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

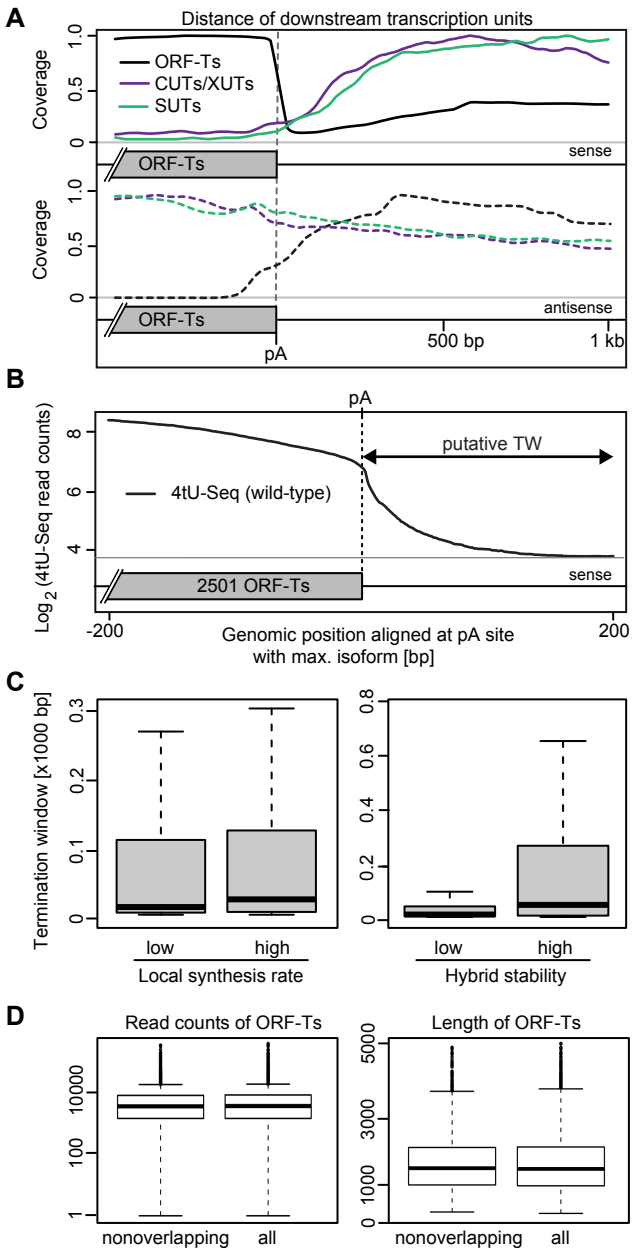

Figure S1, related to Figure 1

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)

