Supplemental Experimental Procedures

S. cerevisiae strains

For strain information please refer to Table S2.

Genome annotation used.

For all experiments described here we used *S. cerevisiae* genome version 3 (sacCer3) and *S. pombe* genome version EF2. We used the following genome sequence (http://goo.gl/1OiBKw) and gene models (http://goo.gl/PW04fg). For accurate representation of untranslated regions (UTRs) we matched experimentally derived UTR (Nagalakshmi et al., 2008) with the genome version used here.

Growth conditions and harvest

Yeast cultures were grown in complete media (YPD for *S. cerevisiae* or YES for *S. pombe*),synthetic dropout medium (SD –LEU for fast Pol II mutants) at 30°C and 250 rpm. For cell growth with galactose as carbon source, cells were grown in YP containing 1% raffinose and 2% D-galactose. Cells were harvested in exponential growth at an OD (595nm) of 0.5-0.6. For nascent RNA extraction, 1 l of cells was pelleted, washed with ice-cold PBS, quick-frozen in liquid nitrogen in 6 aliquots and kept at -80°C. For total RNA extraction, 50 ml of exponentially growing cells were pelleted and used immediately for total RNA extraction.

RNA purification

Nascent RNA was prepared as described (Carrillo Oesterreich et al., 2010) and used for SMIT, long read sequencing and nascent 3' end sequencing experiments on endogenous yeast genes and long read sequencing of the HZ18 reporter. For SMIT on the highly expressed integrated HZ18 reporter total RNA was extracted with Phenol:Chloroform:IAA, 25:24:1, pH 6.6 using the RiboPure RNA Purification Kit, yeast (Life technologies). *S. pombe* nascent RNA was isolated from chromatin analogous to the protocol in *S. cerevisiae* (Carrillo Oesterreich et al., 2010). To minimize the effect of potential changes in Pol II distributions downstream of 3'SSs, we size restricted (< 500

nt) nascent RNAs by polyacrylamide gel extraction for *S. pombe* (Churchman and Weissman, 2012). All RNA samples were treated twice with Turbo DNase (Life technologies) and purified with the RNA Clean & Concentrator-5 kit (Zymoresearch). The RNA column purification included in the RNA Clean & Concentrator-5 kit and/or the final size selection of the SMIT library (SMIT library preparation in Experimental Procedures) potentially eliminates very short transcripts. Hence, we included 3 biological replicates where nascent RNA shorter than 250 nt was isolated by polyacrylamide gel extraction (Churchman and Weissman, 2012). For isolation of those short RNAs all column purifications were substituted by ethanol precipitation at -80°C and no size selection prior sequencing was performed.

Removal of polyA+ RNA

Oligo-dT coated cellulose was used (MicroPolyA Purist kit, Life technologies) and polyA- RNA was separated from polyA+ RNA following the manufacturer's instructions. For nascent 3' end sequencing, polyA- RNA was obtained using oligo-dT coated magnetic beads binding to polyA+ RNA (Dynabeads mRNA DIRECT Micro Purification Kit, Life technologies).

Qualitative and quantitative analysis of nucleic acids

RNA and DNA samples were analyzed by agarose (1-1.5%) or TBE-Urea polyacrylamide (10 or 15%, Invitrogen) gel electrophoresis. DNA and RNA concentrations were determined by UV/Vis spectroscopy with the Nanodrop2000 (ThermoScientific) or fluorometric measurements with the Qubit dsDNA BR Assay or the RNA BR Assay (Life technologies). Paired-end and long read sequencing was done after Bioanalyser, Qubit dsDNA BR assay and Kappa library quantification.

Nascent 3' end sequencing library preparation and sequencing

3' end ligated nascent RNA was heat-fragmented and then reverse transcribed using the SMIT DNA adaptor sequence as RT primer. All library preparation steps, including second strand synthesis with dUTPs, end-repair, A-tailing, adapter ligation and second strand digestion by Uracil-DNA-Glycosylase, were performed according to in-house

protocols of the Yale Center for Genome Analysis, which are available upon request. The samples were sequenced using 75bp paired end sequencing on an Illumina HiSeq 2500 according to Illumina protocols.

