SUPPLEMENTARY FIGURES

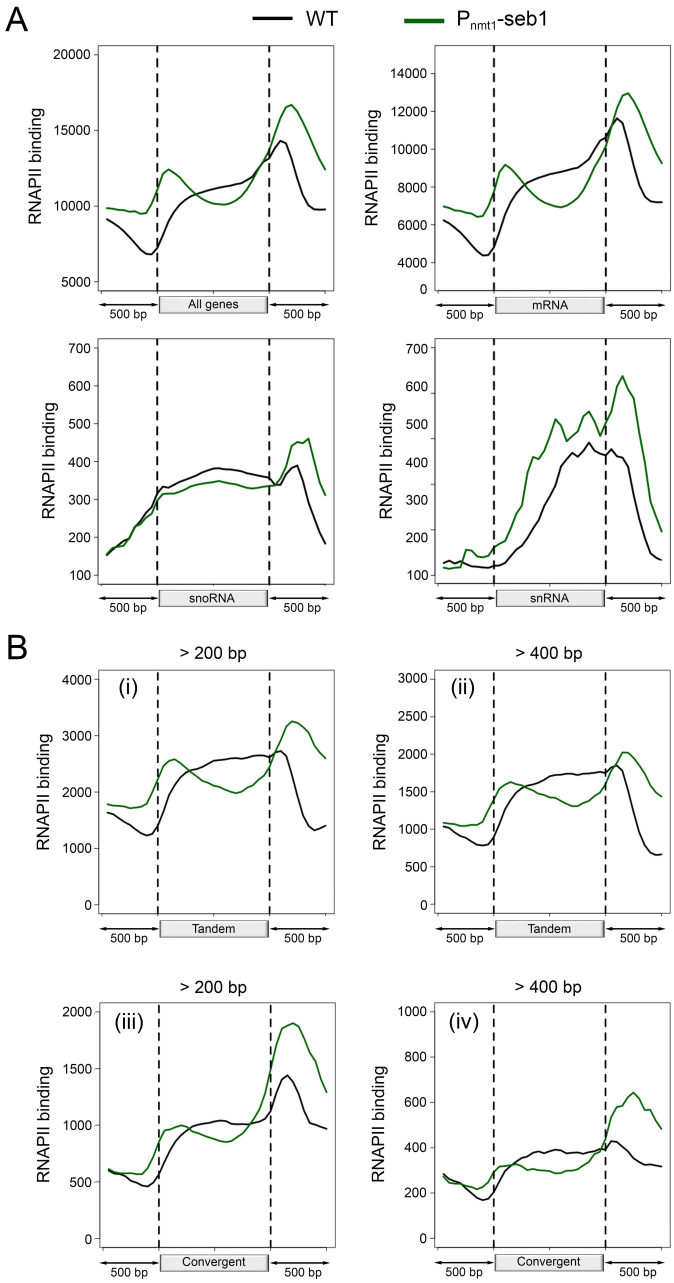

Figure S1. Read-through transcription by RNAPII in Seb1-deficient cells.

(A) Average distribution of RNAPII density in wild-type (WT) and Seb1-depleted cells (P_{nmt1}-seb1), as determined by analysis of ChIP-seq results for all (*n* = 7005), mRNA (*n* = 5123), snoRNA (*n* = 53) and snRNA (*n* = 7) genes. Light grey rectangles represent open reading frames (mRNA genes) or noncoding sequences (snoRNA and snRNA genes) and black arrows 500 base pairs (bp) of upstream and downstream flanking DNA sequences. The absolute levels of RNAPII binding (*y*-axis) depend on the number of genes taken into account. Lower values for snoRNA and snRNA reflect the low number of genes present in these categories as compared to mRNA. **(B)** As in **(A)** but for (i) tandem genes that are more than 200 bp apart (*n* = 1425), (ii) tandem genes that are more than 400 bp apart (*n* = 855), (iii) convergent genes that are more than 200 bp apart (*n* = 452) as well as (iv) convergent genes that are more than 400 bp apart (*n* = 208).

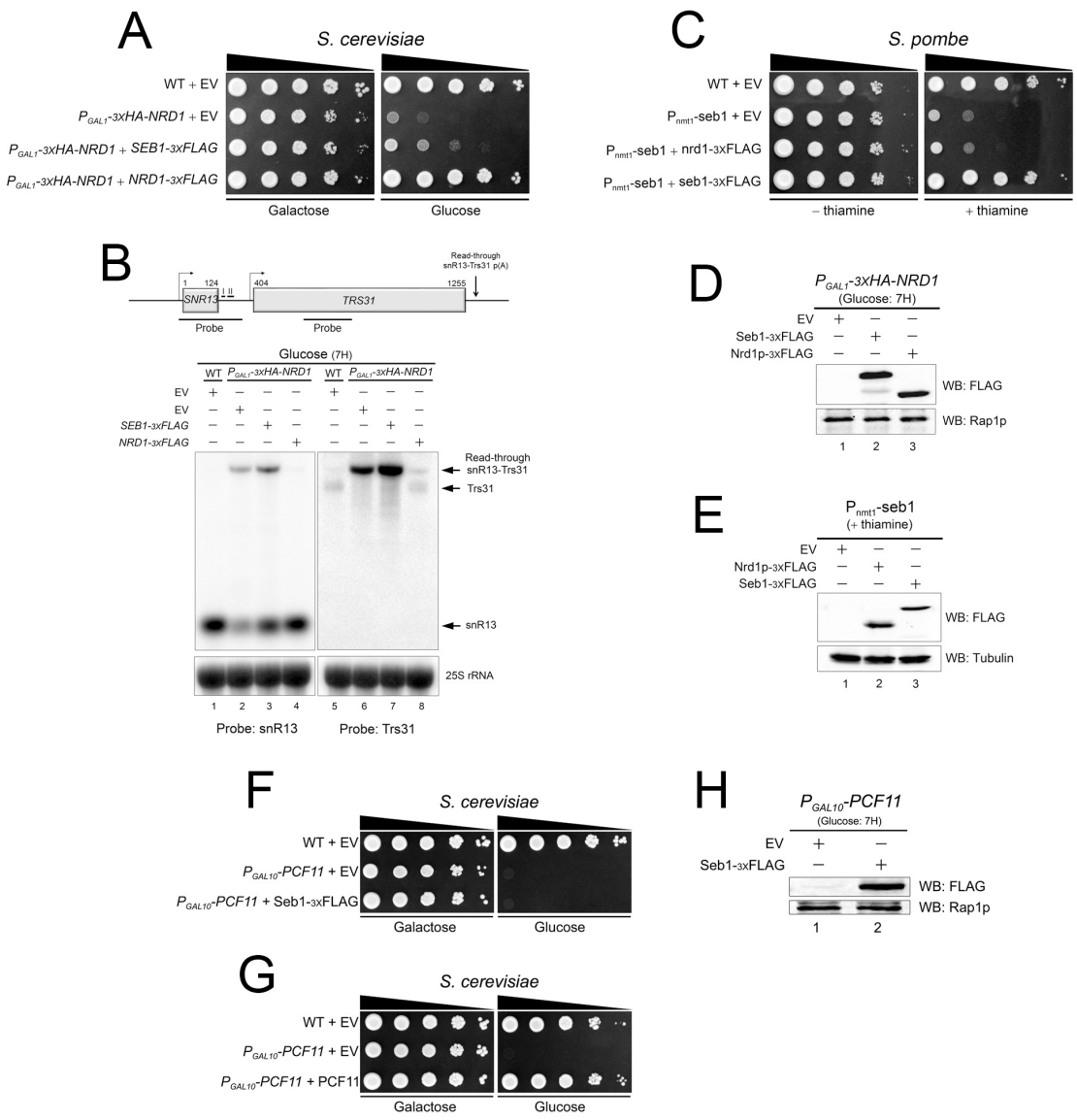

Figure S2. Seb1 is not the functional ortholog of Nrd1p.