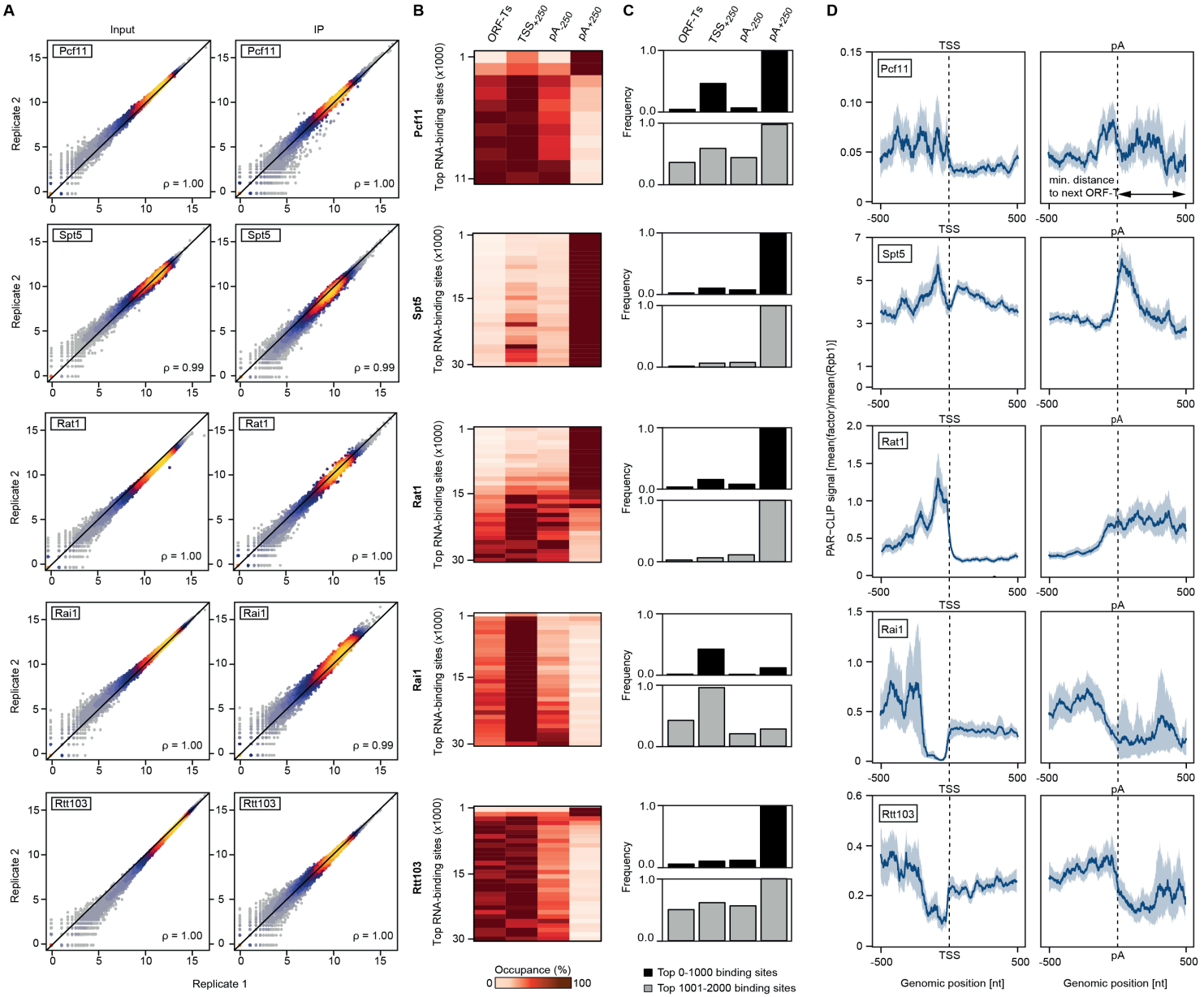

Figure S2, related to Figure 2

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.

