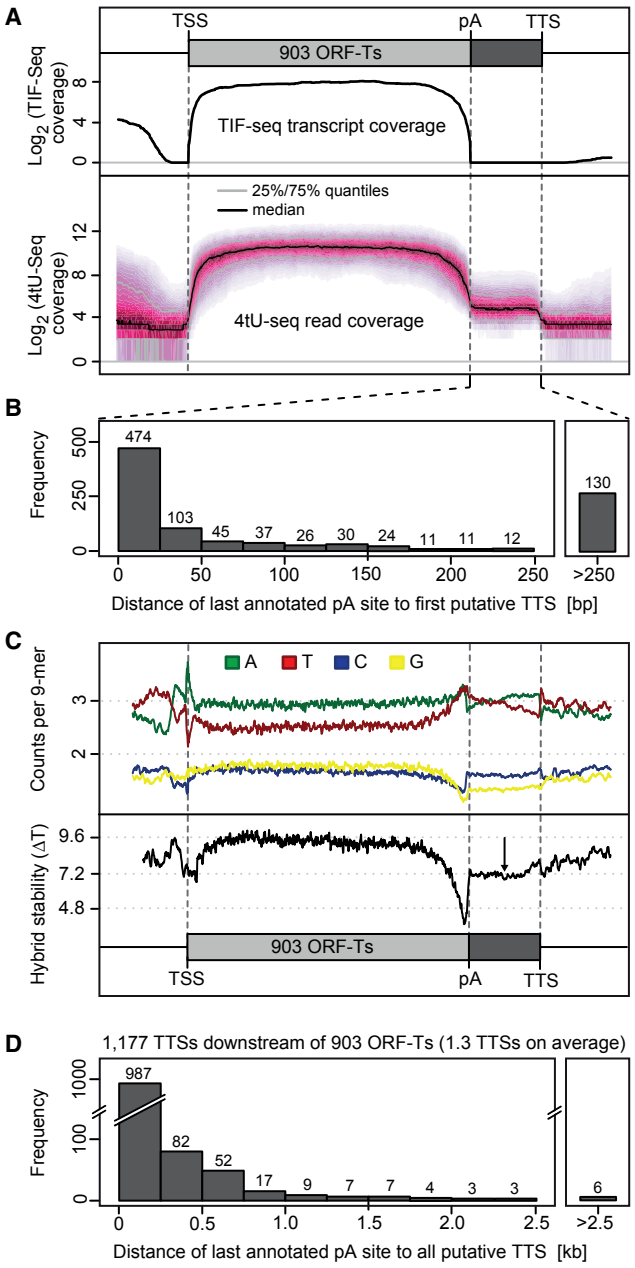


Figure 1. Analysis of the Termination Window

(A) TIF-seq transcript coverage over 903 selected ORF-Ts resulting from transcript isoforms aligned and scaled to the first TSS and last pA of each ORF-T (light gray bar, top) and to the first putative TTS (top), compared to the unique 4tU-seq coverage of wild-type yeast over 25% and 75% quantiles (bottom).

(B) Distribution of estimated first putative TTSs relative to the last annotated pA site for the 903 investigated ORF-Ts (bp).

(C) Mean number of A/C/G/T per 9-mer in the 903 selected ORF-Ts (top) and the predicted stability of the DNA-RNA hybrid (bottom).

(D) Distribution of all determined TTSs relative to the last annotated pA site for the 903 investigated ORF-Ts (kb).

See also Figure S1.

and the TW is usually not larger than a few hundred bp and contains thymine-rich template sequences that destabilize the Pol II elongation complex. The TW in yeast was thus much narrower than in the human genome, where it was estimated to be ~3,300 bp on average (Schwalb et al., 2016) (Figures 1B and S1B). In contrast to human TTSs, which often contain a C-rich motif (Schwalb et al., 2016), the yeast-derived TTSs did not show a sequence motif, resembling the situation with pA sites in yeast, which also do not show a conserved motif.

Differential RNA Interactions of Termination Factors

Having defined the genomic regions where termination occurs, we turned to a genome-wide localization and functional analysis of the key protein factors implicated in termination. Because well over 20 polypeptides are involved in the 3′-transition, we had to focus our analysis on selected factors that reside in the various required multiprotein complexes. We included in our analysis the CFIA subunit Pcf11 and the endonuclease Ysh1, the elongation factor Spt5 and its binding partner Spt4, the torpedo nuclease Rat1, and its interacting factors Rai1 and Rtt103. We concentrated on this selection of key factors involved in the 3′-transition that reside in the various multiprotein complexes required for 3′-RNA processing and transcription termination.

We located factors along the genome with the use of chromatin immunoprecipitation (ChIP)-seq (Figures 2 and S2). We also mapped the binding sites of factors over the transcriptome with the use of photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) (Figures S2B and S2C), a method established in human cells (Hafner et al., 2010) and later adapted to map in vivo protein-RNA interaction sites in yeast (Creamer et al., 2011; Schulz et al., 2013) (STAR Methods). Note that PAR-CLIP reveals direct protein-RNA interactions that are obtained without a crosslinking agent, whereas ChIP reveals factor occupancies on the genome that result after the use of a crosslinking agent and can be indirect; i.e., via other proteins that bind DNA. In particular, proteins that travel with Pol II along genes may produce a ChIP signal via Pol II although they are not near RNA.

All factors showed strong PAR-CLIP signals, leading to smooth metagene profiles (Figure 2, blue traces). The PAR-CLIP data did not provide evidence that any of the investigated factors bind a defined RNA sequence motif. Pcf11 showed strong ChIP and PAR-CLIP peaks in the TW at ~50 bp and ~130 bp downstream of the pA site, respectively (Figure 2). These peaks are consistent with initial recruitment of Pcf11 by the Ser2-phosphorylated Pol II CTD (Licatalosi et al., 2002), followed by additional binding of Pcf11 to nascent RNA downstream of the pA site. Pcf11 bound preferentially in a region downstream of the pA site, which is the part of the RNA that is rapidly degraded following pA site cleavage.

Comparison of the ChIP and CLIP profiles revealed differences for several factors that relate to factor-specific behavior in their binding to the transcribing complex and RNA. In particular, Spt5 showed different ChIP and PAR-CLIP profiles. ChIP-seq revealed the presence of Spt5 within gene bodies as described (Mayer et al., 2010), consistent with direct interaction of Spt5 with Pol II (Martinez-Rucobo et al., 2011). In contrast, PAR-CLIP revealed a peak in the TW ~40 nt downstream of

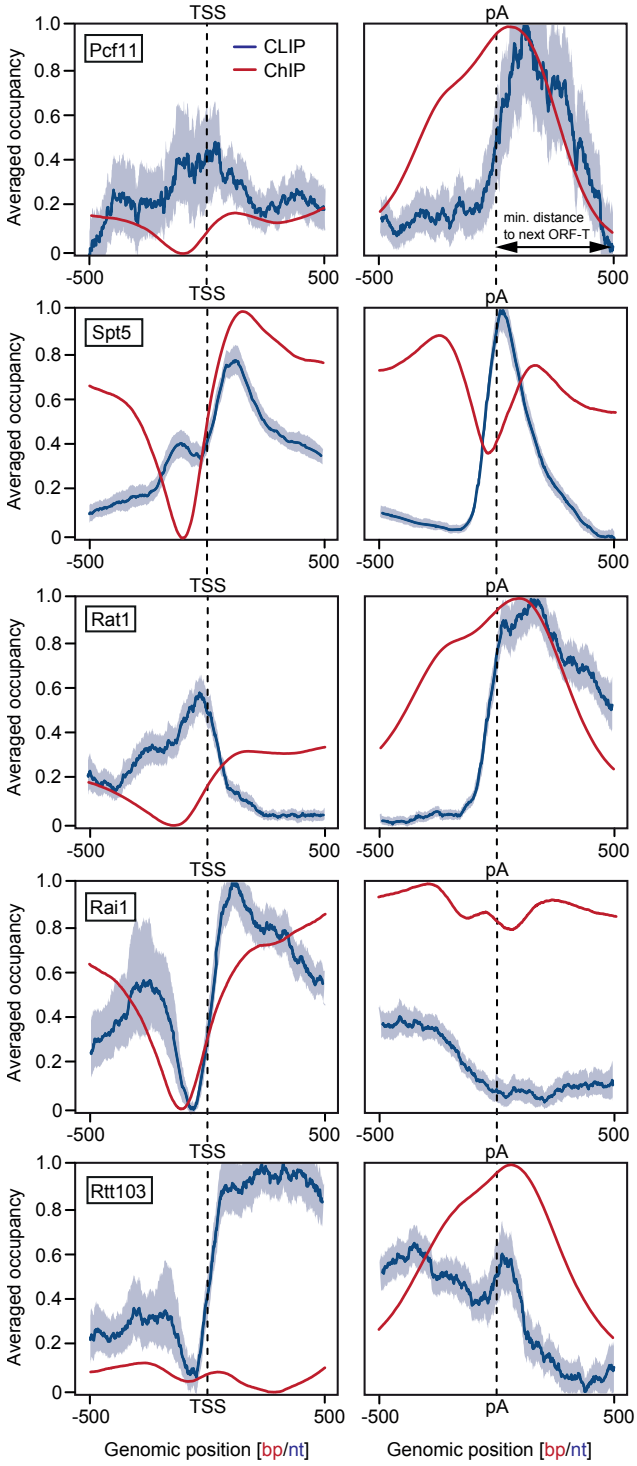


Figure 2. Factors Involved in the 3′-Transition Differ in DNA and RNA Occupancy

Averaged and locus-normalized occupancy profiles obtained from PAR-CLIP (blue) and ChIP-seq (red) experiments of (from top to bottom) Pcf11, Spt5, Rat1, Rai1, and Rtt103. The profiles are aligned at both the TSS and cleavage and pA site of 2,501 ORF-Ts with a minimal distance of 500 bp to the next genomic feature. The shaded areas around the PAR-CLIP traces give symmetric 95% confidence intervals around the median value (blue trace), computed using a bootstrap procedure. See also Figure S2.

the pA site (Figure 2), suggesting that Spt5 interacts more intimately with RNA when Pol II passes over the pA site. To correct for different RNA abundance, we normalized PAR-CLIP data with RNA-seq data. This can lead to artificially high peaks in regions where RNA is unstable and thus less well detected by RNA-seq, in particular downstream of the pA site. We have therefore explored an alternative normalization of the data with the PAR-CLIP signal of Pol II subunit Rpb1 and confirmed the Spt5 peak (Figure S2D).

With respect to Rat1, the ChIP profile showed the presence of the nuclease within gene bodies, but the PAR-CLIP profile showed a peak ~150 nt downstream of the pA site, matching the small width of the TW defined above (Figures 1B and 2). These results are consistent with a model that Rat1 is generally close to the Pol II elongation complex, but only contacts RNA once a free 5′ end is produced after pA cleavage. These results are consistent with the idea that pA site cleavage provides an entry point for Rat1 to degrade RNA from the newly generated 5′ end, and that Rat1 acts as a 5′-3′ torpedo exonuclease in pA-dependent Pol II termination.

The occupancy profiles for Rai1 and Rtt103 were clearly distinct from those of Rat1. PAR-CLIP signals for Rai1 peaked ~100 nt downstream of the transcription start site (TSS), consistent with a role of Rai1 in cap quality control (Jiao et al., 2010). Rtt103 showed PAR-CLIP signals over the entire transcript, including the TW (Figures 2, S2B, and S2C). Although the signals in the transcript body are clearly above noise, the top PAR-CLIP sites are found in the region just downstream of the pA site (Figure S2C). This is consistent with a role of Rtt103 in stabilizing the Rat1-Rai1-Rtt103 complex on transcribed genes. The increasing ChIP signal for Rtt103 at the end of the transcription unit is further consistent with binding of the Rtt103 CTD-interacting domain to Pol II phosphorylated at Ser2 residues (Kim et al., 2004). These data indicate that the Rat1-Rai1-Rtt103 complex can be recruited throughout the transcribed region, but that Rai1 and Rat1 preferentially contact RNA downstream of the TSS and the pA site, respectively.

Pcf11 and Ysh1 Are Required for the 3′-Transition

We next probed the effect of conditional depletion of termination-related protein factors on genome-wide RNA synthesis. We carried out 4tU-seq (Schulz et al., 2013; Sun et al., 2012) before and after nuclear depletion of factors with the anchor-away (AA) technique (Haruki et al., 2008). High correlations between biological replicates demonstrated the reproducibility of the 4tU-seq experiments (Figure S3), and attenuation of cell growth showed the efficiency of AA depletion (Figures S3B, S3C, and STAR Methods).