(A) Ten-fold serial dilutions of *S. cerevisiae* wild-type (WT) cells transformed with an empty vector (EV) as well as P_{GAL1}-*3xHA-NRD1* cells that were transformed with an EV or constructs that express $_{3X}$FLAG-tagged versions of *S. pombe* Seb1 or *S. cerevisiae* Nrd1. Cells were spotted on galactose- and glucose-containing media to allow and repress transcription from the P_{GAL1} promoter, respectively. **(B)** *TOP*: Schematic representation of the *SNR13-TRS31* genomic environment. Bars under the genes show the positions of strand-specific RNA probes used for northern blot analyses. The position of the Trs31/snR31-Trs31 p(A) is illustrated as well as the regions (I/II) known to harbor *cis*-elements directing termination of the *SNR13* gene (Steinmetz and Brow 2003). *BOTTOM*: northern blot analyses of *snR13* (lanes 1-4) and *Trs31* (lanes 5-8) prepared from wild-type (WT) cells transformed with an empty vector (EV) (lanes 1 and 5) or from P_{GAL1}-*3xHA-NRD1* cells that were transformed with an EV (lanes 2 and 6) or constructs that express a $_{3X}$FLAG-tag wild-type version of Seb1 (lanes 3 and 7) or a $_{3X}$FLAG-tag wild-type version of Nrd1p (lanes 4 and 8). The 25S rRNA serves as a loading control. **(C)** Ten-fold serial dilutions of wild-type (WT) *S. pombe* cells transformed with an empty vector (EV) as well as P_{nmt1}-seb1 cells that were transformed with an EV or constructs that express $_{3X}$FLAG-tagged versions of Seb1 or Nrd1. Cells were spotted on minimal medium lacking or containing thiamine to allow and repress transcription from the P_{nmt1} promoter, respectively. **(D)** Immunoblot

analysis of whole-cell extracts prepared from P_{GAL1}-$_{3X}$*HA-NRD1 S. cerevisiae* cells that were transformed with an EV (lane 1) or constructs that express $_{3X}$FLAG-tag versions of Seb1 (lane 2) or Nrd1 (lane 3). Rap1 serves as a loading control. **(E)** As in **(D)** but in P_{nmt1}-seb1 *S. pombe* cells. Tubulin serves as a loading control. (**F, G**) Ten-fold serial dilutions of *S. cerevisiae* wild-type (WT) cells transformed with an empty vector (EV) as well as P_{GAL10}-*PCF11* cells that were transformed with an EV or constructs that express $_{3X}$FLAG-tagged versions of *S. pombe* Seb1 or *S. cerevisiae* Pcf11p. Cells were spotted on galactose- and glucose-containing media to allow and repress transcription from the P_{GAL10} promoter, respectively. **(H)** Immunoblot analysis of whole-cell extracts prepared from P_{GAL10}-*PCF11 S. cerevisiae* cells that were transformed with an EV (lane 1) or constructs that express $_{3X}$FLAG-tag versions of Seb1 (lane 2). Rap1 serves as a loading control.

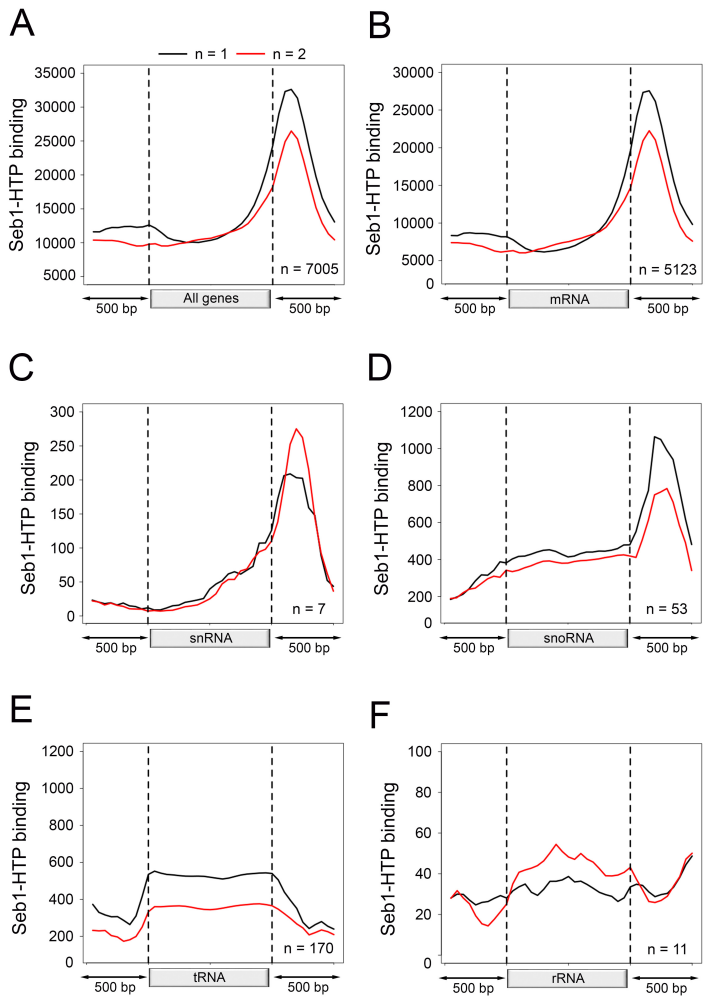

Figure S3. Genome-wide distribution of Seb1 at the 3'-end of RNAPII-transcribed genes.
(A-F) Average gene distribution of Seb1-HTP density as determined by analysis of results from two independent ChIP-seq experiments for **(A)** all, **(B)** mRNA, **(C)** snRNA, **(D)** snoRNA, **(E)** tRNA and **(F)** rRNA genes. Light grey rectangles represent open reading frames (mRNA genes) or noncoding sequences (snoRNA, snRNA, tRNA and rRNA genes) and black arrows 500 base pairs (bp) of upstream and downstream flanking DNA sequences. Sample size for each gene classes is indicated at the bottom right of the graphs. The absolute levels of RNAPII binding (*y*-axis) depend on the number of genes taken into account. Lower values for snoRNA and snRNA reflect the low number of genes present in these categories as compared to mRNA.