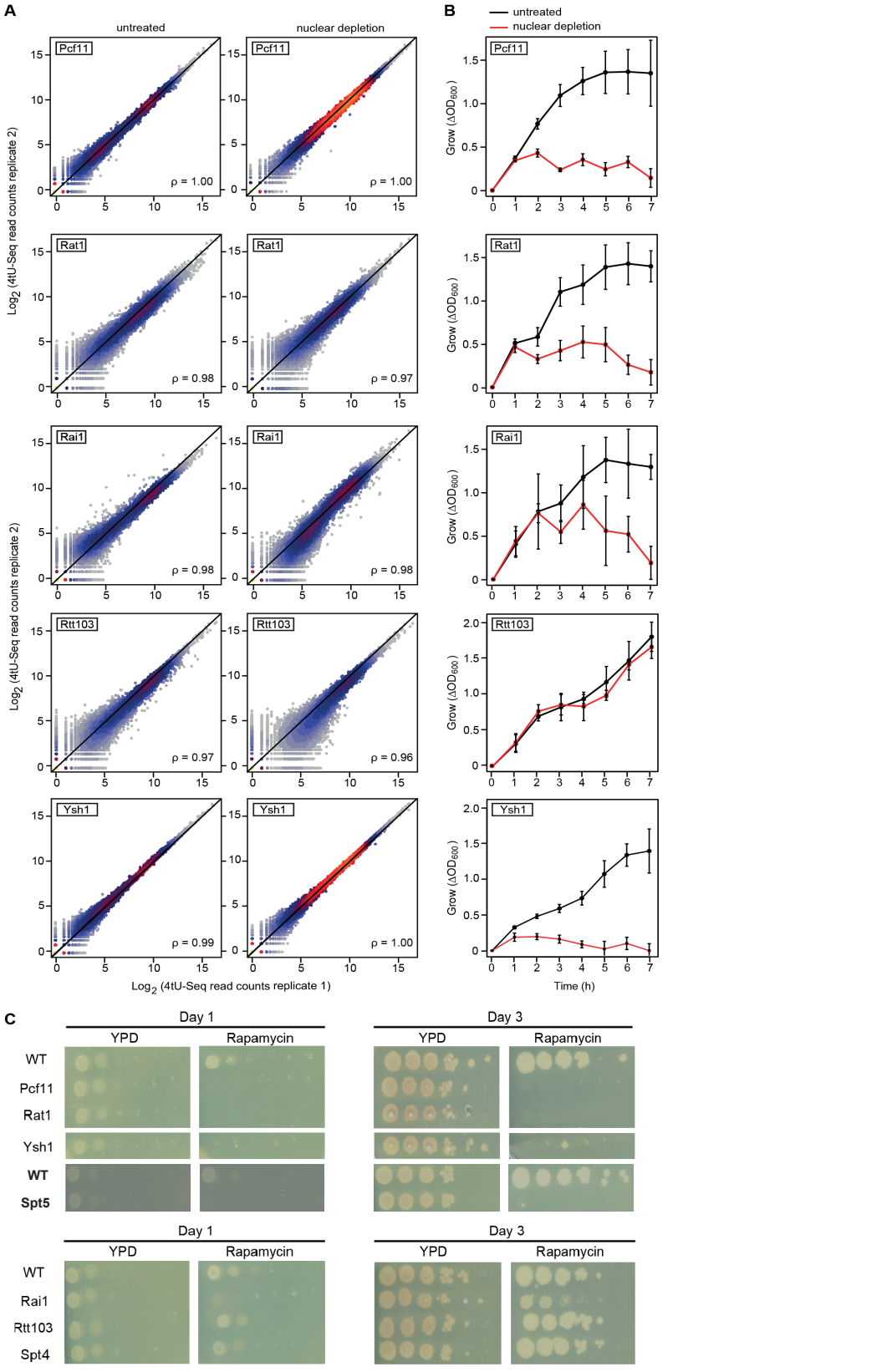

Figure S3, related to Figure 3

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.