We first analyzed Pcf11, which is required for the recognition of the pA site, and Ysh1, which is the endonuclease that cleaves nascent RNA. Examination of the 4tU-seq data revealed very drastic changes upon depletion of these factors (Figures 3A and 3B). In particular, RNA signals did not drop at the end of

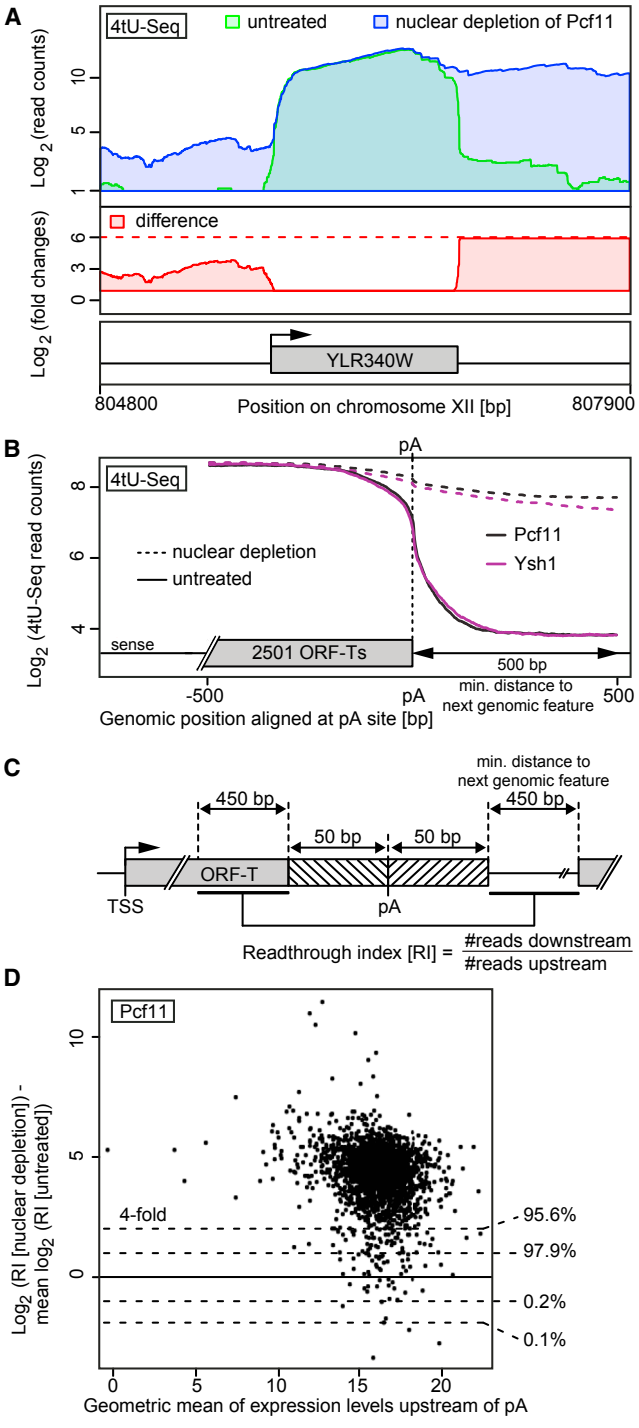


Figure 3. Nuclear Depletion of Pcf11 Abolishes the 3′-Transition

(A) Genome browser view of $\log_2$ read counts measured by 4tU-seq before (green) and after (blue) nuclear depletion of Pcf11, around ORF-T YLR340W. The resulting differences are depicted as fold change (red).

(B) Sense strand 4tU-seq signals ($\log_2$ median position-based read count, quantile normalized) of 2,501 ORF-Ts with a minimal distance of 500 bp to the next genomic feature, aligned at their pA site, before (solid line) and after (dashed line) nuclear depletion of Pcf11 (black) and Ysh1 (purple).

genes, but rather continued into regions that are normally not transcribed, suggesting a dramatic termination defect. To examine the effects globally, we selected 2,501 genes with a minimal distance of 500 bp to the next genomic feature and plotted the metagene signals around the pA site before and after nuclear depletion of Pcf11 or Ysh1 (Figure 3B). We then calculated a "readthrough index" (RI) for each depleted factor at each gene (Figure 3C). The RI measures the ratio between the number of read counts 500 bp downstream and 500 bp upstream of the pA site, excluding a region 50 bp up- and downstream of the pA site (STAR Methods). Depletion of Pcf11 or Ysh1 changed the RI by at least a factor of four for ~96% or ~91% of genes, respectively (Figure 3D).

These data are consistent with a global loss of the 3′-transition. They are also consistent with a global requirement of Pcf11 and Ysh1 for pA site recognition and pA-dependent RNA cleavage, as shown at individual genes (Amrani et al., 1997; Birse et al., 1998; Chanfreau et al., 1996; Garas et al., 2008). The data also indicate that Pcf11 and Ysh1 are required for subsequent termination and RNA degradation downstream of the pA site, consistent with the observation that Pol II release requires Ysh1 (Schaughency et al., 2014). Taken together, nascent RNAs containing a pA site must be recognized and cleaved to achieve a successful 3′-transition in vivo.

Spt5 Functions in the 3′-Transition

We had previously shown that the C-terminal region of Spt5 contributes to the recruitment of CFIA (Mayer et al., 2012b). To investigate the functional role of Spt5 in the 3′-transition, we performed 4tU-seq before and after nuclear depletion of Spt5 (Figures 4B and S4). High correlations between biological replicates demonstrated the reproducibility of the experiments (Figure S4B), and a dramatic drop in cell growth showed the efficiency of the AA approach (Figure S4C and STAR Methods). Nuclear depletion of Spt5 led to an almost complete loss of Spt5 occupancy at two tested genes in vivo (Figure S4A). Depletion of Spt5 caused an increase in RNA reads downstream of the pA site and an at least 2-fold increase of the RI within 41% of the selected genes (Figures 4B and 4C). Depletion of the Spt5-associated protein Spt4 had no effect on the 4tU-seq metagene profiles (Figure 4B) and caused lower change in RI (Figure S4D), consistent with the non-essential nature of this protein (Hartzog et al., 1998).

(C) We defined a RI as the ratio of normalized read counts falling into the (pA+50 bp and pA+500 bp) region over normalized read counts falling into the (pA−500 bp and pA−50 bp) region. Excluding a region of 50 bp up- and downstream of the pA site increased the robustness of our analysis. We then compared the RIs of samples with nuclear depletion (or mutants) and the untreated ones (or wild-type) to characterize the amount of termination defects.

(D) Normalized log readthrough index ($\log_2$ RI[nuclear depletion] − mean $\log_2$ RI[untreated]) after nuclear depletion of Pcf11 versus the geometric mean of expression levels upstream (−500 to −50 bp) of the pA site in both samples. A subset of 2,501 ORF-Ts that have no annotated feature 500 bp downstream of their pA site was used. The indicated percentages facing the dashed horizontal lines correspond to the proportion of ORF-Ts (out of 2,501) that are above or below a 2-fold or 4-fold change in RI.

See also Figure S3.

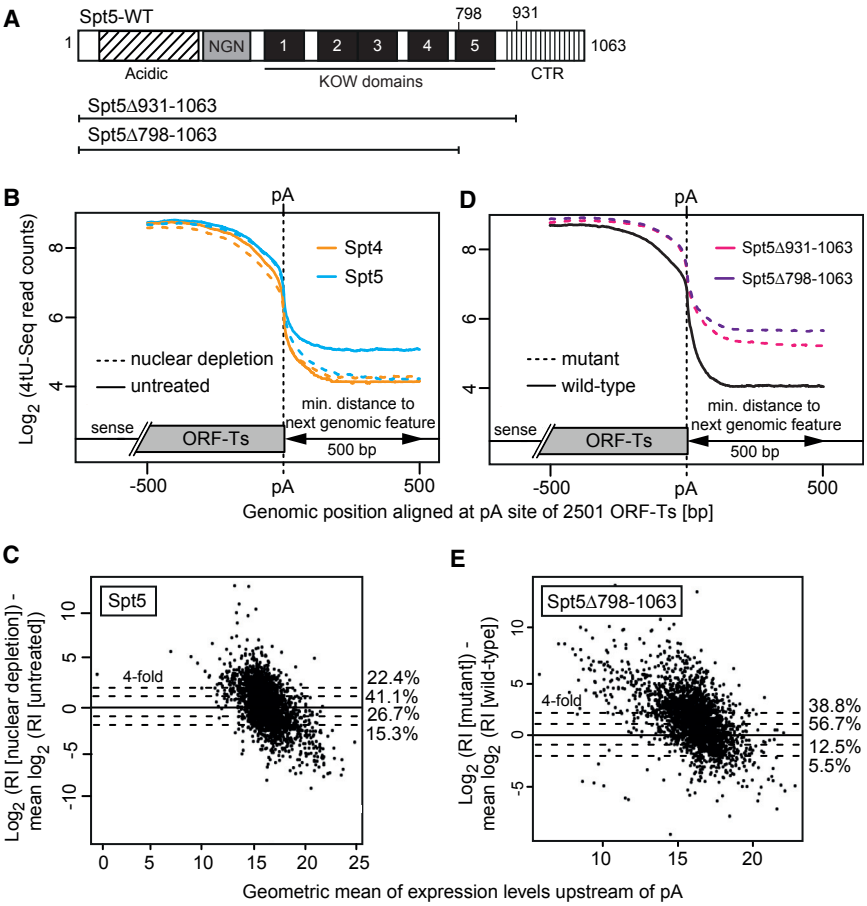


Figure 4. Spt5 Is Involved in the 3′-Transition

(A) Domain architecture of the Spt5 protein and truncated Spt5 variants in mutant strains.

(B) 4tU-seq metagene profiles (log_2 median position-based read count, quantile normalized) of 2,501 ORF-Ts with a minimal distance of 500 bp to the next genomic feature, aligned at the pA site, before (solid lines) and after (dashed lines) nuclear depletion of Spt4 (orange) or Spt5 (blue).

(C) Normalized log RI after nuclear depletion of Spt5 versus the geometric mean of expression levels upstream (−500 to −50 bp) of the pA site.

(D) 4tU-seq metagene profiles as in (B) for the wild-type (solid line) and the mutant strains (dashed lines) Spt5Δ931–1,063 (magenta) and Spt5Δ798–1,063 (purple).

(E) Fold changes of RI as in (D), but with the log_2 RI ratio of the Spt5Δ798–1,063 mutant and the wild-type.

See also Figure S4.

To test whether the effect of Spt5 depletion was due to the function of the Spt5 C-terminal region, we also carried out 4tU-seq with a mutant yeast strain that lacked this region (Spt5Δ931–1,063) and a strain that additionally lacked the neighboring fifth KOW domain (Spt5Δ798–1,063) (Figures 4A, 4D, and S4B). Spt5 depletion had a global negative effect on transcription. In both strains, RNA accumulated downstream of the pA site, with Spt5Δ798–1,063 having an even stronger defect in the 3′-transition (Figures 4D, 4E, S4E, and S5). These results show that Spt5, and in particular its C-terminal region and the fifth KOW domain, are functionally required for the 3′-transition. This function could be the recruitment of CFIA as predicted (Mayer et al., 2012b), but may also include direct interaction of Spt5 with RNA, since Spt5 strongly crosslinks to RNA in the TW (Figures 2 and S2).

Rat1 Exonuclease Is Globally Required for Pol II Release

The above results showed that the 3′-transition leads to a smoothly descending 4tU-seq metagene signal in the TW that is caused by Pol II terminating mainly within 100–200 bp downstream of the pA site. When the 3′-transition is perturbed, characteristic changes of this metagene signal are observed. When pA site recognition or RNA cleavage are defective, a continuous, non-descending RNA signal is obtained. When Spt5 is mutated, the smoothly descending signal is preserved, indicating that termination still occurs, but a higher RNA signal remains downstream of the TW, indicating that a fraction of polymerases are not terminated due to a partial defect in pA site recognition and/or RNA cleavage.

When we depleted the Rat1 exonuclease from the nucleus, a very different shape of the 4tU-seq metagene profile was observed (Figure 5A). Rat1 depletion resulted in a kink exactly at the pA site, instead of a smooth, descending curve (Figure 5A). The lack of the smooth descending curve indicates that Pol II does not terminate. Nuclear depletion of Rat1 also led to reduced Rat1 occupancy at two tested genes in vivo (Figure S2D), providing independent evidence that Rat1 is depleted from the nucleus under the experimental conditions. These observations strongly suggested that Rat1 depletion specifically disrupts Pol II termination.

To confirm that Rat1 depletion really causes a termination defect, we mapped Pol II occupancy by ChIP-seq before and after nuclear depletion of Rat1 (Figures 5B and S5A). Whereas Pol II occupancy decreased in the TW downstream of the pA site under normal conditions as expected, Rat1 depletion led to increased Pol II occupancy in this region, confirming that Rat1 is required for Pol II termination. Pol II occupancy is higher in the TW than in gene bodies, consistent with the reduced speed of Pol II and a resulting higher crosslinkability downstream of the pA site.

The Rat1-specific change in the 4tU-seq metagene profile was not observed when the Rat1-associated factor Rai1 was depleted from the nucleus (Figure 5A). Depletion of Rai1 resulted again in the typical, smoothly descending curve with increased RNA levels. This is consistent with the role of Rai1 in stimulating the exoribonuclease activity of Rat1 (Xiang et al., 2009; Xue et al., 2000) and indicates that Rai1 is not essential for Pol II termination per se. Nuclear depletion of Rat1 and Rai1 affected a similar number of genes, with good overlap between affected genes (Figures 5C

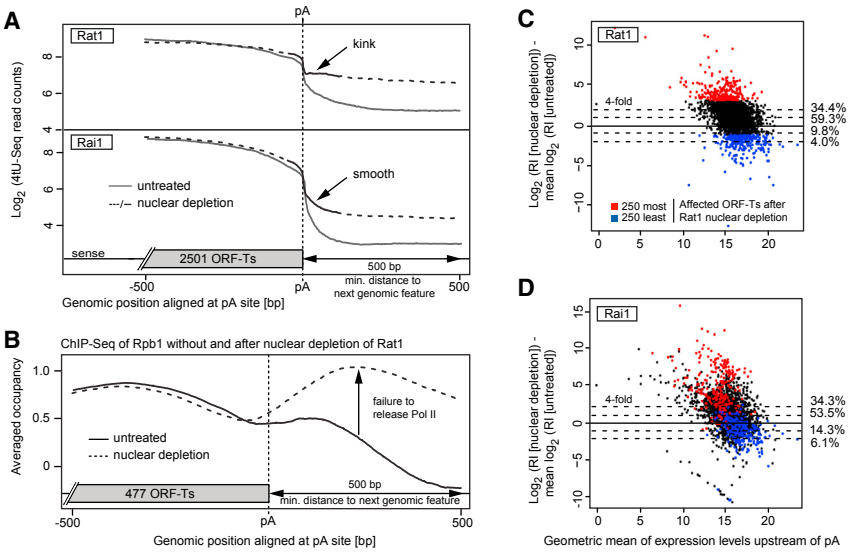


Figure 5. Rat1 Terminates Transcription

(A) Changes in 4tU-seq metagene profiles upon nuclear depletion of Rat1 (top) or its binding partner Rai1 (bottom). Plotted are the $\log_2$ median position-based read counts of 2,501 ORF-Ts with a minimal distance of 500 bp to the next genomic feature, aligned at the pA site.