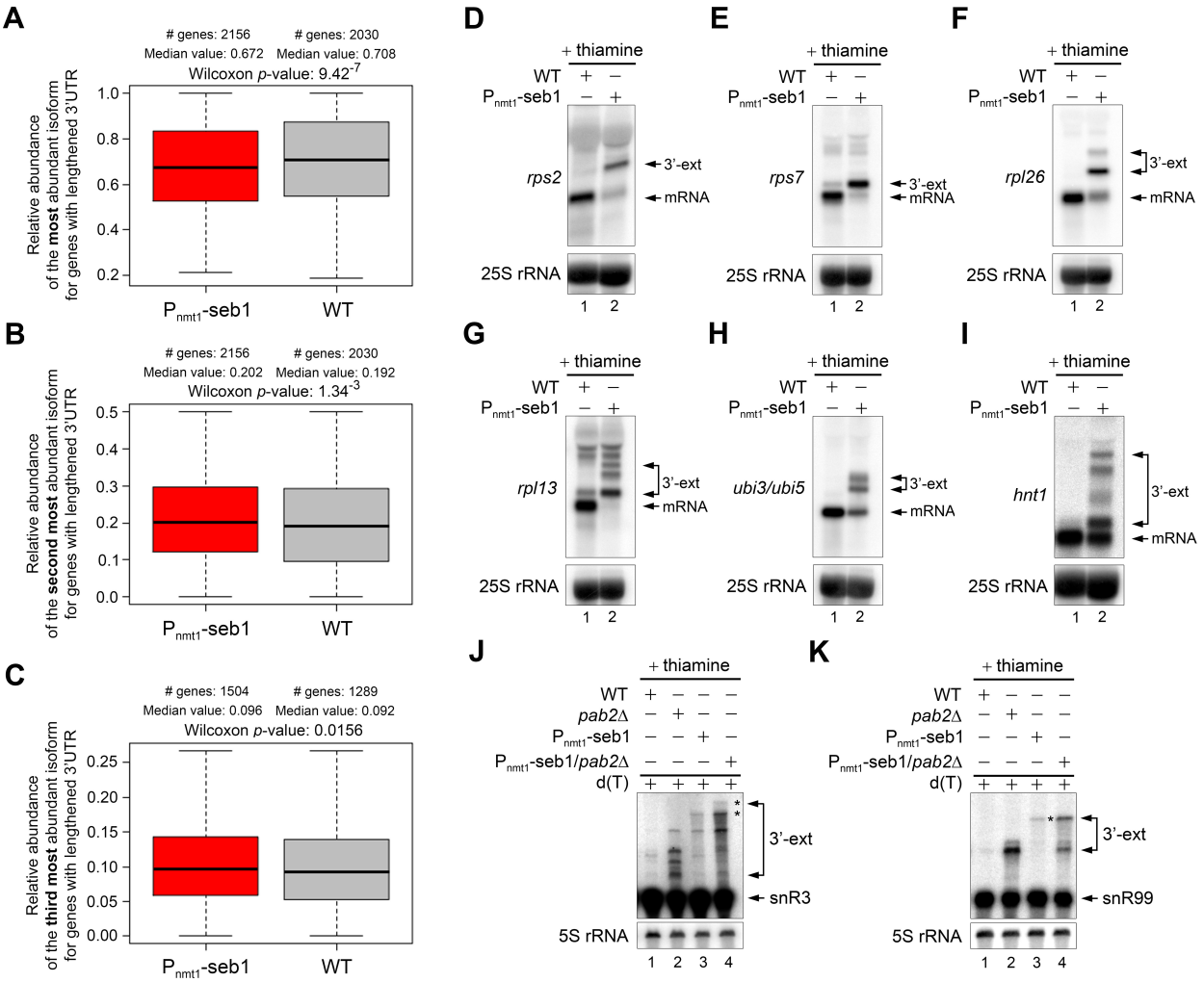

Figure S4. Seb1 levels modulate alternative polyadenylation (APA).

(A-C) Isoform relative abundance (RA) for the top 3 p(A) site (**(A)**: most abundant; **(B)**: second most abundant; **(C)**: third most abundant) for each gene in wild-type (WT) and Seb1-depleted cells (P_{nmt1}-seb1). The RA of each isoform is shown as a fraction of the combined abundance for the top 3 p(A) site. The number of genes used to determine the RA of each isoform, the median RA value in each strain and the degree of significance between the WT and P_{nmt1}-seb1 results are indicated at the top of each panel. The data show that there is a decrease of RA for the most abundant p(A) site in the *seb1* mutant compared to the WT and, consequently, increases of RA for the other 2 p(A) sites in the mutant. **(D-I)** Northern blot analyses of the *rps2* **(D)**, *rps7* **(E)** *rpl26* **(F)**, *rpl13* **(G)**, *ubi3/ubi5* **(H)** and *hnt1* **(I)** mRNA-encoding genes using RNA from WT and P_{nmt1}-seb1 cells. 3'-ext refers to the 3'-extended transcripts that accumulate in Seb1-depleted condition. The 25S rRNA is used as a loading control. **(J-K)** Northern blot analyses of snoRNA-encoding genes using RNA obtained from the indicated strains. Total RNA was treated with RNase H in the presence of a DNA oligonucleotide complementary to H/ACA class snoRNA snR3 **(J)** or snR99 **(K)**. RNase H reactions were performed in the presence of oligo(dT) to collapse the 3'-extended polyadenylated forms (3'-ext) of these snoRNAs which are normally produced as part of their maturation/degradation cycle (Lemay et al. 2010; Larochelle et al. 2012). The *pab2*Δ and P_{nmt1}-seb1/*pab2*Δ strains were used to increase the abundance of these 3'-ext forms (Lemay et al. 2010). The stars (*) indicate the 3'-ext forms that appear specifically following the depletion of Seb1. The 5S rRNA serves as a loading control.

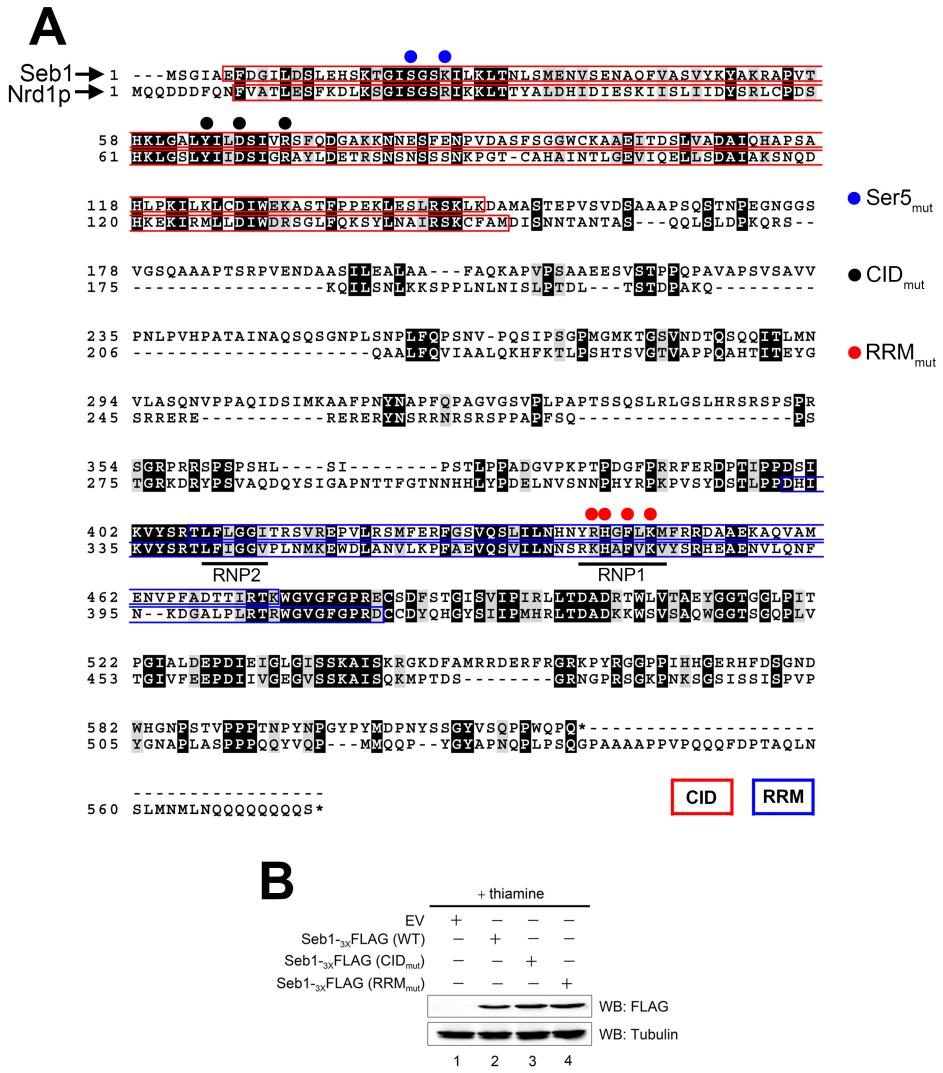

Figure S5. CID and RRM domains of Seb1.

(A) Sequences were aligned using Clustal Omega (Sievers et al. 2011). The RNAPII CTD-interacting domain (CID) and RNA recognition motif (RRM), as defined by the Simple Modular Architecture Research Tool (SMART) algorithm (Schultz et al. 1998), are boxed in red and blue, respectively. Consensus sequences involved in RNA interaction are identified as RNP1/RNP2 (Steinmetz and Brow 1996). Conserved and similar residues are highlighted in black and gray, respectively. The function of the CID and RRM domains of Seb1 was addressed using mutagenesis of conserved and/or similar residues. To create the phospho-Ser5 CID mutant (Ser^{5}_{mut}), serine (S) 22 and lysine (K) 25 were substituted to alanines (S22A/K25A) (blue dots). The same approach was used for the other mutants generated. This includes another mutant of the CID domain (CID_{mut}) where tyrosine (Y) 64, aspartic acid (D) 67, and arginine (R) 71 were all substituted to alanine (Y64A/D67A/R71A) (black dots) as well as the RRM mutant (RRM_{mut}) where arginine (R) 442, histidine (H) 443, phenylalanine (F) 445, and lysine (K) 447 were all converted to alanine (R442A/H443A/F445A/K447A) (red dots). **(B)** Western blot analysis of extracts prepared from P_{nmt1}-seb1 cells transformed with an empty vector or the indicated Seb1 constructs.