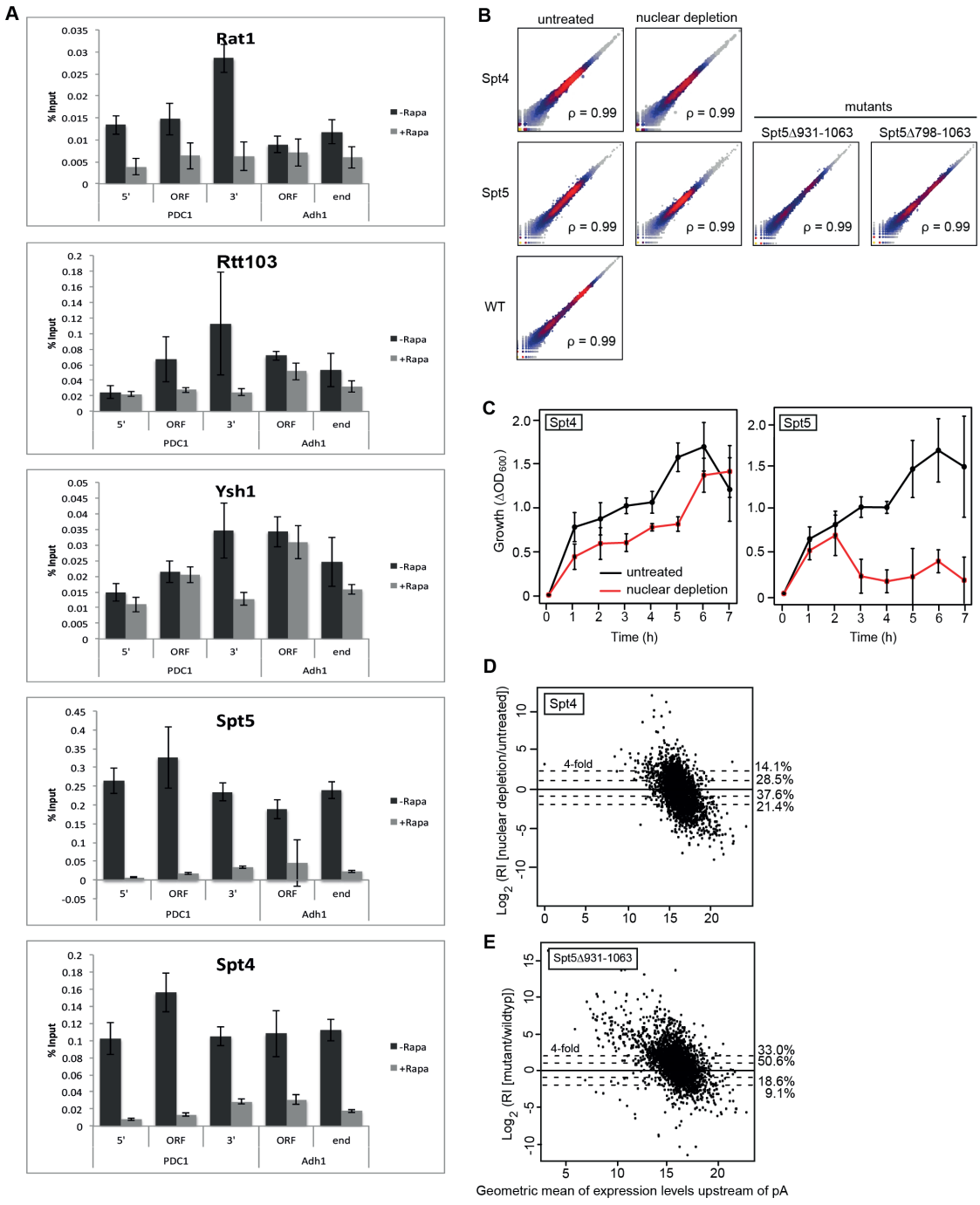

Figure S4, related to Figure 4

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

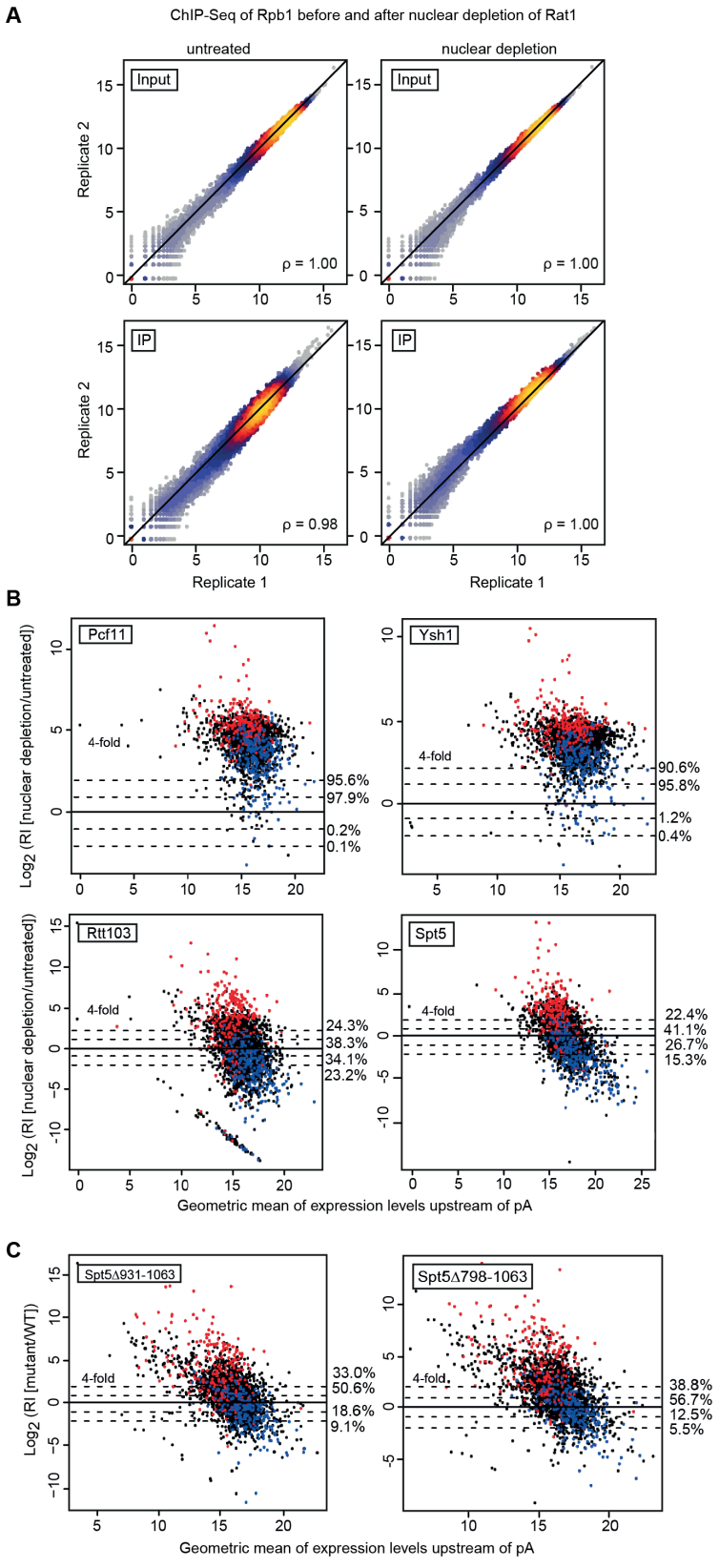

Figure S5, related to Figure 5