(B) Averaged occupancy profiles of Rpb1 (Pol II) obtained from ChIP-seq experiments before (solid line) and after (dashed line) nuclear depletion of Rat1. The profiles are aligned at the pA site of 477 ORF-Ts, with a minimal distance of 500 bp downstream to the next genomic feature on both the sense and antisense strand. This is necessary to avoid spillover effects because ChIP is not strand specific.

(C) Normalized $\log_2$ RI before and after nuclear depletion of Rat1 versus geometric mean of expression levels upstream of the pA site. The 250 most and least affected ORF-Ts after nuclear depletion of Rat1 are in red and blue, respectively.

(D) As in (C), but for nuclear depletion of Rai1.

See also Figure S5.

and 5D), reflecting that these proteins reside in the same biochemical complex. Taken together, in the absence of Rat1, termination fails and Pol II accumulates in the TW, whereas in the absence of Rai1, RNA degradation is slower, but termination still occurs.

Modeling Supports Interpretation of the 3′-Transition

To better rationalize the 3′-transition and distinguish effects of pA site recognition, RNA cleavage, Pol II termination, and RNA degradation, we mathematically modeled how these different events contribute to the shape of the 4tU-seq metagene profiles downstream of the pA site and how various defects change this shape (Figure S6). For a description of our model, please refer to STAR Methods, and for details, please refer to the R script provided for downloading at http://www.mpibpc.mpg.de/15585287/model.zip. Briefly, we model the amount of RNA at a certain position downstream of the pA site. The model takes into account the fraction of pre-mRNA that is cleaved at the pA site, the lifetime of RNA downstream of the pA site, and the efficiency of polymerase termination downstream of the pA site.

The model can rationalize the observed 4tU-seq phenotypes and supports our mechanistic interpretations of changes in the 4tU-seq metagene profiles upon factor depletions. By comparing 4tU-seq experimental data with the theoretical curve shapes obtained from modeling (Figures 6A and S6A), we found that the modeled curve resembled the in vivo data very closely (compare Figures 6A with 3B, 4B, and 5A). This strongly supports our interpretations of changes in metagene profiles after perturbation of various parts of the 3′-transition (Figures 6B and S6B). The modeling was also consistent with our interpretation that the smoothly descending signal indicated termination, and that its loss upon depletion of the exonuclease Rat1 is due to a defect in Pol II termination.

Rat1 and Pcf11 Prevent Deregulation of Downstream Transcription

Finally, we investigated whether defects in the 3′-transition lead to deregulation of transcription at downstream genes by read-through transcription and transcription interference (Figure S7) (Proudfoot, 1986). To measure how defective termination deregulates RNA synthesis at downstream protein-coding genes, we compared the absolute change in length-normalized 4tU-seq read count level within the intergenic region following nuclear depletion (Δy) with the absolute change in length-normalized read count level within the next downstream protein-coding gene (Δz) for a set of 1,632 genes arranged in "tandem" and separated by at most 500 bp (Figure 7A). Note that the observed 4tU-seq signal within the downstream coding gene is a result of both, read-through transcription of the upstream gene and transcription interference; i.e., suppression of transcription of the downstream gene.

We found that the change in 4tU-seq signal at the downstream genes Δz was proportional to the change in the intergenic transcript level Δy (Figures 7B–7D and S7B). A linear regression analysis allowed us to estimate that nuclear depletion of Pcf11 or Ysh1 results in 72% and 81% of non-terminated polymerases that transcribe into downstream genes, respectively (Figures 7B and 7C). The dependence is very strong despite the noise associated with measuring the low levels of readthrough y, as the change in readthrough level Δy explains 65% and 87% (equal to the Pearson correlation) of the total variance of Δz for the knock down of Pcf11 and Ysh1, respectively.

This is consistent with our interpretation that Pcf11 is necessary for pA site recognition and usage, and that the resulting RNA cleavage by Ysh1 endonuclease is essential to render Pol II prone to termination (Figure 7C). After nuclear depletion of Rat1, only around 35% of the non-terminated Pol II enzymes transcribed into downstream genes, and the remaining polymerases were terminated around the downstream gene promoter (Figure 7D). Knock down of Rat1 leads to only weak correlation (0.27) and knock down of Rai1 resulted in a vanishing correlation (Figure S7B). These results demonstrate that only knock down of Ysh1 and Pcf11 and, to a lesser extent of Rat1, lead to significant changes in RNA synthesis at the downstream ORF-Ts and are consistent with a back-up termination mechanism that

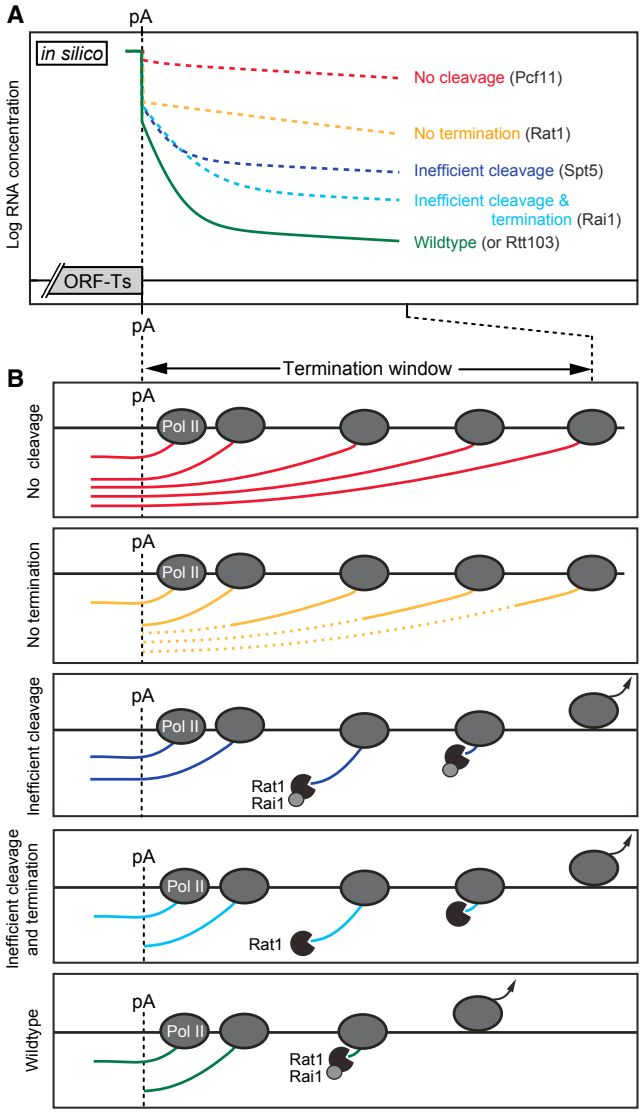


Figure 6. Modeling the 3′-Transition Recapitulates 4tU-Seq Metagene Profile Changes

(A) Mathematically modeled theoretical changes in the logarithm of RNA concentration downstream of the pA site, averaged over many genes. The modeling is based on molecular events including RNA 3′-cleavage, RNA degradation, and Pol II termination (as described in Figure S7 and model.R, http://www.mpibpc.mpg.de/15585287/model.zip).

(B) Cartoon representation of molecular phenotypes of perturbed 3′-transition rationalize changes downstream of the pA site that give rise to specific changes in the RNA metagene curves computed in (A).

See also Figure S6.

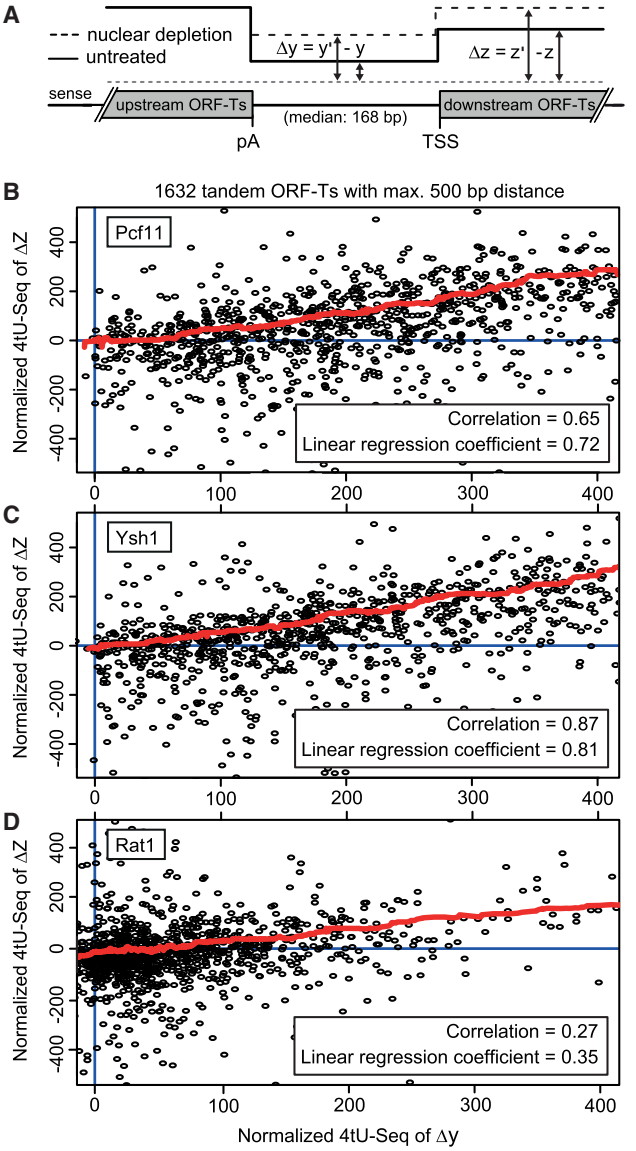


Figure 7. Transcription into Downstream Genes

(A) Schematic representation of how perturbation of the 3′-transition influences the RNA signal in the region downstream of the pA site ($y \rightarrow y' = y+\Delta y$) and the downstream gene ($z \rightarrow z' = z+\Delta z$).

(B) The change between normalized 4tU-seq read counts of the intergenic region before and after nuclear depletion of Pcf11 (Δy) is plotted versus the change in normalized read counts in the downstream ORF-T (Δz). The red line represents the running median over 1,632 tandem ORF-Ts with a median and maximum intergenic distance of 168 bp and 500 bp, respectively.

(C) As in (B), but for nuclear depletion of the endonuclease Ysh1.

(D) As in (B), but for nuclear depletion of the exonuclease Rat1.

See also Figure S7.

attenuates transcription interference (Rondón et al., 2009), maybe as a result of Pol II bumping into DNA-binding factors near the downstream promoter (Colin et al., 2014).

DISCUSSION

Here, we used a combination of functional genomics techniques to analyze the 3′-transition during Pol II transcription of protein-coding genes in yeast. Such a genome-wide in vivo analysis was still lacking, most likely because the 3′-transition of Pol II is difficult to dissect due to the intimate connections between pA site recognition, RNA cleavage, elongation complex changes, RNA degradation, and Pol II termination. Here, we could deconvolute effects on the 3′-transition that result from

various perturbations and could recapitulate the experimental data with a model that describes the 3′-transition as a series of defined molecular events. An advantage of the used anchor-away approach coupled to global analysis is that weak effects can be clearly detected after metagene analysis, but may be hidden when individual genes are analyzed due to noise or gene-specific behavior.

The main findings from this work are the following. First, we found a narrow termination window downstream of the pA site, which matches the high gene density of yeast. Second, we showed that DNA- and RNA-binding profiles of termination-related factors could differ strongly, suggesting rearrangements of factors during the transcription cycle. Third, we showed that the CFIA subunit Pcf11 and the CPF subunit Ysh1 are globally required for the 3′-transition. Fourth, we found that Spt5 has a function in the 3′-transition. Fifth, we observed that the Rat1 exonuclease generally terminates transcription, whereas its partner Rai1 is not required for termination per se, but for efficient RNA degradation. Finally, we showed that an effective 3′-transition is essential to prevent widespread deregulation of transcription at downstream genes and obtained results consistent with a back-up termination mechanism that limits the amount of such deregulation.

Our global analysis converges with a very large body of existing work on a general scheme for the 3′-transition at protein-coding genes. There were two different models for Pol II termination that were suggested, the allosteric model and the torpedo model. In the allosteric model, the binding of RNA 3′-processing factors induces structural rearrangements in the Pol II elongation complex that lead to termination (Logan et al., 1987). In the torpedo model, a nuclease degrades nascent RNA from its 5′ end, catches up with the elongation complex, and displaces Pol II from DNA (Connelly and Manley, 1988; Proudfoot, 1989). A combined model was also suggested that assumes both a change in elongating Pol II that reverses anti-termination effects and the occurring torpedo action (Luo et al., 2006; Schreieck et al., 2014). Our data show that the torpedo nuclease Rat1 is generally required for termination and also provide additional evidence for changes in the elongation complex as Pol II runs over the pA site, arguing for the combined model.

We note that some published results obtained in in vitro assays do not match our in vivo findings. In vitro, RNA cleavage is not strictly required for termination (Zhang et al., 2015), and the Rat1-Rai1-Rtt103 complex only partially removes Pol II from a DNA-RNA scaffold mimicking the nucleic acids in an elongation complex (Dengl and Cramer, 2009; Luo et al., 2006). Such in vitro assays depend on design, because more efficient termination by the Rat1-Rai1 complex was achieved in another in vitro system that used Pol II initiated from a promoter (Pearson and Moore, 2013) or Pol II that was paused after nucleotide misincorporation (Park et al., 2015). Thus, current in vitro systems can recapitulate some of the aspects of the mechanism, including the stimulation of the enzymatic activity of Rat1 by Rai1 (Dengl and Cramer, 2009) that we confirmed here in vivo, but these systems can currently not recapitulate the complex nature of the coupled 3′-transition.