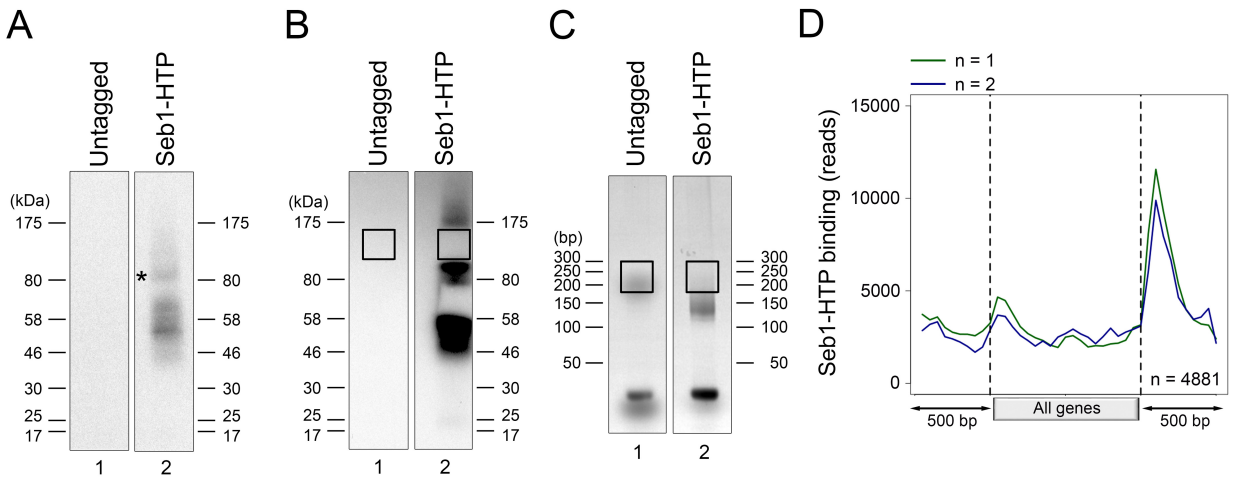

Figure S6. Seb1 RNA-binding profile by CRAC.

(A-C) Purification of Seb1-RNA complexes. **(A)** Immunoblot analysis of Seb1-RNA cross-linked complexes purified via nickel affinity purification under denaturing conditions. A smear with cross-linked RNA products is seen above the signal corresponding to full-length Seb1 (indicated by asterisk). The smear-region was cut out of the membrane, as demonstrated in Fig. S6b, and RNA were isolated for the generation of cDNA libraries. **(B)** Autoradiogram after isolating Seb1 with cross-linked ^{32}P-labelled RNAs of sufficient length that would allow their identification by high-throughput sequencing. **(C)** Agarose gel electrophoresis of PCR products obtained from cDNA libraries. Boxed regions were gel extracted and submitted for high throughput sequencing. **(D)** Average gene distribution of preferential Seb1-HTP binding sites as determined by the analysis of Seb1-bound RNAs from two independent CRAC experiments. Light grey rectangles represent open reading frames or noncoding sequences and black arrows 500 base pairs (bp) of upstream and downstream flanking DNA sequences. The sample size is indicated at the bottom right of the graph.

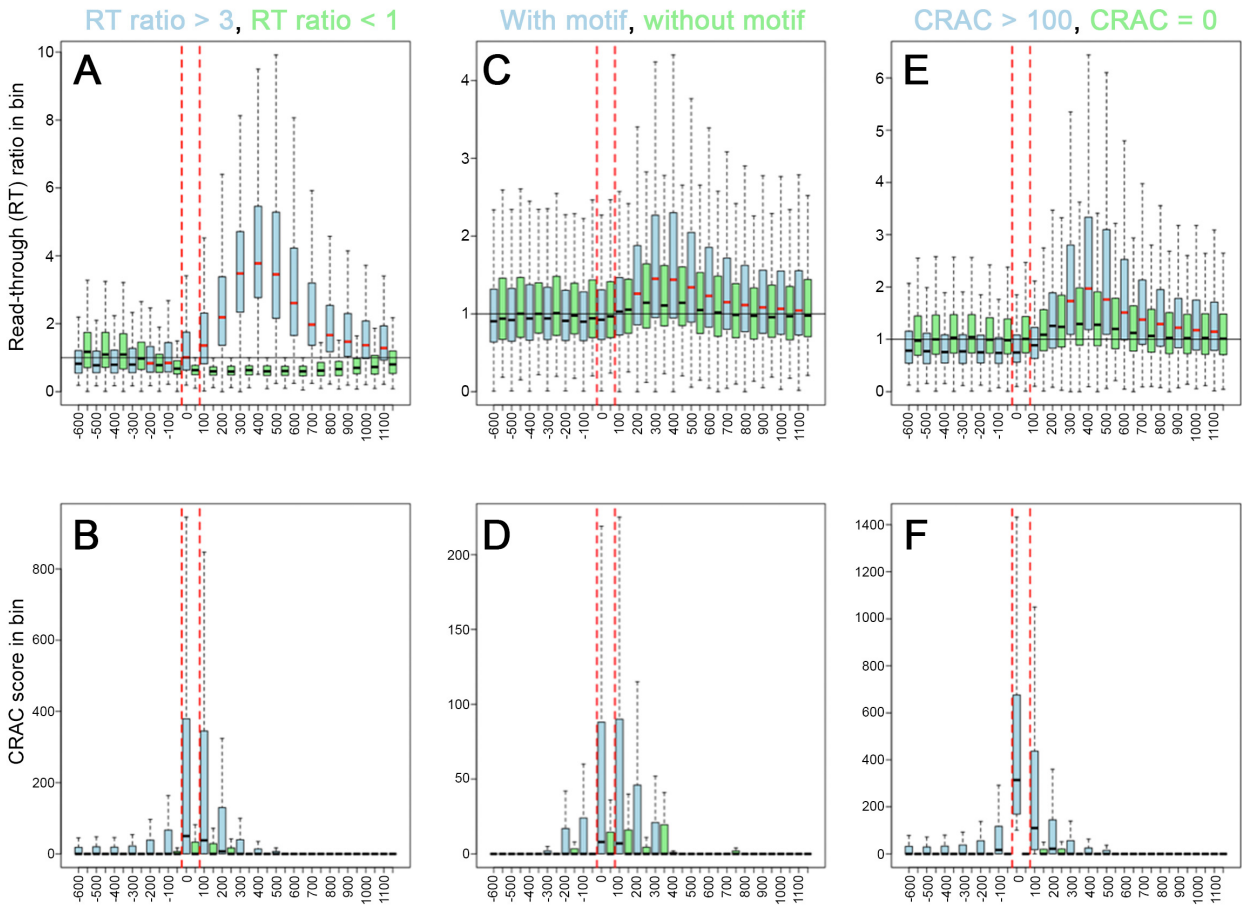

Figure S7. Distribution of read-through ratios (A, C, E) and CRAC scores (B, D, F) in 18 100-nt bins around the 4881 poly(A) sites (position 0 between the dashed red lines) mapped in this study.

Data for 6 gene lists are displayed:

(A, B) Genes with Read-through ratios > 3 (blue) or < 1 (green) in at least two bins spanning nucleotide positions 0-700;

(C, D) Genes with either the GUAG or UGUA Seb1 binding motifs (blue) or neither of these (green) in at least one of three bins spanning nucleotide positions 0-300;

(E, F) Genes with a CRAC score > 100 (blue) or of 0 (green) in the bin spanning nucleotide positions 0-100.

(A, C, and E) Medians of the blue boxplots that are significantly higher than their green counterpart are coloured in red ($P_{wilcox} < 0.05$).