Data processing, mapping and analysis

The R and bash shell scripts can be found on github (https://github.com/carrillo/SMITproject.git).

Long read sequencing data processing and mapping

Pacific Biosciences transcriptome data were obtained in Fastq-format. 3' end linker sequences, Clontech adaptor sequences (SMARTer cDNA synthesis kit, Clontech) and the 5 nt random 3' barcode were removed with cutadapt (Martin, 2011) and the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Processed reads were mapped to the *S. cerevisiae* genome using gmap (Wu and Watanabe, 2005). To ensure the analysis of full-length transcripts, only reads mapped within +/-200 nt of annotated transcription start sites were retained. To remove potential mRNA contaminants reads ending within +/-100 nt of an annotated polyA site and short polyA tails (> 4 nt) were removed from the dataset.

SMIT data processing and mapping.

Fastq files were filtered for read quality with the FASTX toolkit and 3' end linker were trimmed with cutadapt in forward and reverse reads. PCR duplicates were removed with prinseq (Schmieder and Edwards, 2011), followed by 5 nt random 3' barcode removal with the FASTX toolkit. The 3' end reads were mapped with tophat2 (Kim et al., 2013) to the *S. cerevisiae* genome. To infer splicing state, the junction reads were mapped with bowtie2 (Langmead and Salzberg, 2012) to bowtie indices designed to contain spliced (exon-exon, EEJ) and unspliced (exon-intron-exon, EIJ) sequences. After quantifying correlation of biological replicates with respect to the Pol II position counts (Figure S1H), samples were pooled for position-based splicing analysis.

Nascent 3’ end sequencing data processing and mapping.

Fastq files were filtered for read quality with the FASTX toolkit and 3’ end linker were trimmed with cutadapt in forward and reverse reads. Only reads containing the 3’ end linker were kept. The 5 nt random 3’ barcode was removed (FASTX). Paired-end reads were mapped with tophat2. Nascent 3’ end reads were extracted from mapped paired-end data using samtools (Li et al., 2009). Nascent RNA 3’ end coverage profiles were generated using a custom shell script (https://github.com/carrillo/SMITproject) and subsequent alignment over terminal exons of the 87 SMIT genes was done using bedtools (Quinlan and Hall, 2010) and R. For visualization of Pol II density over terminal exons positions were grouped in 20 nt bins, total read count determined and values color-coded (Log10 scale, see color map). Individual rows represent terminal exons (padded by 100 nt up- and downstream) sorted by increasing length.

Splicing reporters

The pHZ18 splicing reporter constructs with consensus 5’ SS sequences as well as MS2-HA fusion protein expressing plasmid were a generous gift from the Rosbash lab (Lacadie et al., 2006). Single nucleotide substitutions were introduced into reporter constructs by site directed mutagenesis (QuikChange Kit, Strategene) using primers given in Table S3. Reporters were integrated into the URA3 locus using standard experimental procedures, integration primers are given in Table S3.

Chromatin immunoprecipiation (ChIP)

For MS2-protein ChIP, yeast strains containing either integrated or episomal splicing reporter constructs, were transfected with MS2-HA protein expression plasmid (*pHA-MS2*). Alternatively, for splicing factor ChIP, yeast strains containing either integrated or episomal splicing reporter constructs in combination with c-terminally tagged (HA) endogenous protein coding splicing factor genes were used. ChIP experiments were carried out similarly to (Abruzzi et al., 2004). A 200 ml yeast culture was grown in inducing condition (2% galactose) to OD600 = 0.5 and crosslinked by addition of 20 ml formaldehyde solution (11% formaldehyde, 0.1 M NaCl, 1 mM EDTA, 50 mM HEPES-KOH pH 7.5) for 20 minutes. Crosslinking was stopped by addition of 30 ml quenching