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.

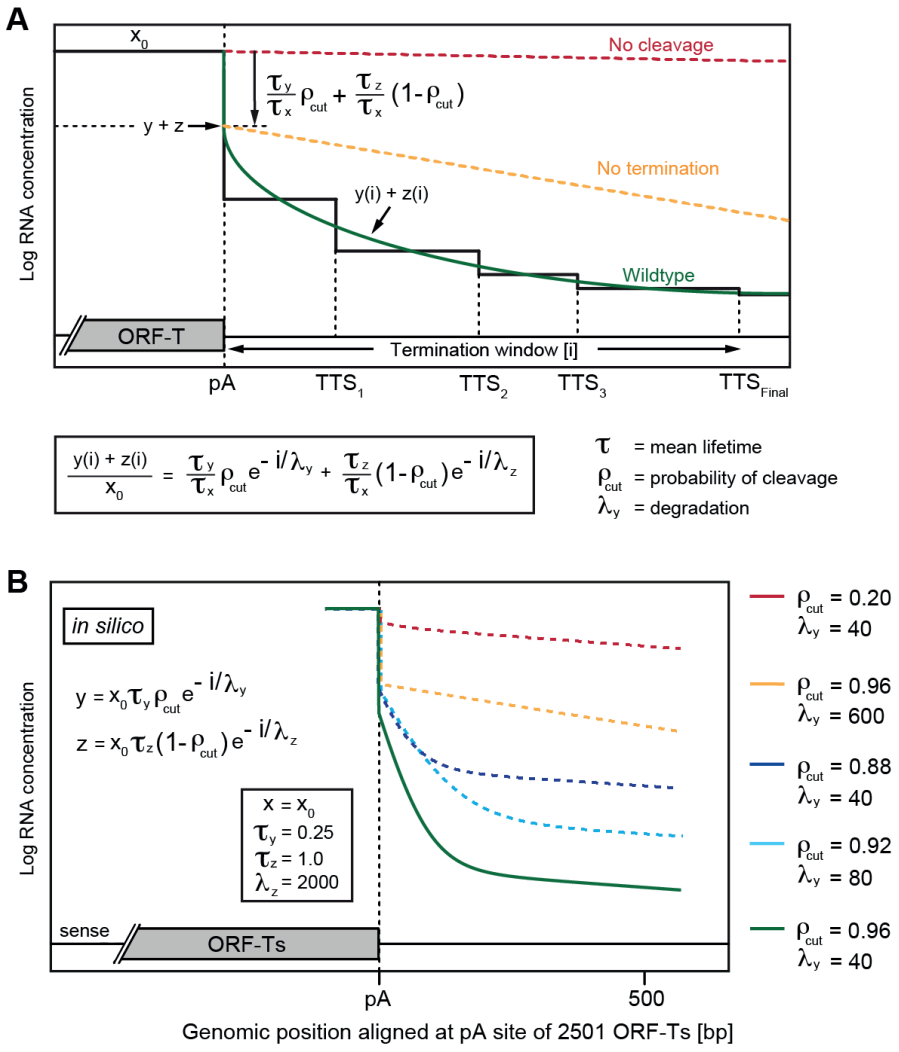

Figure S6, related to Figure 6

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)