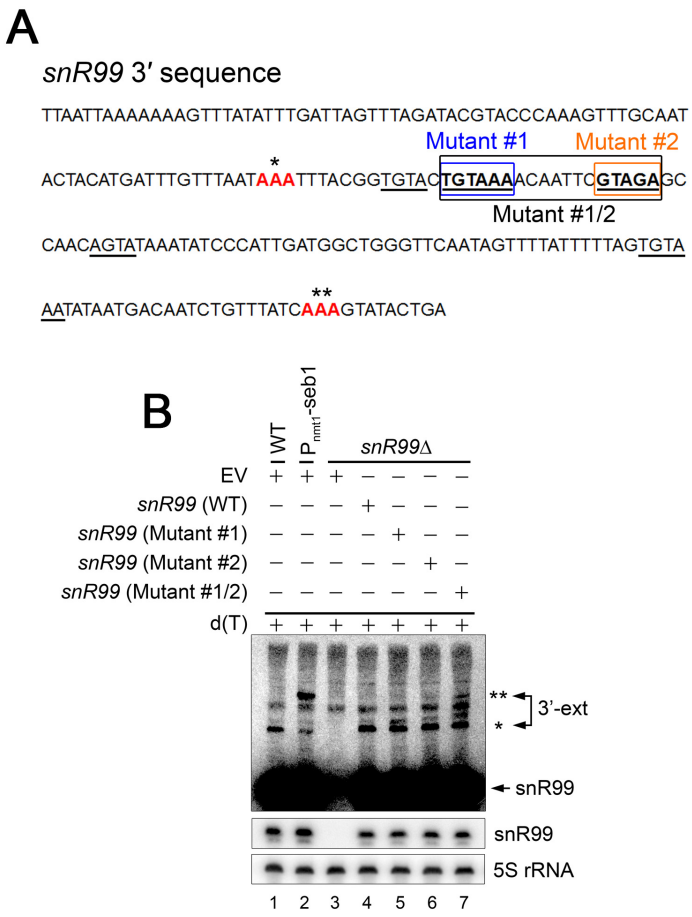

Figure S8. Functional Seb1 binding motifs are important for snoRNA 3' end processing.
(A) 200-nt of genomic sequence downstream of the *snR99* snoRNA gene with consensus motifs of the Seb1 binding site underlined. Regions indicated by colored boxes represent Seb1 binding sites that were deleted independently (blue box: Mutant #1; orange box: Mutant #2) or in combination (black box: Mutant #1/2). The red nucleotides indicate the major poly(A) site position of *snR99* polyadenylated precursors (3'-ext) detected in a wild-type (*) and in Seb1-depleted (**) cells, as determined by 3' RACE. **(B)** RNase H cleavage assay using *snR99Δ* cells (lanes 3-7) that contain an empty vector control (lane 3) or that express wild-type (lane 4) and mutant versions (lanes 5-7) of the *snR99* gene with deletions in the Seb1-binding motifs. WT (lane 1) and P_{nmt1}-seb1 (lane 2) cells transformed with an empty vector (EV) were analyzed as controls. Total RNA was treated with RNase H in the presence of a DNA oligonucleotide complementary to the *snR99* transcripts and an oligo(dT) to collapse the heterogeneous polyadenylated precursor population into a discrete product: *, normal 3' end processing site; **, 3' end processing site used in Seb1-deficient cells. Mature snR99 snoRNA is indicated by an arrow.

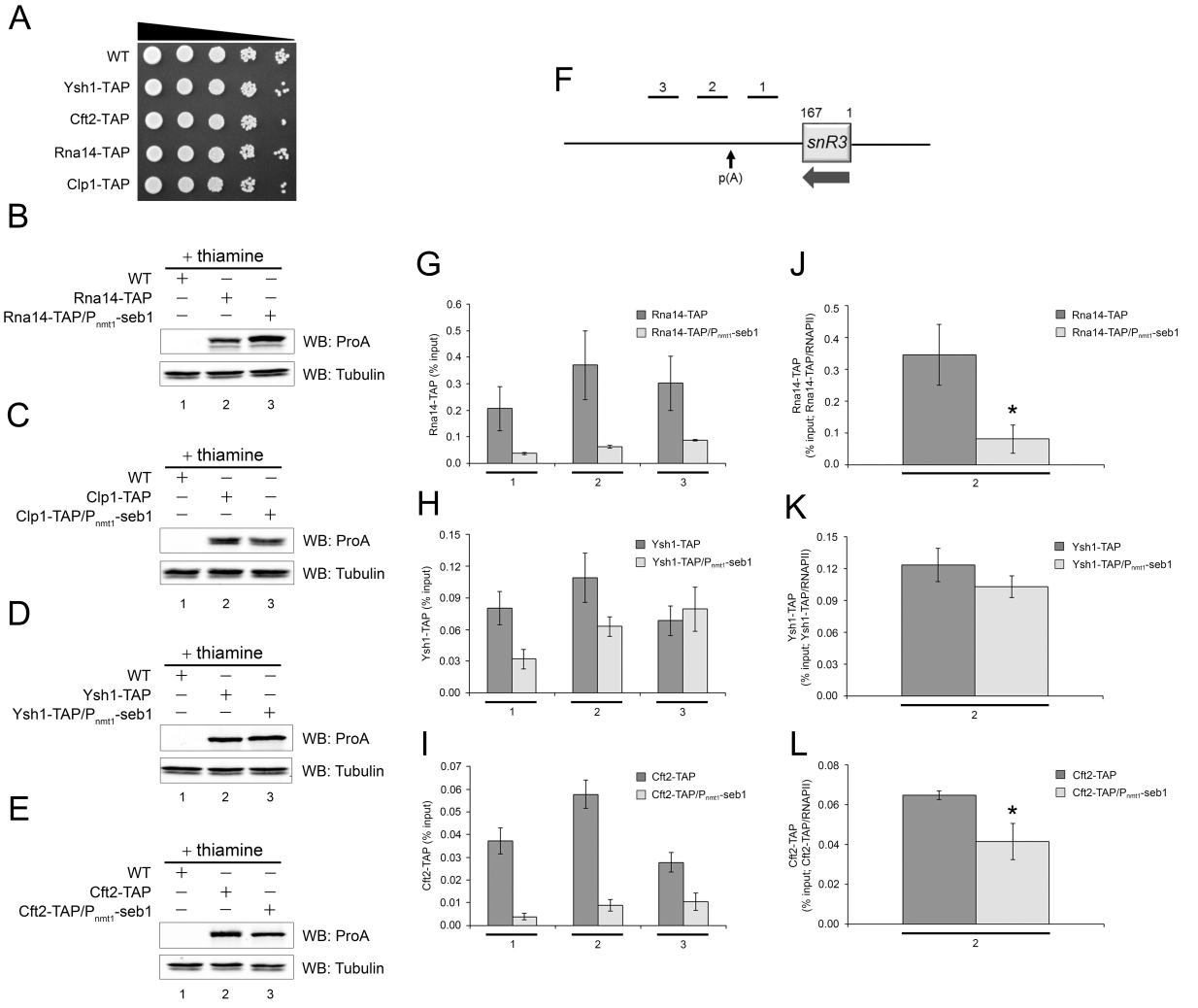

Figure S9. Recruitment of 3' end processing factors is impaired in Seb1-depleted cells.

(A) Ten-fold serial dilutions of the indicated strains were spotted on rich medium. **(B-E)** Immunoblot analyses of extracts from control untagged cells (WT) or from cells harboring a TAP-tag version of Rna14 **(B)**, Clp1 **(C)**, Ysh1 **(D)**, and Cft2 **(E)** in the wild-type or conditional Seb1 (P_{nmt1}-seb1) background. TAP-tagged proteins were revealed by an antibody directed against the protein A domain of the tag. Tubulin serves a as loading control. **(F)** Schematic representation of the *snR3* locus. Bars above the gene show the positions of PCR products used for ChIP analyses throughout Supplementary Figure 7. For *snR3*, p(A) refers to the polyadenylation sites of its 3'-extended precursors (Lemay et al. 2010). **(G-I)** ChIP assays on TAP-tagged versions of Rna14-TAP **(G)**, Ysh1-TAP **(H)** and Cft2-TAP **(I)** in wild-type (WT) or Seb1-depleted condition (for example: Rna14-TAP versus Rna14-TAP/P_{nmt1}-seb1). **(J-L)** 3'-end processing factors recruitment normalized to RNAPII density. Only amplicon 2 is shown as it represents the region where 3'-end processing factors recruitment is maximal. Error bars represent the standard deviation from three independent biological replicates. *: $p < 0.05$ (Student's t-test).