Due to the current limitations in studying the 3′-transition in vitro, structural knowledge is limited and many of the mechanistic aspects remain to be explored. We envision that the mechanism underlying the 3′-transition may be tackled by focusing on three specific questions. First, how are CPF and CFIA assembled on pre-mRNA to recognize the pA site and provoke RNA cleavage? Second, how is the elongating Pol II complex changed after its passage over the pA site to render it prone to termination, and what are the roles of CFIA and Spt5 in this process? Third, what is the state and conformation of Pol II at the termination site just before it is released from DNA and RNA, and how do termination factors provoke Pol II release? To address these questions, large and transient macromolecular complexes must be analyzed biochemically and structurally.

The role of Spt5 in the 3′-transition is of particular interest, because Spt5 is the only polymerase-associated factor that is conserved in archaea and bacteria. The Spt5 counterparts in archaea and bacteria (NusG) suppress polymerase pausing and stimulate elongation (Herbert et al., 2010; Hirtreiter et al., 2010). Alteration or release of Spt5 is thus very likely a prerequisite for termination. In the bacterial system, NusG interacts with the termination factor Rho (Burmann et al., 2010; Cardinale et al., 2008; Chalissery et al., 2011; Peters et al., 2012). Although there is no homolog of Rho in yeast, Spt5 may similarly interact with factors required for Pol II termination. We show that yeast Spt5 binds RNA in vivo, consistent with results in *Drosophila* (Missra and Gilmour, 2010). RNA binding by Spt5 occurs in the TW, near the CFIA subunit Rna15 that binds ~20 nt downstream of the pA site (Baejen et al., 2014; Mayer et al., 2012b), and thus Spt5 may interact with CFIA on RNA.

Finally, it is likely that the 3′-transition of yeast Pol II investigated here is a good model for understanding the 3′-transition of human Pol II. First, most factors involved in the 3′-transition are highly conserved in the yeast and human systems. Second, consistent with our results, knock down of CPSF73 (counterpart of Ysh1) or CstF64 (counterpart of CFIA subunit Rna15) in human cells impairs termination (Nojima et al., 2015). Third, our results agree with a global termination defect observed by ChIP-seq of human Pol II after catalytic inactivation of XRN2, the human counterpart of Rat1 (Fong et al., 2015). A strong difference between the human and yeast systems is, however, the width of the TW. We suggest that the TW for human Pol II is much wider (Schwalb et al., 2016), to respond to the need for more elaborate cotranscriptional RNA processing. Termination of human Pol II further downstream of the pA site could open a time window to complete RNA processing before transcription termination.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Rapamycin-Dependent Depletion of Yeast Proteins from the Nucleus
  - PAR-CLIP of Yeast Proteins
  - ChIP-Seq and ChIP-qPCR Experiments

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2017.02.009.

AUTHOR CONTRIBUTIONS

C.B. and P.C. designed the study. C.B. planned and coordinated experiments. K.C.M., S.E., and P.R. cloned strains and performed anchor-away and 4tU-seq experiments. K.C.M., P.R., and C.B. carried out growth rate experiments. A.B. and C.B. performed PAR-CLIP experiments. S.B. performed ChIP-seq experiments. K.C.M. performed ChIP-qPCR experiments. J.A., P.T., B.S., J.S., C.B., and P.C. planned data analyses. J.A. and B.S. carried out 4tU-seq data analyses. J.A. carried out ChIP-seq data analyses. P.T., C.B., and M.L. carried out PAR-CLIP data analyses. J.S. supervised data analysis and modeling. P.C. supervised research. C.B. and P.C. wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

We thank Helmut Blum, Stefan Krebs, and Alexander Graf (LAFUGA, Gene Center Munich) for Illumina sequencing. We thank Mai Sun (EMBL, Heidelberg) for cloned FRB strains and Amelie Schreieck for Spt5 mutant strains. J.A. was supported by a Fellowship for Postdoctoral Researchers from the Alexander von Humboldt Foundation. J.S. was supported by the Bavarian Center for Molecular Biosystems (BioSysNet grant), by grant e:BIO-M1-115 from the German Ministry for Education and Research (BMBF), and by the Deutsche Forschungsgemeinschaft (SFB646). M.L. was supported by the Center for Innovative Medicine (CIMED) at Karolinska Institutet and by the Science for Life Laboratory (SciLifeLab) in Stockholm. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860), the Advanced Grants "TRANSIT" and "TRANSREGULON" of the European Research Council, the Volkswagen Foundation, the Center for Innovative Medicine (CIMED) at Karolinska Institutet, and the Science for Life Laboratory (SciLifeLab) in Stockholm.

Received: June 21, 2016
Revised: October 6, 2016
Accepted: February 9, 2017
Published: March 16, 2017

REFERENCES

Amrani, N., Minet, M., Wyers, F., Dufour, M.E., Aggerbeck, L.P., and Lacroute, F. (1997). PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. Mol. Cell. Biol. *17*, 1102–1109.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Baejen, C., Torkler, P., Gressel, S., Essig, K., Söding, J., and Cramer, P. (2014). Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. Mol. Cell *55*, 745–757.

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. Science *280*, 298–301.

Burmann, B.M., Schweimer, K., Luo, X., Wahl, M.C., Stitt, B.L., Gottesman, M.E., and Rösch, P. (2010). A NusE:NusG complex links transcription and translation. Science *328*, 501–504.

Cardinale, C.J., Washburn, R.S., Tadigotla, V.R., Brown, L.M., Gottesman, M.E., and Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. Science *320*, 935–938.

Chalissery, J., Muteeb, G., Kalarickal, N.C., Mohan, S., Jisha, V., and Sen, R. (2011). Interaction surface of the transcription terminator Rho required to form a complex with the C-terminal domain of the antiterminator NusG. J. Mol. Biol. *405*, 49–64.

Chanfreau, G., Noble, S.M., and Guthrie, C. (1996). Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). Science *274*, 1511–1514.

Colin, J., Candelli, T., Porrua, O., Boulay, J., Zhu, C., Lacroute, F., Steinmetz, L.M., and Libri, D. (2014). Roadblock termination by reb1p restricts cryptic and readthrough transcription. Mol. Cell *56*, 667–680.

Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. *2*, 440–452.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. *7*, e1002329.

Dengl, S., and Cramer, P. (2009). Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. J. Biol. Chem. *284*, 21270–21279.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dye, M.J., and Proudfoot, N.J. (1999). Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Mol. Cell *3*, 371–378.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. Mol. Cell *60*, 256–267.

Garas, M., Dichtl, B., and Keller, W. (2008). The role of the putative 3′ end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. RNA *14*, 2671–2684.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol. Cell. Biol. *26*, 3986–3996.

Grosso, A.R., de Almeida, S.F., Braga, J., and Carmo-Fonseca, M. (2012). Dynamic transitions in RNA polymerase II density profiles during transcription termination. Genome Res. *22*, 1447–1456.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev. *12*, 357–369.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell *31*, 925–932.

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. Mol. Cell *49*, 55–66.

Herbert, K.M., Zhou, J., Mooney, R.A., Porta, A.L., Landick, R., and Block, S.M. (2010). E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. J. Mol. Biol. *399*, 17–30.

Hirtreiter, A., Damsma, G.E., Cheung, A.C., Klose, D., Grohmann, D., Vojnic, E., Martin, A.C., Cramer, P., and Werner, F. (2010). Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. Nucleic Acids Res. *38*, 4040–4051.

Hyman, L.E., and Moore, C.L. (1993). Termination and pausing of RNA polymerase II downstream of yeast polyadenylation sites. Mol. Cell. Biol. *13*, 5159–5167.

Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. J. Comput. Graph. Stat. *5*, 299–314.

Jiao, X., Xiang, S., Oh, C., Martin, C.E., Tong, L., and Kiledjian, M. (2010). Identification of a quality-control mechanism for mRNA 5′-end capping. Nature *467*, 608–611.

Kazerouninia, A., Ngo, B., and Martinson, H.G. (2010). Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro. RNA *16*, 197–210.

Keller, W., and Minvielle-Sebastia, L. (1997). A comparison of mammalian and yeast pre-mRNA 3′-end processing. Curr. Opin. Cell Biol. *9*, 329–336.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. *14*, 2452–2460.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science *332*, 475–478.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II. Mol. Cell *9*, 1101–1111.

Lidschreiber, M., Leike, K., and Cramer, P. (2013). Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. Mol. Cell. Biol. *33*, 3805–3816.

Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc. Natl. Acad. Sci. USA *84*, 8306–8310.

Luo, W., and Bentley, D. (2004). A ribonucleolytic rat torpedoes RNA polymerase II. Cell *119*, 911–914.

Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Dev. *20*, 954–965.

Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M., and Strathern, J.N. (2006). Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil. Genetics *172*, 2201–2209.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. Nature *444*, 953–956.

Manley, J.L., and Takagaki, Y. (1996). The end of the message–another link between yeast and mammals. Science *274*, 1481–1482.

Martin, G., Gruber, A.R., Keller, W., and Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. Cell Rep. *1*, 753–763.

Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C., and Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J. *30*, 1302–1310.

Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol. Cell *17*, 831–840.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol. *17*, 1272–1278.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012a). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723–1725.

Mayer, A., Schreieck, A., Lidschreiber, M., Leike, K., Martin, D.E., and Cramer, P. (2012b). The spt5 C-terminal region recruits yeast 3′ RNA cleavage factor I. Mol. Cell. Biol. *32*, 1321–1331.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature *430*, 223–226.

Minvielle-Sebastia, L., Beyer, K., Krecic, A.M., Hector, R.E., Swanson, M.S., and Keller, W. (1998). Control of cleavage site selection during mRNA 3′ end formation by a yeast hnRNP. EMBO J. *17*, 7454–7468.

Mischo, H.E., and Proudfoot, N.J. (2013). Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. Biochim. Biophys. Acta *1829*, 174–185.

Missra, A., and Gilmour, D.S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. Proc. Natl. Acad. Sci. USA *107*, 11301–11306.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038–1042.

Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell *161*, 526–540.

Park, J., Kang, M., and Kim, M. (2015). Unraveling the mechanistic features of RNA polymerase II termination by the 5′-3′ exoribonuclease Rat1. Nucleic Acids Res. *43*, 2625–2637.

Pearson, E.L., and Moore, C.L. (2013). Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor Rat1. J. Biol. Chem. *288*, 19750–19759.

Pearson, E., and Moore, C. (2014). The evolutionarily conserved Pol II flap loop contributes to proper transcription termination on short yeast genes. Cell Rep. *9*, 821–828.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127–131.

Pelechano, V., Wei, W., Jakob, P., and Steinmetz, L.M. (2014). Genome-wide identification of transcript start and end sites by transcript isoform sequencing. Nat. Protoc. *9*, 1740–1759.

Peters, J.M., Mooney, R.A., Grass, J.A., Jessen, E.D., Tran, F., and Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in Escherichia coli. Genes Dev. *26*, 2621–2633.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat. Rev. Mol. Cell Biol. *16*, 190–202.

Proudfoot, N.J. (1986). Transcriptional interference and termination between duplicated alpha-globin gene constructs suggests a novel mechanism for gene regulation. Nature *322*, 562–565.

Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem. Sci. *14*, 105–110.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–191.

Ray-Soni, A., Bellecourt, M.J., and Landick, R. (2016). Mechanisms of bacterial transcription termination: all good things must end. Annu. Rev. Biochem. *85*, 319–347.

Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. Mol. Cell *36*, 88–98.

SantaLucia, J., Jr. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA *95*, 1460–1465.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-wide mapping of yeast RNA polymerase II termination. PLoS Genet. *10*, e1004632.

Schreieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L.A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. Nat. Struct. Mol. Biol. *21*, 175–179.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075–1087.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228.

Sidorenkov, I., Komissarova, N., and Kashlev, M. (1998). Crucial role of the RNA:DNA hybrid in the processivity of transcription. Mol. Cell *2*, 55–64.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. *22*, 1350–1359.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114–117.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5′–> 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522–525.

Whitelaw, E., and Proudfoot, N. (1986). Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human alpha 2 globin gene. EMBO J. *5*, 2915–2922.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. (2009). Structure and function of the 5′–>3′ exoribonuclease Rat1 and its activating partner Rai1. Nature *458*, 784–788.

Xue, Y., Bai, X., Lee, I., Kallstrom, G., Ho, J., Brown, J., Stevens, A., and Johnson, A.W. (2000). Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. Mol. Cell. Biol. *20*, 4006–4015.

Zhang, Z., and Gilmour, D.S. (2006). Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Mol. Cell *21*, 65–74.

Zhang, H., Rigo, F., and Martinson, H.G. (2015). Poly(A) signal-dependent transcription termination occurs through a conformational change mechanism that does not require cleavage at the Poly(A) site. Mol. Cell *59*, 437–448.