solution (3 M glycine, 20 mM Tris) and incubation at 30°C for 5 minutes. Cells were harvested by centrifugation (4°C, 5 min at 1100g, JLA 8.1000) and washed twice with ice-cold 200 ml PBS and once with 10 ml FA lysis buffer (0.1% SDS, 50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate). Cell pellets were snap frozen in liquid nitrogen, resuspended in 1 ml ice-cold FA lysis buffer and transferred to a 2 ml plastic tube containing ~1.5 ml glass beads. Cells were lysed by 15 ∗ 30 sec pulses of bead-beating interrupted by 30 sec pauses on ice. Beads were washed with 3 ml FA lysis buffer and separated from the lysate by centrifugation. The lysate was transferred to two 2 ml plastic tubes, centrifuged (4°C, 10000 g, 10 min) and the chromatin pellet washed twice with 4 ml FA lysis buffer. The pellet was resuspended in 3 ml FA lysis buffer and transferred to a 15 ml plastic tube. Sonication was performed to yield DNA fragments of ~ 200 nt. Sonicated lysate was centrifuged (4°C, 10000 g, 10 min). 800 µl of the supernatant were transferred to a fresh 1.5 ml plastic tube, for each experiment and NaCl solution (4 M) added to a final concentration of 275 mM. 10 µl of protein A beads (Invitrogen) were prewashed with 1 ml FA lysis buffer (4 min, rotating wheel at room temperature). 15 µl anti-HA antibody (12CA5, Abcam) and 200 µl TE pH 8.0 (10 mM Tris-Cl pH 8.0, 1 mM EDTA) was added and incubated at 4°C for 1 h (rotating wheel). After washing with 1 ml TE pH 8.0, 750 µl chromatin sample was added and incubated at 4°C overnight (rotating wheel). 50 µl of the chromatin sample were left untreated (input). After incubation, beads were washed (4 min, rotating wheel, followed by 1 min at 1000 rpm) with i) 1.4 ml FA lysis buffer 275 mM NaCl, ii) 1.4 ml FA lysis buffer 500 mM NaCl, iii) 1.4 ml washing buffer 3 (10 mM Tris-Cl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% sodium deoxycholate) and iv) 1.4 ml TE. 250 µl elution buffer (50 mM Tris-Cl pH 7.5, 10 mM EDTA, 1% SDS) was added to the pellet and incubated at 65°C for 10 min. The supernatant was transferred to a fresh 1.5 ml plastic tube and the beads washed with 250 µl TE. For the input sample, only 500 µl TE was added. 15 µl pronase solution (20 mg/ml, Sigma) was added and incubated at 42°C for 1 h, followed by incubation at 65°C for 4 h. DNA was purified (QIAPrep Spin Miniprep, Qiagen) and eluted with 100 µl elution buffer (Qiagen). Quantification was performed by qPCR with primer pairs listed

in (Table S3). Signals gained by qPCR were corrected for primer efficiency and normalized to values gained from input samples.

List of Supplemental Tables (provided separately):

Table S1 List of gene-specific parameters
Table S2 List of yeast strains
Table S3 List of oligonucleotides
Table S4 List of gene-specific parameters fast Pol II

Supplemental References

Abruzzi, K.C., Lacadie, S., and Rosbash, M. (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. The EMBO journal *23*, 2620-2631.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent rna reveals transcriptional pausing in terminal exons. Molecular Cell *40*, 571-581.

Churchman, L.S., and Weissman, J.S. (2012). Native elongating transcript sequencing (NET-seq). Curr Protoc Mol Biol *Chapter 4*, Unit 4 14 11-17.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

Lacadie, S.A., Tardiff, D.F., Kadener, S., and Rosbash, M. (2006). In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. Genes and Development *20*, 2055-2066.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Martin, M. (2011). Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal *17*, 10--12.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science (New York, NY) *320*, 1344-1349.

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

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863-864.

Wu, T.D., and Watanabe, C.K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics *21*, 1859-1875.