SUPPLEMENTARY TABLES

Table S1.
List of proteins identified in the Seb1-HTP purification. *See supplementary excel file.*

Table S2.
List of proteins associated with 3'-end processing and transcription termination of mRNAs identified following the purification of Seb1-HTP. *See supplementary excel file.*

Table S3.
List of genes displaying 3' UTR lengthening upon Seb1 depletion. *See supplementary excel file.*

Table S4.
List of k-mers and their associated Z-score derived from pyMOTIF analysis of Seb1-HTP CRAC experiments. *See supplementary excel file.*

Table S5.
List of strains and plasmids used in this study. *See supplementary excel file.*

SUPPLEMENTARY METHODS

Yeast strains and media

A list of all *S. pombe* and *S. cerevisiae* strains used in this study is provided in Table S5. Fission yeast cells were routinely grown at 30°C in yeast extract medium with adenine, uracil and amino acid supplements (YES) or in Edinburgh minimal media (EMM) supplemented with adenine, uracil and appropriated amino acids. Conditional strain in which the genomic copy of *seb1* is expressed from the thiamine-repressible *nmt1* promoter (P_{nmt1}) was generated as described previously(Lemieux et al. 2011). P_{nmt1}-dependent gene expression was repressed by the addition of 60 µM of thiamine to the growth medium for 10-15 hrs. To inhibit transcription elongation, 6-azauracil (6-AU) was added to the culture medium of FBY1712 and FBY2063 strains at a final concentration 0.08 mg/ml for 4h. Gene disruptions and C-terminal tagging of proteins were performed by PCR-mediated gene targeting(Bahler et al. 1998), using the lithium acetate method for cell transformation. Knockouts strains were confirmed by RT-PCR, tagged proteins by western blotting and promoter swap by PCR screening and/or growth assay. Growth of *S. cerevisiae* strains were carried out in Synthetic Defined media (SD) supplemented with 2 % galactose, adenine, uracil and appropriated amino acids with the exception of leucine. Metabolic depletion of Nrd1p and Pcf11p from the P_{GAL1}-*3X*HA-*NRD1* and the P_{GAL10}-*PCF11* strains, respectively, was performed for 7 hrs in 2% glucose-containing medium as described previously (Thiebaut et al. 2006; Haddad et al. 2012). Primer sequences used throughout this study are available upon request.

Growth (spot) assays

For growth assays, exponential cell cultures were adjusted to an optical density (OD_{600nm}) of 1.0, and serially diluted 10-fold with water. For *S. pombe* experiments, each sample was spotted (3 µl/spot) on EMM plates with or without 15 µM of thiamine and incubated at 30°C for 3 and 7 days. For *S. cerevisiae* spot assays, each sample was spotted (3 µl/spot) on SD leu- or ura- plates containing either 2% galactose or glucose and incubated at 30°C for 2-3 days.

RNA preparation and analyses

Total yeast RNA was extracted using the hot-acid phenol method, resolved on agarose-formaldehyde gels or on polyacrylamide-urea gels in case of RNase H experiments and subsequently transferred onto nylon membranes prior to RNAs crosslinking. Pre-hybridization and overnight hybridization were carried out in Church buffer at 42°C and 65°C for DNA and RNA probes, respectively. Strand-specific RNA probes were designed to be complementary to ORF sequences. RNA probes were generated by *in vitro* transcription using the T7 riboprobe system (NEB) and internally labeled with [α-32^P]-UTP. DNA probes were generated by 5' end labeling of single-stranded oligonucleotides using the T4 polynucleotide kinase and [γ-32^P]-ATP. For DNA probes, 2× 15 min washes were done in 2× SSC/0.1 % SDS followed by 2× 5 min washes in 0.1× SSC/0.1 % SDS; for RNA probes, 3× 15 min washes were performed in 1× SSC/0.1 % SDS followed by one 15 min wash in 0.1× SSC/0.1 % SDS. Membranes were exposed using Phosphor Screens followed by visualization and quantification of signals using a Typhoon Trio instrument and ImageQuant TL, respectively (GE Healthcare).

RNase H cleavage assays

RNase H assays were performed as described previously (Lemay et al. 2010).

Protein analyses

Analysis of protein expression.

Total cell extracts were prepared by harvesting mid-log cells in ice-cold lysis buffer (50 mM Tris (pH 7.5), 5 mM $MgCl_2$, 150 mM NaCl and 0.1 % NP-40) containing a cocktail of protease

inhibitors (1× PMSF, 1× PLAAC and 1× cOmplete) prior to lysis with glass beads using a FastPrep instrument (MP Biomedicals). Clarified lysates were normalized for total protein concentration using the Bradford protein assay. Proteins were separated by SDS–PAGE, transferred to nitrocellulose membranes, and analyzed by immunoblotting using either, a mouse monoclonal antibody specific to the FLAG peptide (Sigma-Aldrich, F1804; 1:500 (v/v) dilution), a mouse monoclonal antibody specific to α-tubulin (Sigma-Aldrich, T5168; 1:1,000 (v/v) dilution), a mouse monoclonal antibody against the hemagglutinin protein (HA) (Roche, 11 583 816 001; 1:500 (v/v) dilution) or a rabbit polyclonal antibody against protein A (Sigma-Aldrich, P3775; 1:10,000 (v/v) dilution) in the case of HTP-tagged Seb1 or TAP-tagged proteins. Membranes were then probed with goat anti-rabbit or anti-mouse secondary antibodies conjugated to IRDye 800CW (LI-COR, 926-32213; 1:15,000 (v/v) dilution) and AlexaFluor 680 (Life Technologies, A-21057; 1:15,000 (v/v) dilution), respectively. Detection of the proteins was performed using an Odyssey infrared imaging system (LI-COR).

Small-scale affinity purification.
For standard co-immunoprecipitation experiments, 10 mg of total proteins were subjected to immunoprecipitation with 25 µl of Pan Mouse IgG Dynabeads (ThermoFisher Scientific, 11041), that were pre-equilibrated in ice-cold lysis buffer (50 mM HEPES-KOH (pH 7.5), 5 mM $MgCl_2$, 100 mM KCl, 0.25 % NP-40, 10 % glycerol, 1 mM EDTA and 0.5 mM DTT) containing a cocktail of phosphatase and protease inhibitors (1× PhosSTOP, 1× PMSF, 1× PLAAC and 1× cOmplete), and gently mixed for 2 hrs. When needed, 1 µl of the DNA/RNA nuclease Benzonase (Sigma-Aldrich; E1014) (final concentration: 250 U/ml) was added to the mix at the beginning of the immunoprecipitation. The efficiency of the benzonase treatment was systematically tested for all protein purifications by monitoring RNA integrity by RT-qPCR on a handful of abundant RNAs. Co-immunoprecipitated proteins were eluted by heat (95°C/5 min) in 1× SDS sample buffer. Proteins were separated by SDS–PAGE, transferred to nitrocellulose membranes, and analyzed by immunoblotting using, a mouse monoclonal antibody specific to the MYC peptide (Santa Cruz Biotechnology, sc-40; 1:500 (v/v) dilution) and a rabbit polyclonal antibody against protein A. Secondary antibodies and protein detection are as described above.