Zhelkovsky, A., Tacahashi, Y., Nasser, T., He, X., Sterzer, U., Jensen, T.H., Domdey, H., and Moore, C. (2006). The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3′-end processing in Saccharomyces cerevisiae. RNA *12*, 435–445.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
mTOR (human FRB domain) polyclonal antibody	Enzo	ALX-215-065-1
Rabbit IgG antibody	Sigma-Aldrich	A0545
y-80 Rpb1 antibody	Santa Cruz	sc-25758
Chemicals, Peptides, and Recombinant Proteins		
Rapamycin	Sigma-Aldrich	R0395
4-thiouracil	Sigma-Aldrich	440736
EZ link Biotin-HPDP	Thermo Fisher Scientific	21341
Dithiothreitol	Sigma-Aldrich	43815
T4 PNK Reaction Buffer A	Fermentas	EK0031
5x First-Strand Buffer	Invitrogen	18080-044
Antarctic Phosphatase Reaction Buffer	NEB	M0289
Rnal2tr ligase buffer	NEB	M0351
T4 RNA Ligase 1 Reaction Buffer (10X)	NEB	B0204
KAPAHiFi PCR Kit	PeqLab	07-KK2100-01
RNase T1	Thermo Fisher Scientific	EN0541
RNase A	Fermentas	N/A
T4 Polynucleotide Kinase	Fermentas	EK0031
Antarctic Phosphase	NEB	M0289
T4 RNA Ligase 2 (K227Q)	NEB	M0351
Proteinase K	NEB	P8102
T4 RNA Ligase 1	NEB	M0204
Phusion HF DNA Polymerase	NEB	M0530
Bacterial Alkaline Phosphatase	Invitrogen	18011-015
RNase OUT	Invitrogen	10777-019
SuperScript III RT	Invitrogen	18080-093
Ambion TURBO Dnase	Invitrogen	AM2238
Critical Commercial Assays		
Ovation Universal RNA-Seq System Library Kit	NuGen	343
Encore Complete RNA-Seq Library Systems	NuGen	311
ThruPLEX DNA-seq Kit	Rubicon Genomics	N/A
Deposited Data		
Raw and processed sequencing data	This study	GEO: GSE79222
Experimental Models: Organisms/Strains		
BY4741, MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Euroscarf	N/A
Pcf11 TAP, BY4741 PCF11-TAP::HIS	Open Biosystems	N/A
Rai1 TAP, BY4741 RAI1-TAP::HIS	Open Biosystems	N/A
Rat1 TAP, BY4741 RAT1-TAP::HIS	Open Biosystems	N/A
Rtt103 TAP, BY4741 RTT103-TAP::HIS	Open Biosystems	N/A
Spt5 TAP, BY4741 SPT5-TAP::HIS	Open Biosystems	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Spt5Δ798-1063, BY4741 Spt5Δ798-1063::KAN	This study	N/A
Spt5Δ931-1063, BY4741 Spt5Δ931-1063::KAN	Mayer et al., 2012b	N/A
HHY168, MATalpha tor1-1 fpr1::NAT RPL13A-2xFKB12::TRP1	Euroscarf	N/A
Pcf11 FRB, HHY168 PCF11-FRB::KAN	This study	N/A
Rai1 FRB, HHY168 RAI1-FRB::KAN	This study	N/A
Rat1 FRB, HHY168 RAT1-FRB::KAN	This study	N/A
Rtt103 FRB, HHY168 RTT103-FRB::KAN	This study	N/A
Spt4 FRB, HHY168 SPT4-FRB::KAN	This study	N/A
Spt5 FRB, HHY168 SPT5-FRB::KAN	This study	N/A
Ysh1 FRB, HHY168 YSH1-FRB::KAN	This study	N/A
Oligonucleotides		
3′ adaptor (5′ rApp TGGAATTCTCGG GTGCCAAGG-3 ddC 3′, IDT)	Schulz et al., 2013	N/A
5′ adaptor (5′ HO-GUUCAGAGUUCU ACAGUCCGACGAUC-OH 3′, IDT)	Schulz et al., 2013	N/A
RT primer (5′ HO-CCTTGGCACCCG AGAATTCCA-OH 3′, IDT)	Schulz et al., 2013	N/A
NEXTflex barcode and universal primer	Bioo Scientific	5138-07/ 512911
Recombinant DNA		
pFA6a-FRB-KanMX6	Schulz et al., 2013	N/A
Software and Algorithms		
STAR version 2.3.0	Dobin et al., 2013	https://www.code.google.com/archive/p/rna-star/downloads
SAMtools	Li et al., 2009	http://www.htslib.org/
R	Ihaka and Gentleman, 1996	https://www.r-project.org/
Bioconductor	Gentleman et al., 2004	http://www.bioconductor.org/
DESeq	Anders and Huber, 2010	https://www.bioconductor.org/packages/release/bioc/html/DESeq.html
Bowtie version 1.1.1	Langmead et al., 2009	https://www.sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.1/
Bowtie version 2.2.3	Langmead and Salzberg, 2012	https://www.sourceforge.net/projects/bowtie-bio/files/bowtie2/2.2.3/
BEDtools	Quinlan and Hall, 2010	http://www.bedtools.readthedocs.io/
deepTools	Ramírez et al., 2014	https://www.deeptools.github.io/
model.R script	This paper	http://www.mpibpc.mpg.de/15585287/model.zip
Other		
2200 TapeStation	Agilent Technologies	N/A
Dynabeads Protein G	Invitrogen	10003D
Covaris S220 system	ThermoFisher Scientific	4465653
Qubit 1.0	Invitrogen	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Patrick Cramer (patrick.cramer@mpibpc.mpg.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cultivation of *S. cerevisiae* BY4741 strains was performed at 30°C and 160 rpm using either Yeast extract Peptone Dextrose (YPD) or Synthetic Complete (SC) medium supplemented with 2% glucose. Therefore, cells were plated on YPD agar and cultivated at 30°C for 1-2 days. One colony was subsequently used to inoculate a 30 mL YPD pre-culture. Cell density was photometrically determined using a spectrometer at 600 nm. One optical density unit (OD_{600}) corresponds to ~2.5×10^7 yeast cells. For CLIP experiments, *S. cerevisiae* cells expressing the TAP-tagged protein were grown at 30°C and 160 rpm from OD_{600} ~0.1 to OD_{600} ~0.5 in one liter of CSM minimal medium (Formedium) supplemented with 10 mg/l uracil, 100 mM 4-thiouracil (4tU) and 2% glucose. After reaching OD_{600} ~0.5, another 900 mM 4tU were added and cells were grown further for 4 hr (OD_{600} ~1.3-1.6). For anchor-away experiments, the FRB strains were grown in YPD overnight at 30°C. Cultures were diluted to OD_{600} ~0.1 and grown until OD_{600} ~0.6 before being labeled with 4-thiouracil.

METHOD DETAILS

Rapamycin-Dependent Depletion of Yeast Proteins from the Nucleus

Anchor-away experiments and 4-thiouracil (4tU)-Sequencing were performed as described (Schulz et al., 2013). The *Saccharomyces cerevisiae* parental anchor-away (AA) strain (W303; MATalpha ade2-1 trp1-1 can1-100 leu2-3,-112 his3-11,-15 ura3 GAL psi+ tor1-1 fpr1::NAT, RPL13A-2 × FKBP12::TRP1) and plasmid pFA6a-FRB-KanMX6 were obtained from Euroscarf. FRB-KanMX6 constructs were amplified using appropriate primers, inserted into the genome by homologous recombination, and transformants were selected on G418/Clonat plates. The presence of the FRB fragment was confirmed by PCR. For anchor-away experiments, the FRB strains were grown in yeast extract peptone dextrose (YPD) overnight at 30°C. Cultures were diluted to OD_{600}~0.1 and grown until OD_{600} = 0.6. Then, 4tU labeling was performed as described (Sun et al., 2012). One-half of the culture was supplemented with rapamycin to a final concentration of 1 μg/ml. Samples for 4tU-Seq were taken after 60 min from untreated cultures and cultures that were treated with rapamycin. Total RNA was extracted and labeled RNA purified. 100 ng of labeled RNA was used as input for the Ovation Universal RNA-Seq System Library Kit (NuGen) and processed following the manufacturer's protocol. Libraries were quality checked on a 2200 TapeStation (Agilent Technologies) and quantified with Qubit 1.0 (Invitrogen). Libraries were pooled and sequenced on an Illumina sequencer (HiSeq 1500).

To measure growth rates for anchor-away efficiency determination, the FRB strains were grown in YPD overnight at 30°C. Cultures were diluted to OD_{600}~0.1, incubated until an OD_{600}~0.6, and divided into two 35 mL cultures. Again, one-half of the culture was supplemented with rapamycin, the other half was supplemented with an equal volume of DMSO. The OD_{600} was then measured hourly for three biological replicates, to determine averaged growth rates. For spot dilutions, cells were grown overnight from fresh plates and then inoculated at OD_{600} ~0.3 in 5 mL YPD. After 4-5 hr, concentrations were adjusted for all strains to OD_{600} ~1.0, cell were washed twice with water, serially diluted five times 1:10 and spotted onto YPD and rapamycin plates (final conc. 1μg/mL). Each strain was plated in duplicate, plates were scanned every day for 3 days (Epson Perfection V700 Photo).

Growth curves alone are insufficiently informative with respect to the time point to be used for data acquisition after rapamycin addition during the anchor-away procedure. Therefore, we tested anchor-away strains for depletion of the factors of interest by ChIP-qPCR at several loci using an antibody against the FRB tag and the same experimental conditions as for the genome-wide analysis, i.e., after one hour of rapamycin treatment (Figure S4A). This analysis showed substantial decreases of the depleted factors from chromatin under the conditions used. The Pcf11 and Rai1 anchor-away strains could not be tested because, for unknown reasons, the FRB-tagged proteins did not produce strong ChIP signals. However, these two strains showed strong phenotypes in genome-wide experiments, suggesting that nuclear depletion was successful.

PAR-CLIP of Yeast Proteins

S. cerevisiae strains BY4741 containing C-terminally tandem affinity purification (TAP)-tagged genes (Open Biosystems) were tested for expression of the correct tagged protein by western blotting. Cells were lysed and diluted lysate was run on a pre-cast NuPAGE Bis-Tris gel (Invitrogen). Following SDS-PAGE, samples were blotted onto a PVDF membrane (Bio-Rad) and the membrane was probed with a primary PAP antibody against the TAP tag (Sigma-Aldrich). Antibody detection was performed using Pierce enhanced chemiluminescence (ECL) western blotting substrate (Thermo Scientific) and either an Amersham Hyperfilm ECL (GE Healthcare) or the Advanced Fluorescence Imager (Intas).

PAR-CLIP and data acquisition were performed as described (Baejen et al., 2014). Briefly, *S. cerevisiae* cells expressing the TAP-tagged protein were grown at 30°C from OD_{600} ~0.1 to OD_{600} ~0.5 in CSM minimal medium (Formedium) supplemented with 10 mg/l uracil, 100 μM 4-thiouracil and 2% glucose. After reaching OD_{600}~0.5, another 900 μM 4-thiouracil were added and cells were grown further for 4 hr (OD_{600}~1.3-1.6). 4tU-labeled yeast cells were UV-irradiated with an energy dose of 12 J/cm^2 at 365 nm. Cells were lysed by bead beating and the lysate was sonicated in a Covaris S220 system. Immunoprecipitation (IP) was performed overnight at 4°C with rabbit IgG-conjugated Protein G magnetic Dynabeads (Sigma-Aldrich, Invitrogen). Crosslinked RNA was partially digested with RNase T1 (Thermo Fisher Scientific) and prepared for cDNA library preparation. Following adaptor ligation, RNA was recovered by Proteinase K digestion (NEB), and subsequent acidic phenol/chloroform extraction. Reverse transcription was done using

SuperScript III RTase (Invitrogen). Generated cDNA was amplified, size-selected, and quantified using a 2200 TapeStation System (Agilent Technologies). Samples were sequenced on an Illumina machine (Genome Analyzer IIx or HiSeq 1500).

ChIP-Seq and ChIP-qPCR Experiments

Chromatin immunoprecipitation sequencing (ChIP-Seq) was performed as described (Mayer et al., 2010), with modifications. Yeast strains containing TAP-tagged versions of the proteins were grown in 600 mL YPD medium to mid-log phase (OD_{600} ~0.8). Yeast cultures were treated with 1% formaldehyde (Sigma-Aldrich) for 20 min at 20°C and subsequently quenched with 375 mM glycine for 5 min at 20°C. Cells were washed and cell pellets were flash-frozen in liquid nitrogen and stored at −80°C. Cells were lysed by bead beating and chromatin was sonicated in a Covaris S220 system. Immunoprecipitation was performed for 3 hr at 4°C with rabbit IgG-conjugated Protein G magnetic Dynabeads (Sigma-Aldrich, Invitrogen). Beads were washed in ChIP wash buffer and TE, and finally eluted in ChIP elution buffer for 10 min at 70°C. Eluted chromatin was treated with RNase A (Fermentas) and subsequently digested with Proteinase K (Sigma-Aldrich). The reversal of cross-links was performed overnight at 65°C. DNA was purified with a QIAquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. Purified DNA from IP and input samples was then used for ChIP-Seq library preparation with the ThruPLEX DNA-seq Kit (Rubicon Genomics), according to the manufacturer's protocol. Libraries were analyzed on a 2200 TapeStation System (Agilent Technologies), quantified with a Qubit 1.0 (Invitrogen), and sequenced on an Illumina machine (HiSeq 1500).

To conduct ChIP-Seq after rapamycin-dependent nuclear depletion of Rat1, the Rat1 AA-strain was grown from OD_{600} ~0.1 to ~0.7 before then supplemented with rapamycin (final concentration of 1 μg/ml). After 1 hr of incubation, the cell culture was treated with a final concentration of 1% formaldehyde (Sigma-Aldrich) and subsequently quenched. As a negative control, cultures were supplemented with an equal volume of DMSO instead of rapamycin. The ChIP-Seq experiments were performed as above. IP was performed using the y-80 Rpb1 antibody (Santa Cruz).

For ChIP analysis of factor depletion at individual genes after rapamycin addition, the same protocol as for ChIP-seq was used with the following exceptions. After inoculation cells were grown to OD_{600} ~0.6, the sample was split in half, to + sample rapamycin was added to final conc. 1 μg/mL, to – sample the same amount DMSO was added. Cells were grown for 1 hr at 30°C and then cross-linked. The IP was performed overnight with the mTOR (human FRB domain) polyclonal antibody (Enzo). qPCR was performed using 3 technical and 2 biological replicates on a Biometra qTower.