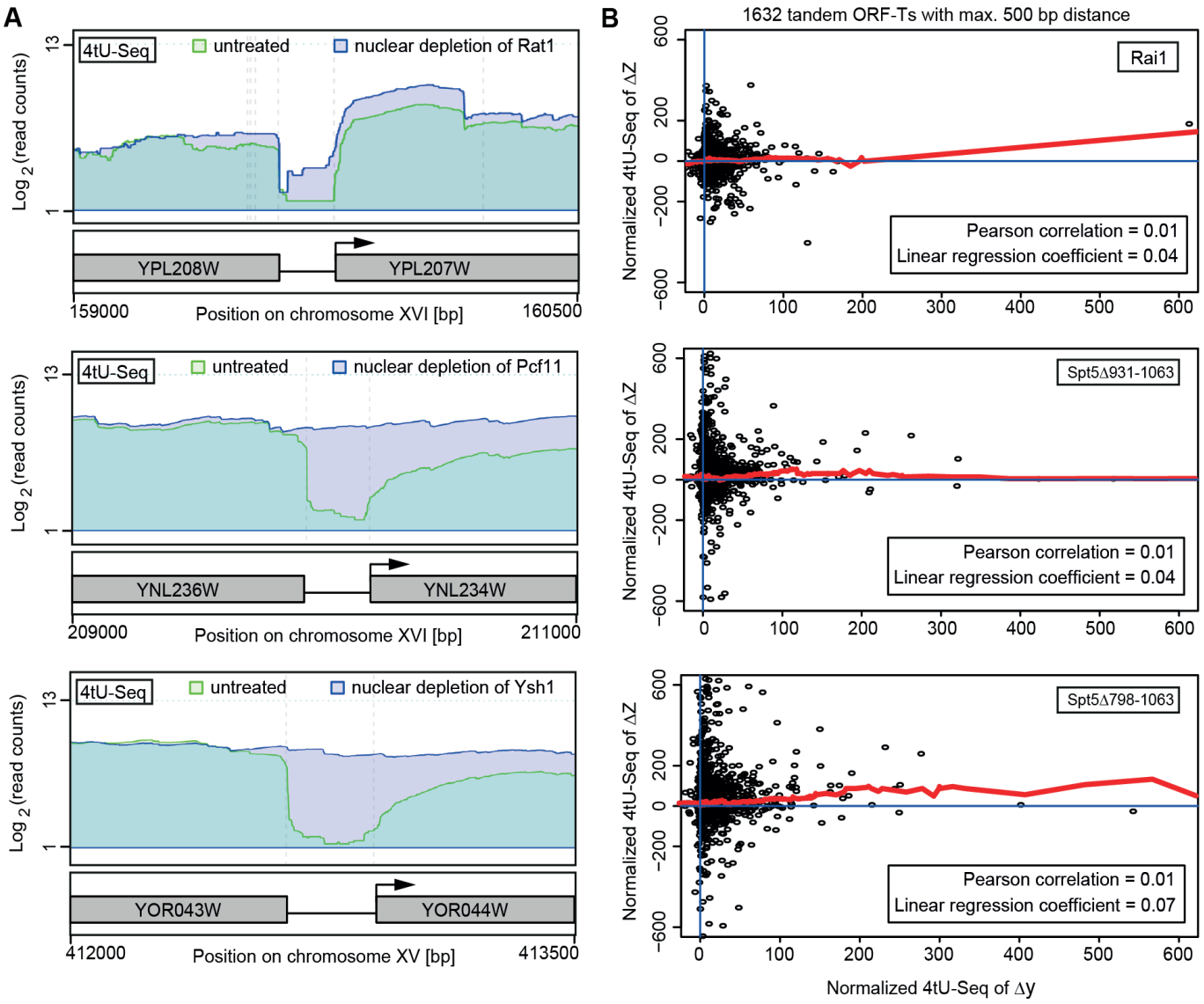

Figure S7, related to Figure 7

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.