Large-scale affinity purification and mass spectrometry.
A 2 L culture of WT and Seb1-HTP cells (OD_{600nm} of 1) grown in YES media was pelleted and resuspended in a volume of lysis buffer (50 mM HEPES-KOH (pH 7.5), 5 mM $MgCl_2$, 100 mM KCl, 0.25 % NP-40, 10 % glycerol, 1 mM EDTA, 0.5 mM DTT, 1× PhosSTOP, 1× mM PMSF, 1× PLAAC and 1× cOmplete) equal to one-fourth of the cell pellet volume and subsequently added dropwise to liquid nitrogen. Cells were disrupted by cryo-grinding with a Freezer/Mill 6870 (SPEX SamplePrep) using the following setup: 4 rounds of 2 min of grinding at 15 pulses/second followed by 2 min of cooling. Cell powder was resuspended in a volume of IP buffer (lysis buffer without DTT) equal to 1.2× the cell pellet volume and stirred for 15 min in a cold room. The lysate was then clarified by two successive centrifugations at 12,000 rpm at 4°C (25 and 15 min for the first and second centrifugation respectively). The supernatant of each sample was incubated for 2 hrs at 4°C on a rocker with 275 µl of Pan Mouse IgG Dynabeads that were pre-equilibrated in IP buffer. Beads were washed 3× with 1 ml of IP buffer for 1 min each followed by one 5 min wash. Proteins were eluted with freshly prepared 0.5 M NH_4OH at 37°C for 20 min and samples were dried overnight using a SpeedVac concentrator. Immunoprecipitated proteins were resuspended in 45 µl of buffer composed of 50 mM Tris (pH 6.8), 1 % SDS and 10 mM DTT and allow to reduce at 56°C for 30 min. Upon reaching room temperature, alkylating agent iodoacetamide was added to a final concentration of 50 mM and samples were incubated for 40 min in the dark prior to boiling in SDS sample buffer. The eluates were ultimately run on a 4-12 % Bis-Tris Novex mini-gel (Life Technologies) using MOPS buffer followed by Coomassie-staining (Simply Blue SafeStain; ThermoFisher Scientific). The entire

protein gel lanes were excised and cut into 8 slices. Gel bands were destained and subjected to in-gel digestion with 12.5 µg/ml of freshly prepared trypsin Gold (Promega, V5280) overnight at 30°C. Digested peptides were extracted from gels bands, dried using a SpeedVac concentrator and resuspended in 0.1 % TFA (trifluoroacetic acid). The samples were then loaded onto mini-C_{18} columns (ZipTips; EMD Millipore) and process according to manufacturer's instructions but with slights modifications: the wetting and elution solution being respectively changed to 100 % acetonitrile and 1 % FA (formic acid)/50 % acetonitrile. Eluted peptides were dried using a SpeedVac concentrator and resuspended in 1 % FA. Purified trypsin digested samples were analyzed by liquid-chromatography (LC)-MS/MS as described previously (Grenier St-Sauveur et al. 2013).

GST pull-down assays

Cell pellets from 100-ml cultures of Seb1-GFP (FBY1615), Rmt3-GFP (FBY8), and a control untagged strains that were previously transformed with a plasmid expressing either NLS-GST-CTD (Materne et al. 2015) or a control GST-fusion protein were lyzed with glass beads using a Fastprep instrument (MP Biomedicals) in ice-cold lysis buffer (see section small-scale affinity purification). Two-mg of total protein were subjected to affinity purification using 50 µl of Glutathione Sepharose beads (GE Helathcare) that were pre-equilibrated in ice-cold lysis buffer and gently mixed for 1h. Bound proteins were eluted in 1x SDS followed by SDS-PAGE separation, transfer to nitrocellulose membranes, and immunoblotting using a rabbit polyclonal anti-GST (Santa Cruz Biotechnology, 1:1000 (v/v) dilution), a mouse monoclonal anti-tubulin (Sigma-Aldrich, T5168; 1:1,000 (v/v) dilution), and a mouse monoclonal anti-GFP (Roche, 1:1000 (v/v) dilution).

***seb1/NRD1* expression constructs**

seb1 expression constructs were created by a 3-step cloning procedure using the *ade6* integration vector (Fusby et al. 2015), pFB366, as the host vector and contain 1 kbp of *seb1* 5' UTR sequences followed by the *seb1* open reading frame (ORF) and an additional 1 kbp of *seb1* 3' UTR sequences. For complementation studies, the ORF of *nrd1* was cloned between seb1's 5' and 3' UTR sequences. Insertion of a C-terminal $_{3X}$FLAG tag to both Nrd1p and Seb1 was performed using the Q5 site-directed mutagenesis kit (NEB) with primers containing the tag sequence resulting in pFB941 and pFB943 respectively. To create the C-terminal interacting domain (CID) mutant (CID_{mut}) of Seb1 (pFB944), tyrosine (Y) residue 64, aspartic acid (D) residue 67 and arginine (R) residue 71 were all changed to alanine (Y64A/D67A/R71A). For the Ser5 mutant ($Ser5_{mut}$), serine (S) residue 22 and lysine (K) residue 25 were both substituted to alanine (S22A/K25A = pFB1037). All mutations were performed by site-directed mutagenesis using pFB943 as template. The same approach was used for the RNA recognition motif (RRM) mutant (RRM_{mut}) (pFB946), where arginine (R) residue 442, histidine (H) residue 443, phenylalanine (F) residue 445 and lysine (K) residue 447 were all changed to alanine (R442A/H443A/F445A/K447A). All constructs were confirmed by sequencing. For single integration into the *ade6* locus, pFB366, pFB941, pFB943, pFB946 and pFB1037 were linearized and transformed into the conditional P_{nmt1}-seb1 strain (FBY915). Positive integrants were confirmed by western blot. For complementation studies in *S. cerevisiae*, *NRD1-$_{3X}$FLAG* and *SEB1-$_{3X}$FLAG* cDNAs were cloned between 1 kbp of *NRD1*'s 5' and 3' sequence using pRS315 (pFB950) as the host vector resulting in pFB970 and pFB1007 respectively. All constructs were confirmed by sequencing. pFB950, pFB974 and pFB1007 were maintained episomally into the conditional *P_{GAL1}-$_{3X}$HA-NRD1* strain (FBY1802) (Thiebaut et al. 2006). Expressions were confirmed by western blot.

***rps2*-GFP-*rps2* and *snR99* constructs with alteration in Seb1-binding elements**

To generate the *rps2*-GFP-*rps2* wild-type construct, GFPS65T was PCR amplified with 522 bp of *rps2* promoter and 977 bp of *rps2* terminator sequences and cloned between ApaI and SpeI sites into pFB366 to generate pFB612. Mutant versions of *rps2*-GFP-*rps2* were created by site directed mutagenesis of pFB612 using the Q5 site-directed mutagenesis. In mutant #1, the GUA core of three Seb1 motifs located upstream of the WT *rps2* cleavage site (+17 to +19, +36 to +38 and +46 to +48 from GFP stop codon) were replaced by CAC, generating plasmid pFB1141. In mutant #2, the GUA core of eight Seb1 motifs located downstream of the WT *rps2* cleavage site (+109 to +111, +116 to +118, +150 to +152, +154 to +156, +172 to +174, +179 to +181, +188 to +190 and +259 to +261 from GFP stop codon) were replaced by CAC, resulting in pFB1143. All constructs were confirmed by DNA sequencing. For single integration into the *ade6* locus, pFB612, pFB1141 and pFB1143 were linearized and transformed into a WT and/or the conditional P_{nmt1}-seb1 strains. Positive integrants were confirmed by growth selection on EMM agar plates lacking adenine and RT-PCR.

The *snR99* mutants designed to alter the binding of Seb1 to nascent snR99 transcripts (*snR99* Mutant #1, #2 and #1/2) were generated using the *ade6* integration vector (pFB366) as the host vector and contain 1 kbp of *snR99* 5' sequences followed by the *snR99* noncoding region and an additional 1 kbp of *snR99* 3' sequences, to result in pFB600. Deletion of nucleotides 91-96 (inclusive) (Mutant #1), 104-108 (Mutant #2) and 91-108 (Mutant #1/2) of the *snR99* 3' sequences were performed using the Q5 site-directed mutagenesis kit with pFB600 as template to result in pFB1057, pFB1059 and pFB1061 respectively. All constructs were confirmed by DNA sequencing. For single integration into the *ade6* locus, pFB600, pFB1057, pFB1059 and pFB1061 were linearized and transformed into a *snR99*Δ strain. Positive integrants were confirmed by growth selection on EMM agar plates lacking adenine while similar snR99 transcript level between the WT and mutated versions of *snR99* were confirmed by RT-PCR using primers located within the *snR99* noncoding region.