In Silico Model to Describe the 3′-Transition by RNA Levels Downstream of pA Site

We use the following notation: x: concentration of mRNA upstream of the pA site; y, z = mean lifetime for *cut* and *uncut* RNA downstream of the pA site, respectively; τ_x, τ_y, τ_z: mean lifetime of mRNA upstream of pA site and cut and uncut RNA downstream of the pA site; γ = synthesis rate of mRNA; p_{cut} = fraction of mRNAs that are cut at the pA site. At equilibrium, the rates of synthesis must equal the rates of decay: $dx/dt = \gamma - x/\tau_x = 0 \Rightarrow x = \tau_x \gamma$ and $dy/dt = p_{cut}\gamma - y/\tau_y \Rightarrow y = \tau_y p_{cut} \gamma$ and $dz/dt = (1\text{-}p_{cut})\gamma - z/\tau_z \Rightarrow y = \tau_y (1\text{-}p_{cut})\gamma$. Now we look at the positional dependence after the pA site. We assume that the polymerase terminates with an exponential decay length τ_y if the mRNA was properly cut. When the mRNA was not cut properly, the polymerase takes much longer to terminate and fall off. This is reflected in a longer decay length λ_z ? λ_y. Hence the concentration of RNA at position i after the pA site is modeled by $y(i) + z(i) = y_0 \exp(-i/\lambda_y) + z_0 \exp(-i/\lambda_z)$ and, using the previous equations for x, y, and z, we obtain $(y(i) + z(i))/x_0 = \tau_y/\tau_x p_{cut} \exp(-i/\lambda_y) + \tau_z/\tau_x (1\text{-}p_{cut}) \exp(-i/\lambda_z)$. Setting the parameters shown in Figure S6B produces the traces shown, with characteristic shapes indicating a change in mRNA cut efficiency (p_{cut}), in termination efficiency of uncut RNA (λ_z) or of cut RNA (λ_y). For details see the R script provided for downloading at http://www.mpibpc.mpg.de/15585287/model.zip.

QUANTIFICATION AND STATISTICAL ANALYSIS

Detection of Transcription Termination Sites (TTSs)

For each of the 4,928 protein-coding gene (ORF-Ts) instances identified by TIF-seq (Pelechano et al., 2013), the genomic region for site determination (potential termination window, PTW) was set from the last annotated pA site [according to all given isoforms of the respective ORF-T instance in the TIFseq annotation (Pelechano et al., 2013) to the annotated first TSS of the subsequent downstream feature [according to all given isoforms in the TIFseq annotation (Pelechano et al., 2013). PTWs exceeding 10 kb were trimmed down to 10 kb. For each PTW, putative TTSs were estimated by finding borders of abrupt coverage drops between consecutive segments via fitting a piecewise constant curve to the coverage profiles (whole fragments) for both replicates using the segmentation method from the R/Bioconductor package "tilingArray."

We selected 903 ORF-Ts for TTS determination based on two criteria, (i) the distance from the start of the PTW and the TTS was at least 5 bp, and (ii) the average coverage of the two replicates in the first segment of the PTW was at least 5. The TTS was set to be the border dividing two segments where the mean coverage dropped at least to 20% of its initial value. In total we determined 1,177 TTSs were only the ones with the maximal drops (for 903 instances) were selected for sequence analysis. The mean melting temperature for each termination window (TW, set from the last pA to the estimated TTS) was calculated as the gene-wise position based estimate of the melting temperature of 8-base pair DNA-RNA hybrids (SantaLucia, 1998). TWs were split by mean melting temperature into two groups and compared by length. This was done analogously for the average coverage of the TWs giving an approximation for a local synthesis rate.

We selected 2,501 genes with a minimal distance of 500 bp to the next genomic feature on the same strand, including ORF-Ts (Pelechano et al., 2013) and non-coding RNAs such as CUTs and SUTs (Neil et al., 2009; Pelechano et al., 2013), XUTs (van Dijk et al., 2011), NUTs (Schulz et al., 2013). This selection represents about half of the total 4,928 ORF-Ts identified by TIF-seq (Pelechano et al., 2013). We checked that this selection did not introduce any length or gene expression biases (Figure S1D).

4tU-Seq Data Acquisition and Analysis

Acquisition of data obtained from 4tU-Seq experiments was performed as described (Schulz et al., 2013), with minor modifications. Briefly, paired-end 50 base reads with additional 6 base reads of barcodes were obtained for labeled RNA. Reads were demultiplexed and aligned to the *S. cerevisiae* genome (sacCer3, version 64.2.1) using STAR (version 2.3.0) (Dobin et al., 2013). SAMTools was used to quality filter SAM files (Li et al., 2009). Alignments with MAPQ smaller than 7 (-q 7) were skipped and only proper pairs (-f99, -f147, -f83, -f163) were selected. Further processing of the data was carried out using R/Bioconductor (Gentleman et al., 2004; Ihaka and Gentleman, 1996).

Piled-up counts for every genomic position were summed up over replicates, using fragment coverage i.e., counting both sequenced bases covered by reads and unsequenced bases spanned between proper mate-pair reads. Size factors for each condition were calculated with the DESeq package (Anders and Huber, 2010) using only reads falling into ORF-T regions defined from TIF-Seq experiments (Pelechano et al., 2013). Those size factors were used to correct for library size and sequencing depth variations.

For Figures 3B, 4B, and 4D, we applied quantile normalization (scaling) to the 4tU-seq data for untreated samples in order to overlay the ''untreated'' profiles and to be able to compare the profiles following nuclear depletion with a common reference. To this effect, for untreated samples, we sorted coverage values and the values for the same ranks were replaced by the average value for the ranked dataset. We then scaled coverage values for each sample with nuclear depletion with the same transformation: We ranked the unnormalized coverage values of untreated and treated samples together and transformed the coverage values of the treated sample into the interpolation of the quantile-normalized values of the untreated sample with ranks just above and below.

To correlate absolute changes in 4tU-Seq read count level in the intergenic region and in downstream genes (Figures 7 and S7), we calculated cumulated read counts in those regions and normalized them by the length of the relevant region. The size factor normalization assumes that transcription levels stay constant between experiments, whereas in fact the knock-down of termination factors leads to read-through into the next gene and therefore to higher transcription levels. We estimated that about half of the genes are affected by read-through. Therefore, we assumed that total transcription levels in the anchor-away strains would be $0.5 \times x\%$ higher than in wild-type strains, where x% is the fraction of Pol II transcribing into downstream genes (estimated from the linear regression coefficient of Δz relative to Δy, obtained before applying this correction). This ensured that for $\Delta y = 0$ we had on average also $\Delta z = 0$.

PAR-CLIP Data Analysis

Data analysis of PAR-CLIP data was performed as described (Baejen et al., 2014). Adaptor sequences are first trimmed from the raw sequencing files. The quality filter then discards all reads containing unidentified nucleotides (N), Phreds scores below 30, reads shorter than 15 nucleotides (nt), or reads that are flagged by Illumina's internal chastity filter. Quality-trimmed reads are aligned to the *S. cerevisiae* genome (sacCer3, version 64.2.1) using the short read aligner Bowtie (version 1.1.1) (Langmead et al., 2009) with a maximum of one mismatch and taking unique matches only (options: -q -p 4 -S -sam -nohead -v 1 -e 70 -l 28 -y -a -m 1 -best -strata -phred33 -quals). The resulting SAM files are then converted into BAM and Pileup format using SAMTools (Li et al., 2009).

We calculated the P values for true cross-linking sites using our previously described statistical model (Baejen et al., 2014). We quantitatively modeled the null hypothesis, i.e., the probability that the T-to-C mismatches observed in reads covering a certain T nucleotide in the genome are not caused by cross-links between the immune-precipitated factor and RNA but are due to the other sources of mismatches. We estimated this null model distribution by fitting a two-component binomial mixture distribution to the frequency of the other 11 mutations. The first binomial component models the sequencing errors, while the second component models SNPs. We could then calculate the P value for any genomic T site, given the total number of reads covering the site (''coverage'') and given the fraction of these reads which show the T-to-C mismatch. Owing to the exquisite sensitivity of our experimental PAR-CLIP procedure, we could set a very stringent P value cut-off of 0.005 and a minimum coverage threshold of 4. For true cross-linking sites passing our stringent thresholds, the PAR-CLIP-induced T-to-C transitions strongly dominate over the contributions by sequencing errors and SNPs.

The number of observed T-to-C transitions at a site i in the transcriptome will be proportional to the occupancy of the immunoprecipitated factor on site i of the RNAs times the concentration of RNAs containing site i. Therefore, the occupancy of the factor on the RNA is proportional to the number of reads showing the T-to-C transition divided by the concentration of RNAs covering the T site. This concentration was estimated from the RNA-Seq read coverage measured under comparable conditions as described (Baejen et al., 2014) and was used to obtain normalized occupancies. The meta-transcript profiles shown in Figure 2 were smoothed within a window of ± 100 nt around each position. To obtain confidence intervals in Figure 2, we used a bootstrap analysis, drawing 500 times a set of 2501 of the genes with replacement, computing the mean metagene plots for these 500 sets and plotting the region between the 2.5% and 97.5% quantiles of these 500 traces as shaded areas. For Figure S2D, the meta-transcript profiles for each factor were

alternatively normalized (divided) by the Rpb1 meta-transcript profile, i.e., the averaged and smoothed coverage profile (raw data, number of T-C transitions) of each factor was divided by the averaged and smoothed Rpb1 coverage profile. For the bootstrap analysis, this was done for each bootstrap sample.

ChIP-seq Data Analysis

For genome-wide occupancy profiling, paired-end reads were aligned to the *S. cerevisiae* genome (sacCer3, version 64.2.1) using the short read aligner Bowtie (version 2.2.3) (Langmead and Salzberg, 2012) with options: -p 8–no-discordant. SAMTools was used to quality-filter SAM files (Li et al., 2009). Alignments with MAPQ smaller than 7 were skipped (-q 7). Duplicate reads were removed. The BEDTools toolset (Quinlan and Hall, 2010) was used to obtain coverage tracks that were subsequently imported into R/Bioconductor where further processing of the data was carried out. Normalization between IP and Input was done using the signal extraction scaling (SES) factor obtained with the estimateScaleFactor function from deepTools (Ramírez et al., 2014) with options: –l 100 –n 100000 and a median fragment size of 200 (-f 200). The ChIP-Seq enrichment at each genomic position was obtained by dividing ChIP intensities of unique, SES-normalized reads by the corresponding input intensities: $\log_2$(IP/Input). The final meta-gene profiles were obtained using a 5% trimmed mean of the ChIP-seq enrichment.

DATA AND SOFTWARE AVAILABILITY

Data

The accession number for the raw and processed data reported in this paper is NCBI Gene Expression Omnibus (Edgar et al., 2002): GSE79222.

Software

A custom R script containing the details of the computational model used to describe the 3′-transition is provided for download at http://www.mpibpc.mpg.de/15585287/model.zip.

Molecular Cell, Volume *66*

Supplemental Information

Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes

Carlo Baejen, Jessica Andreani, Phillipp Torkler, Sofia Battaglia, Bjoern Schwalb, Michael Lidschreiber, Kerstin C. Maier, Andrea Boltendahl, Petra Rus, Stephanie Esslinger, Johannes Söding, and Patrick Cramer

SUPPLEMENTARY FIGURES

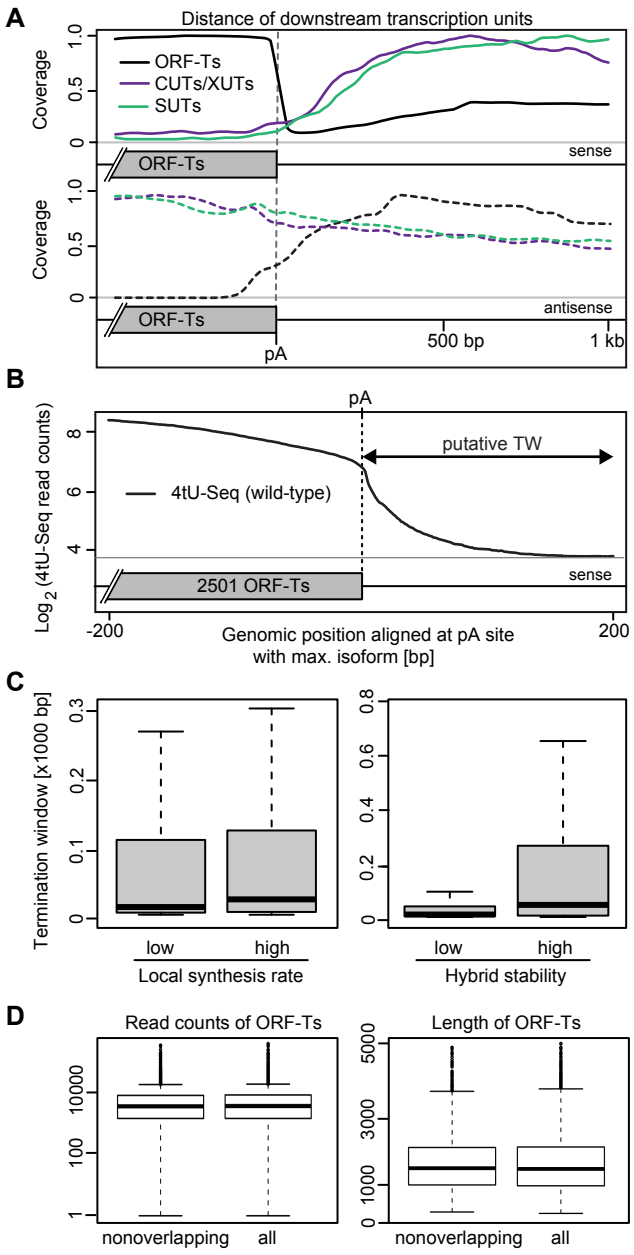


Figure S1, related to Figure 1

A. Occurrence of transcription units (ORF-Ts, CUTs/XUTs and SUTs) in sense (top) and antisense (bottom) direction downstream of the preferred polyadenylation (pA) sites of ORF-Ts. Relative average transcript densities were calculated via piling up every transcript of the TIFseq transcript annotation (Steinmetz) reflecting the theoretical situation where each transcript isoform has been completely and uniformly sequenced exactly once. Accumulation of these transcript frequencies in a metagene-wise fashion aligned at the designated loci results in the depicted densities for the given transcript classes.
The plot is done for all 4,928 ORF-Ts instances (aligned at their pA) identified by TIF-seq (Pelechano & Steinmetz 2013) but is also showing the CUTs/XUTs and SUTs categories.

B. Sense strand 4tU-Seq signals of the wild-type strain (log2 median position-based read count) of 2501 ORF-Ts with a minimal distance of 500 bp to the next genomic feature, aligned at the pA site with the max. isoform (annotated by TIF-Seq).

C. The size of the termination window is determined by both the local synthesis rate (left) and the hybrid stability (right). The boxplots compare the lower and higher quantiles below and above the median, respectively.

D. Comparison of the 2501 selected ORF-Ts with the full set of 4925 ORF-Ts identified by TIF-seq: Distribution of read counts per ORF-T in the two sets (left) and distribution of lengths (right)

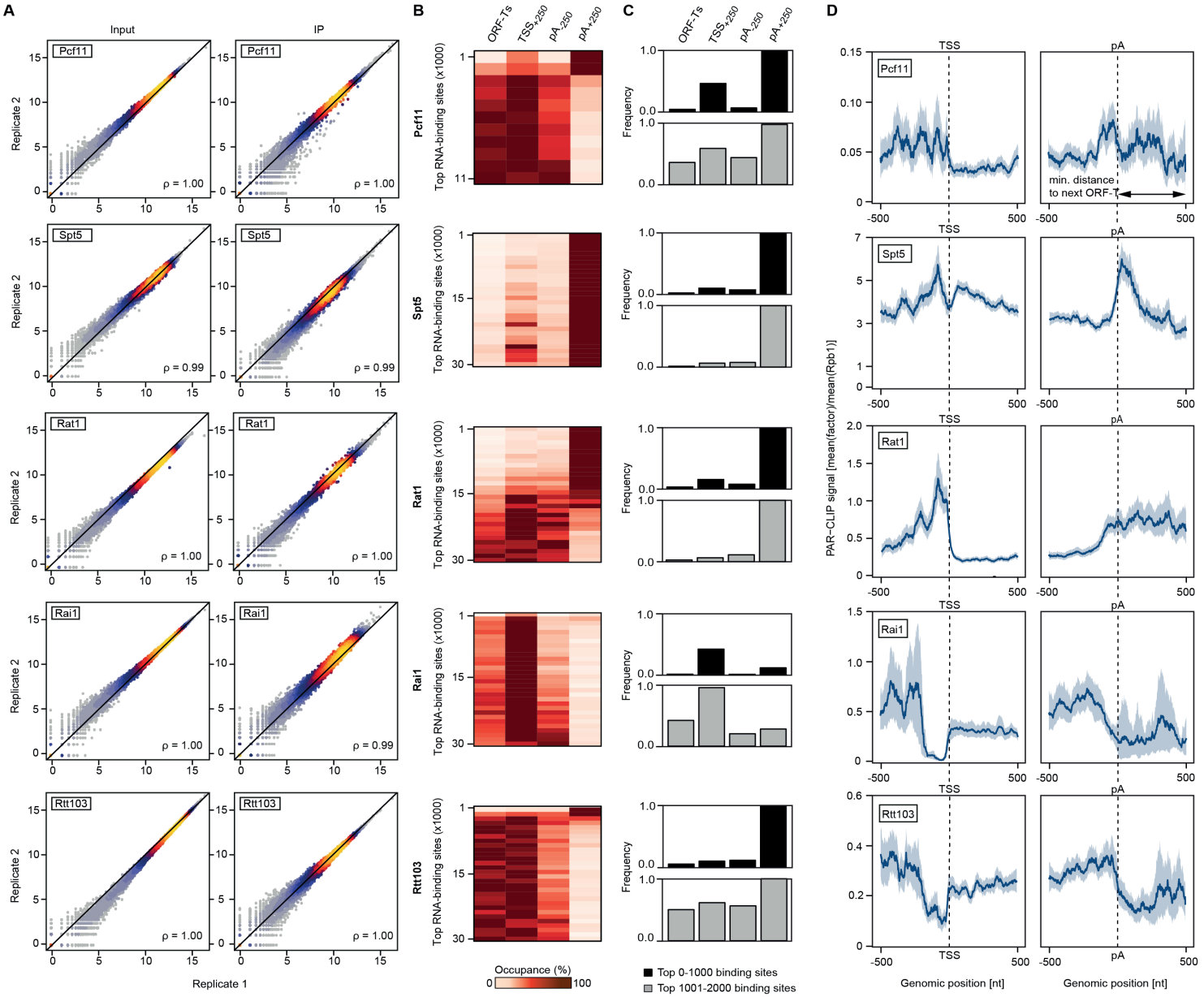


Figure S2, related to Figure 2

A. Comparison of replicate measurements for ChIP-Seq of the input and IP (from top to bottom) Pcf11, Spt5, Rat1, Rai1, and Rtt103. The scatterplot compares read counts of all annotated features (ORF-Ts, CUTs, and SUTs) using Spearman correlation.

B. Distribution of the top RNA-binding sites with respect to the TSS and the preferred polyadenylation (pA) sites. Each line represents a bin of 1000 RNA-binding sites sorted by occupancy. The color code shows the occupancy within the four selected regions: within the ORF-T, a window 250 nt downstream of the TSS (TSS+250), and two windows flanking the preferred polyadenylation (pA) sites 250 nt each.

C. Comparison of the first (top) and second (bottom) bin from panel B, corresponding to the top 1000 and top 2000 RNA-binding sites, respectively.

D. Rpb1-normalized occupancy profiles obtained from PAR-CLIP experiments of (from top to bottom) Pcf11, Spt5, Rat1, Rai1, and Rtt103. Normalization using Rpb1 PAR-CLIP data was performed by taking the ratio of the meta-transcript profile of each factor by the meta-transcript profile of Rpb1 (see STAR Methods). Profiles are aligned at both the transcription start site (TSS) and cleavage and polyadenylation (pA) site of 2501 ORF-Ts with a minimal distance of 500 bp to the next genomic feature. The shaded areas around the PAR-CLIP traces give symmetric 95% confidence intervals around the median value (blue trace), computed using a bootstrap procedure.

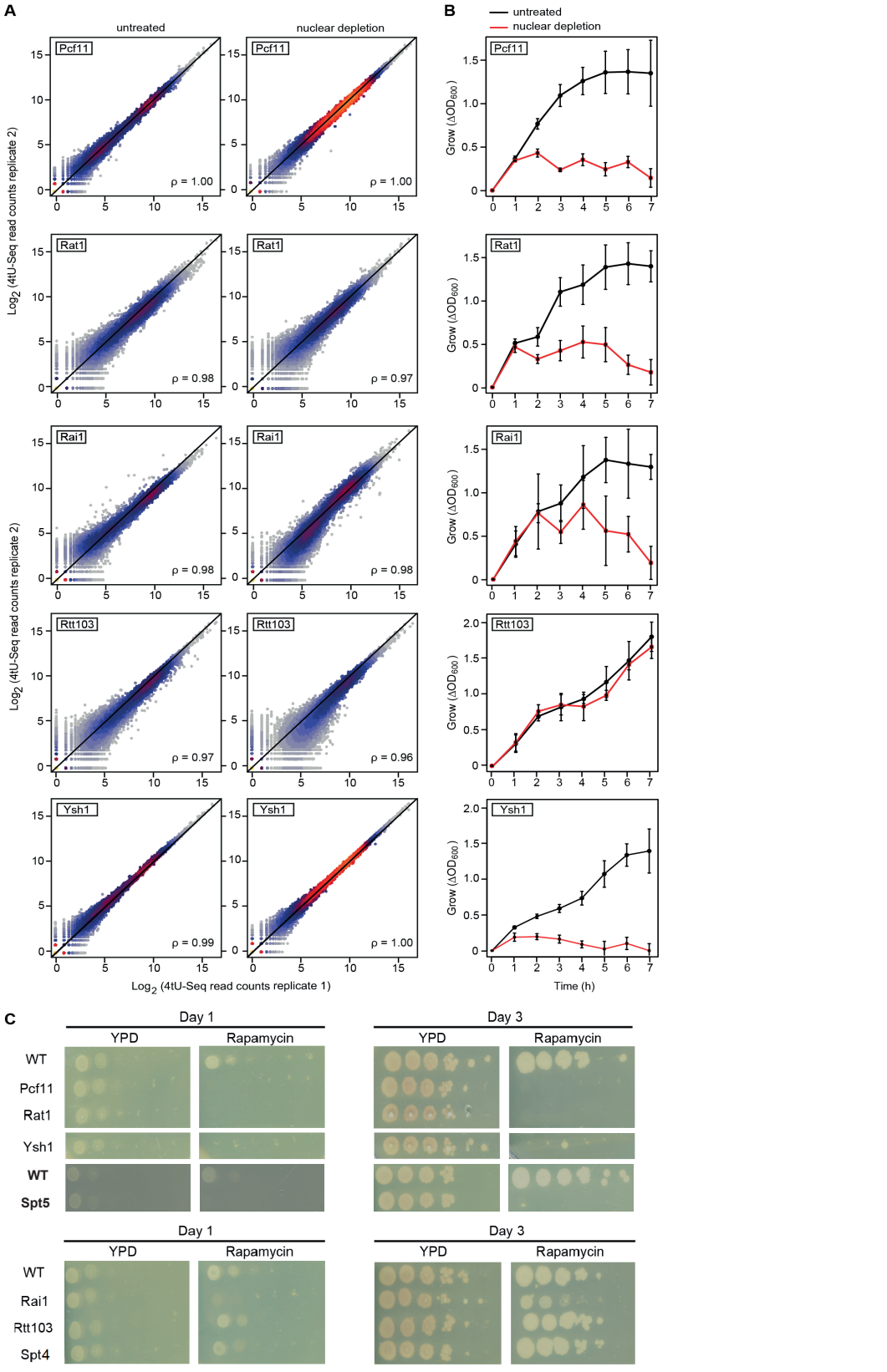


Figure S3, related to Figure 3

A. Comparison of replicate measurements for 4tU-Seq of the untreated sample and after nuclear depletion of (from top to bottom) Pcf11, Rat1, Rai1, Rtt103 and Ysh1. The scatterplots compare read counts of all annotated features (ORF-Ts, CUTs, and SUTs) using Spearman correlation.

B. Compared are the growth curves of anchor-away strains before (black) and after (red) nuclear depletion averaged for three biological replicates.

C. Comparison of spot dilutions of essential (top) and non-essential (bottom) factors on YDP plates and plates enriched with rapamycin after 1 and 3 day.

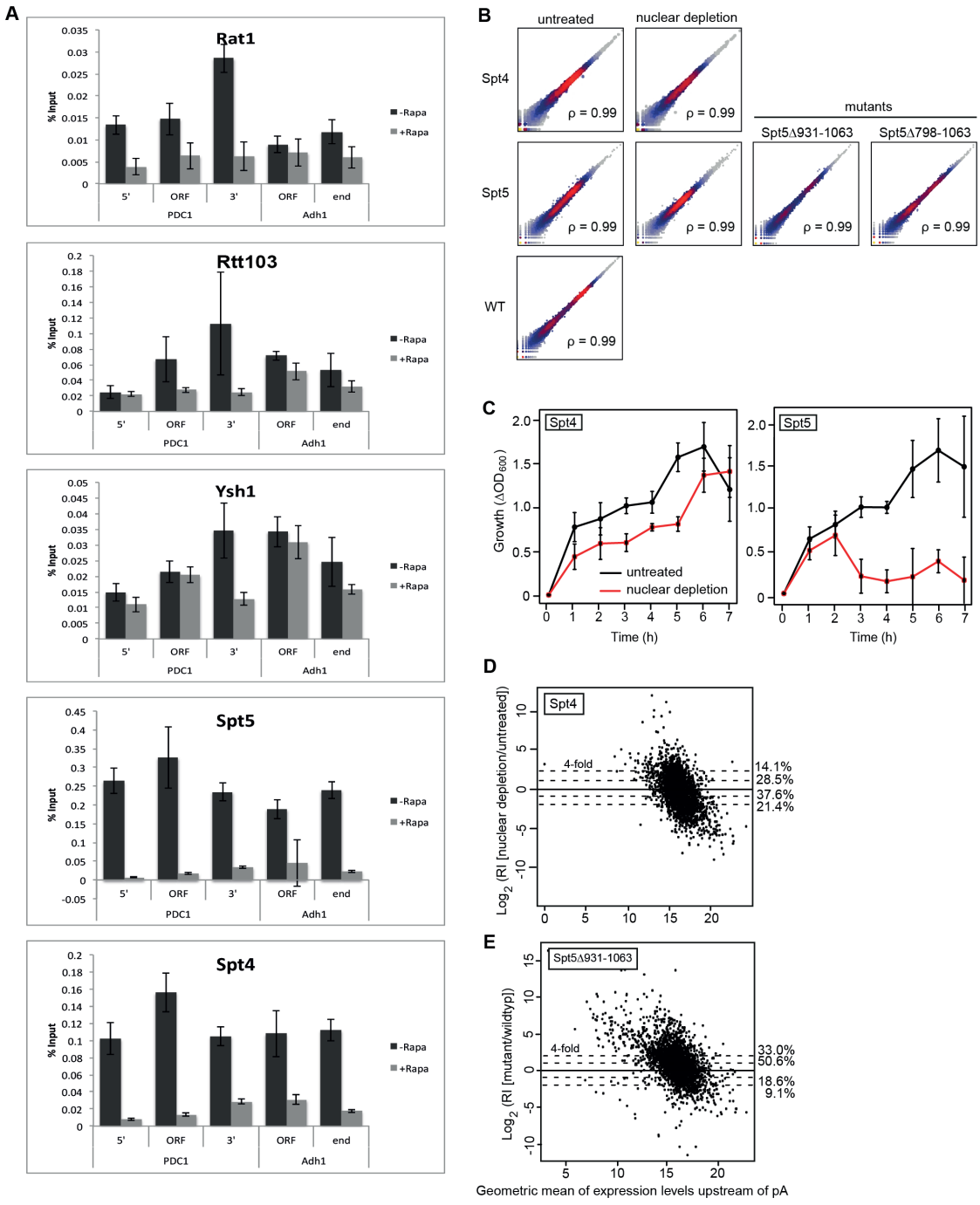


Figure S4, related to Figure 4

A. ChIP-qPCR validation of indicated anchor-away strains for selected genes PDC1 and ADH1. Three biological replicates were averaged for both untreated samples (dark grey) and after nuclear depletion by incubation with rapamycin for one hour (light grey).