Chromatin immunoprecipitation (ChIP) assays

ChIP-qPCR and ChIP-seq experiments were performed as described previously (Lemay et al. 2014). In the case of P_{nmt1}-seb1 cells grown in thiamine-supplemented condition, the volume of cell culture was doubled to get sufficient RNAPII-bound chromatin for library generation. Chromatin was immunoprecipitated directly with Pan Mouse IgG Dynabeads in the case of TAP/HTP-tagged proteins or with anti-FLAG magnetic agarose beads (Sigma-Aldrich, M8823) when the Seb1-$_{3X}$FLAG strain and its CID, Ser5 and RRM mutants versions were used. ChIP assays were also performed using Pan Mouse IgG Dynabeads coated with antibodies against Rpb1 (8WG16; Convance, MMS-126R) or the hemagglutinin protein (HA) when using the Rpb3-$_{3X}$HA strain. Note that the 8WG16 antibody is known to recognize the C-terminal domain (CTD) of Rpb1 and measures total RNAPII along genes in yeast (Bataille et al. 2012). Control ChIP assays with untagged strains or with isotype matched control antibody were performed.

Transcription run-on assays

Transcription run-on assays were performed as described previously (Lemay et al. 2014)

Library preparation and Illumina sequencing

RNA-seq Libraries were prepared from total *S. pombe* RNA and sequenced with Illumina HiSeq technology as described (Lemieux et al. 2011). For ChIP-seq analyses, libraries were prepared as described previously (Lemay et al. 2014).

RNA-seq data analysis

RNA-seq data were analysed as in (Lemay et al. 2014).

ChIP-seq data analysis
Data where analysed as described in (Lemay et al. 2014). Average gene and reads accumulation analyses were performed with in house R scripts using TTS coordinates generated in this study (poly(A) mapping). TSS coordinates were retrieved from the pombase database (www.pombase.org). Heatmaps of read coverage were generated using the deepTools software (Ramirez et al. 2014) using TSS and TTS coordinates reported in the pombase database. Single gene profiles were generated using the Integrative Genomics Viewer (Robinson et al. 2011).

3' READS analysis
The 3'READS method used in this study was performed as previously described using total *S. pombe* RNA (Hoque et al. 2013).

CRAC assays, libraries preparation and Illumina sequencing
S. pombe cells were grown in YES medium to an OD_{600nM} of 0.45-0.5 and UV-irradiated in the Megatron UV cross-linker (Granneman et al. 2011) for 220 seconds. CRAC was performed as previously described (Granneman et al. 2009) with slight modifications: we used App_PE 3' adapter (5'-App-NAGATCGGAAGAGCACACGTCTG-ddC) and App_RT (5'-CAGACGTGTGCTCTTCCGATCT) oligonucleotides to make the libraries compatible with paired-end sequencing. The cDNA libraries were PCR amplified using P5 (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT) and barcoded P3 oligonucleotides (5'-CAAGCAGAAGACGGCATACGAGAT-6nt_barcode-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT) and CRAC libraries were paired-end sequenced on a HiSeq2500 at Edinburgh Genomics, University of Edinburgh. Data from two biological replicates (untagged control and Seb1-HTP samples) were subsequently merged.

***CRAC data analysis*.**
Raw fastq files were pre-processed using the pyCRAC package (Webb et al. 2014) Raw reads were demultiplexed (script pyBarcodeFilter.py) and filtered (script pyFastqDuplicateRemover.py). Processed reads were mapped to the fission yeast genome (ASM294v2.23) using the novoalign software (www.novocraft.com). Mapped files were further analysed using the pyCRAC package. Briefly, reads were mapped to the fission yeast genome annotation including all annotated coding and non-coding transcripts (www.pombase.com). Untranscribed intergenic regions were included in this genomic annotation as independent features. Taking "intergenic" features into account allows to screen for potential binding motifs enrichment downstream of transcribed units using the pyCRAC software (see below).

Reads were mapped to genome annotation:

```
pyReadCounters.py -f myReads.novo --gtf=myAnnotation.gtf -r 0 -v --unique
```

Reads were then clustered to define regions of enriched binding:

```
pyClusterReads.py -f myReads_count_output_reads.gtf --gtf=myAnnotation.gtf --
cic=10 --ch=10 --co=5 -r 0 --mutsfreq=10 -o
Clusters_0_myReads_count_output_reads.gtf
```

Significant regions, or clusters, were defined with an FDR of 5%:

```
pyCalculateFDRs.py -f Clusters_0_myReads_count_output_reads.gtf -o
FDR_Clusters_0_myReads_count_output_reads.gtf -c myGenome.txt -m 0.05 --min=5
--gtf=my.Annotation.gtf
```

Motifs significantly enriched at SEB1 binding sites were extracted from significant clusters:

```
pyMotif.py -f  FDR_Clusters_0_myReads_count_output_reads.gtf --
gtf=myAnnotation.gtf --tab=Schizosaccharomyces_pombe.ASM294v2.23.tab --
k_min=4 --k_max=8 -r 0
```

Coordinates of CRAC reads binding and of crosslinking-induced nucleotide deletions were extracted from the "myReads_count_output_reads.gtf" files (output of pyReadCounters.py) and plotted around poly(A) sites using in house R scripts as for ChIP-seq data. Statistics of read numbers binding to specific genomic features were extracted from the "myReads_hittable_reads.txt" files (output of pyReadCounters.py). Reads covering two genomic features were assigned once to each feature and therefore counted twice. Average gene analysis of CRAC data was performed as for ChIP-seq datasets using in house scripts.

SUPPLEMENTARY REFERENCES

Bahler J, Wu JQ, Longtine MS, Shah NG, McKenzie A, 3rd, Steever AB, Wach A, Philippsen P, Pringle JR. 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast* **14**: 943-951.

Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F. 2012. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* **45**: 158-170.

Fusby B, Kim S, Erickson B, Kim H, Peterson ML, Bentley DL. 2015. Coordination of RNA Polymerase II Pausing and 3' End Processing Factor Recruitment with Alternative Polyadenylation. *Mol Cell Biol* **36**: 295-303.

Granneman S, Kudla G, Petfalski E, Tollervey D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 9613-9618.

Granneman S, Petfalski E, Tollervey D. 2011. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *EMBO J* **30**: 4006-4019.

Grenier St-Sauveur V, Soucek S, Corbett AH, Bachand F. 2013. Poly(A) tail-mediated gene regulation by opposing roles of Nab2 and Pab2 nuclear poly(A)-binding proteins in pre-mRNA decay. *Mol Cell Biol* **33**: 4718-4731.

Haddad R, Maurice F, Viphakone N, Voisinet-Hakil F, Fribourg S, Minvielle-Sebastia L. 2012. An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination. *Nucleic acids research* **40**: 1226-1239.

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* **10**: 133-139.

Larochelle M, Lemay JF, Bachand F. 2012. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic acids research* **40**: 10240-10253.

Lemay JF, D'Amours A, Lemieux C, Lackner DH, St-Sauveur VG, Bahler J, Bachand F. 2010. The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. *Mol Cell* **37**: 34-45.

Lemay JF, Larochelle M, Marguerat S, Atkinson S, Bahler J, Bachand F. 2014. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat Struct Mol Biol* **21**: 919-926.

Lemieux C, Marguerat S, Lafontaine J, Barbezier N, Bahler J, Bachand F. 2011. A Pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. *Mol Cell* **44**: 108-119.

Materne P, Anandhakumar J, Migeot V, Soriano I, Yague-Sanz C, Hidalgo E, Mignion C, Quintales L, Antequera F, Hermand D. 2015. Promoter nucleosome dynamics regulated by signalling through the CTD code. *eLife* **4**: e09008.

Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**: W187-191.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature biotechnology* **29**: 24-26.

Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 5857-5864.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

Steinmetz EJ, Brow DA. 1996. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol* **16**: 6993-7003.

Steinmetz EJ, Brow DA. 2003. Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol Cell Biol* **23**: 6339-6349.

Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. 2006. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell* **23**: 853-864.

Webb S, Hector RD, Kudla G, Granneman S. 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* **15**: R8.