B. Comparison of replicate measurements for 4tU-Seq before and after nuclear depletion of Spt4 (top) and Spt5 (middle left), as well as the 4tU-Seq replicates of the Spt5 mutant Δ931-1063 and Δ798-1063 (middle right) and the wild-type reference (bottom). The scatterplots compare read counts of all annotated features (ORF-Ts, CUTs, and SUTs) using Spearman correlation.

C. Growth curves of the Spt4 (left) and Spt5 (right) anchor-away strains. Compared are three replicates before (black) and after (red) nuclear depletion.

D. Normalized log2 RI before and after nuclear depletion of Spt4 versus the geometric mean of expression levels upstream (-500 to -50 bp) of the pA site in both samples. A subset of 2501 ORF-Ts that have no annotated feature 500 bp downstream of their pA site was used. The indicated percentages facing the dashed horizontal lines correspond to the proportion of ORF-Ts (out of 2501) that are above or below a 2-fold or 4-fold change in RI.

E. As in panel D but for the Spt5 mutant Δ931-1063

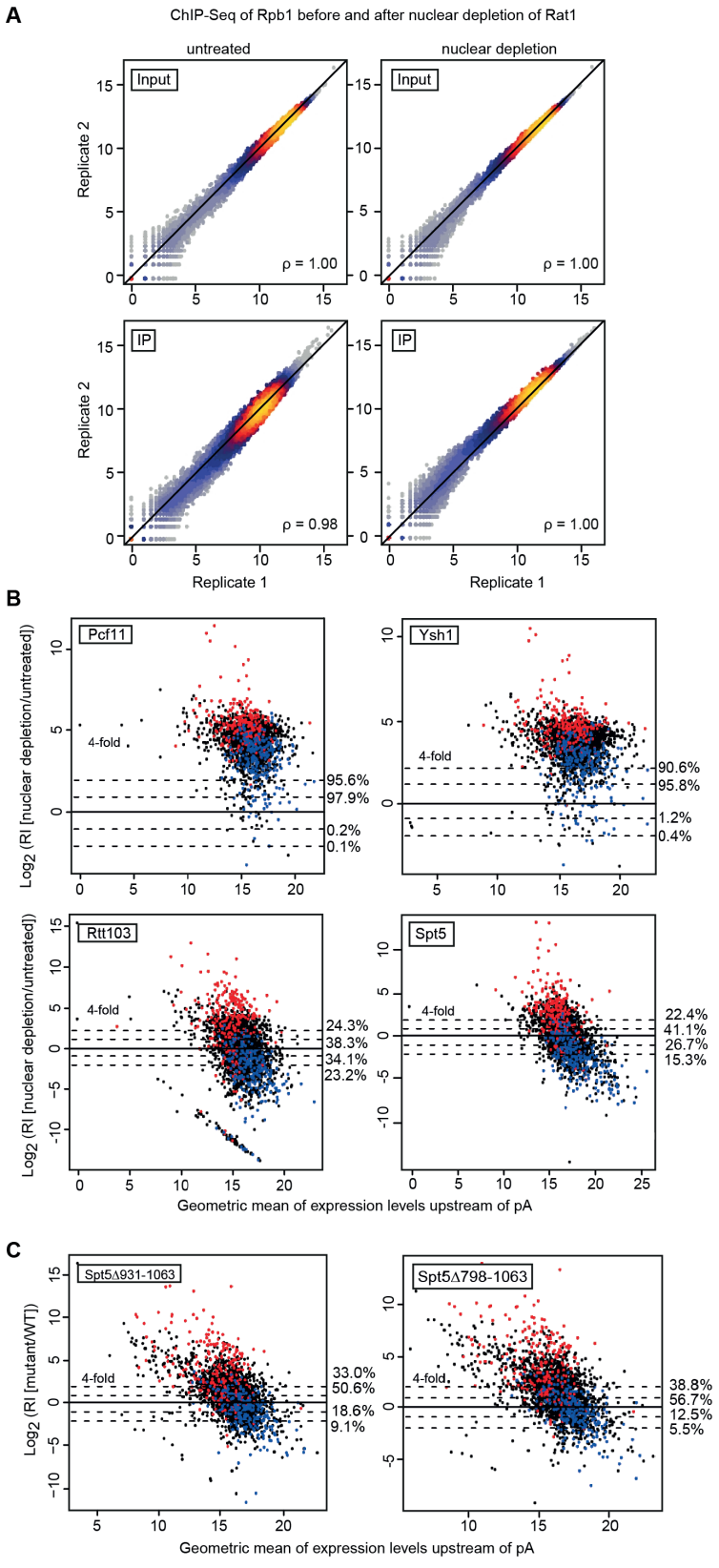


Figure S5, related to Figure 5

A. Compared are the replicate measurements for ChIP-Seq of Rpb1 (Pol II) without and after nuclear depletion of the exonuclease Rat1. The scatterplots compare read counts of all annotated features (ORF-Ts, CUTs, and SUTs) using Spearman correlation.

B. Normalized log RI before and after nuclear depletion of Pcf11, Ysh1, Rtt103, and Spt5 versus geometric mean of expression levels upstream of the pA site. The 250 most and least affected ORF-Ts after nuclear depletion of Rat1 are in red and blue, respectively.

C. As in panel B but for the Spt5 mutants Δ931-1063 and Δ798-1063.

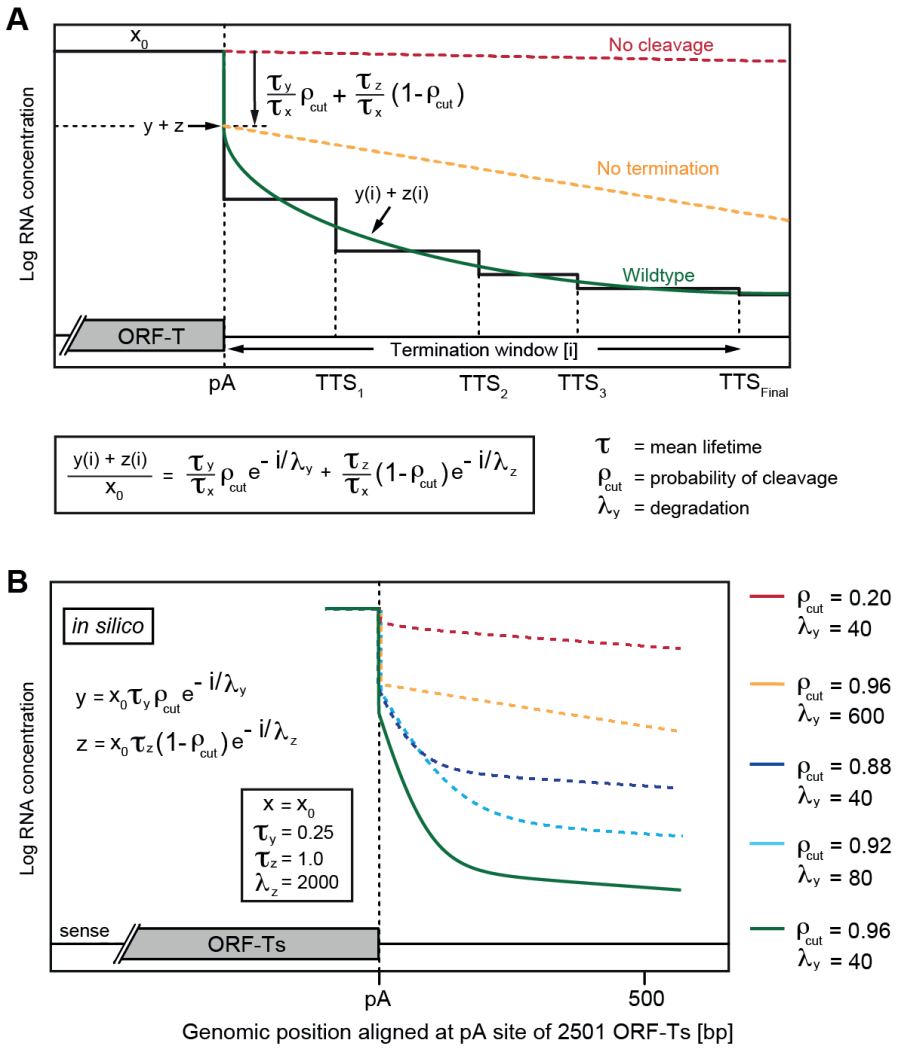


Figure S6, related to Figure 6

A. Summary of the in silico model for the RNA levels downstream of the pA site (for further details consult the STAR Methods). We defined x_0 as the concentration of mRNA upstream of the pA site, and y(i) and z(i) as the concentration of cleaved and uncleaved mRNA downstream of the pA site, respectively. The sum of y(i) and z(i) models the shape of the curve just after the pA site, which can now be compared to our measured 4tU-Seq traces. We modeled the decrease of the RNA level after the pA site by an exponential function exp(-i/λ) and observed that it is much slower for uncleaved than for cleaved RNA (λz >> λy). We assumed a constant rate of synthesis of mRNA ɣ for x, ρ_{cut} ɣ for y and (1-ρ_{cut}) ɣ for z, where ρcut is the fraction of mRNA cleaved at the pA site. This led to three time-dependent differential equations for x(t), y(t,i) and z(t,i) for which we calculated the equilibrium concentrations, x, y(i), and z(i).

B. Applied values for mathematically modeling the theoretical curves (as in Figure 6A)

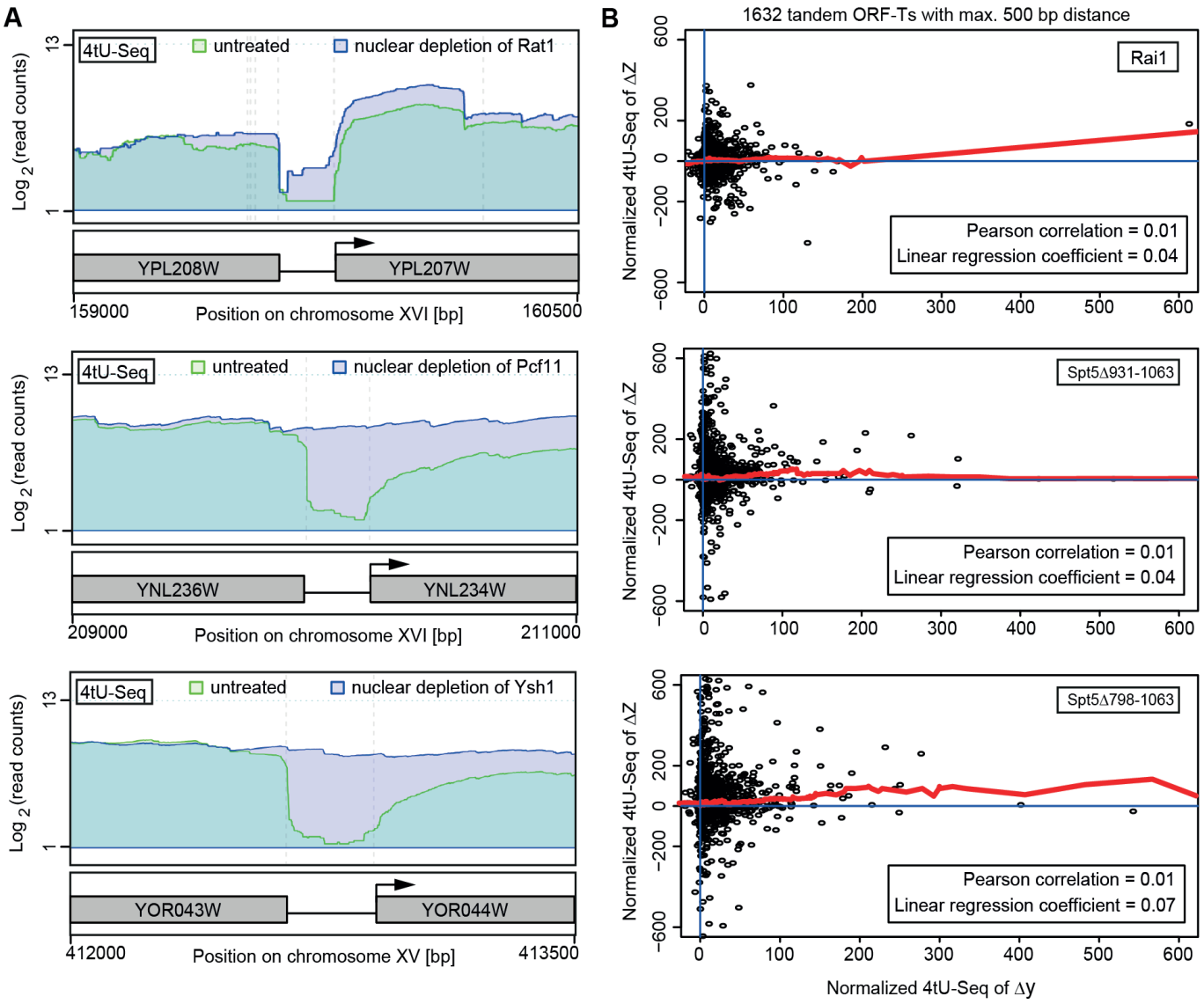


Figure S7, related to Figure 7

A. Genome browser view of log2 read counts measured by 4tU-Seq before (green) and after (blue) nuclear depletion of (from top to bottom) Rat1, Pcf11 and Ysh1, for selected pairs of tandem ORF-Ts.

B. The change between normalized 4tU-Seq read counts of the intergenic region (from top to bottom) before and after nuclear depletion of Rai1 and for the Spt5 mutants Spt5Δ931-1063 and Spt5Δ798-1063 compared to wild-type reference (Δy) is plotted versus the change in normalized read counts in the downstream ORF-T (Δz). The red line represents the running median over 1632 tandem ORF-Ts with a median and maximum intergenic distance of 168 bp and 500 bp, respectively.