Supplemental Material – Herzel, Straube and Neugebauer

Table of contents

1. Supplemental figures and legends

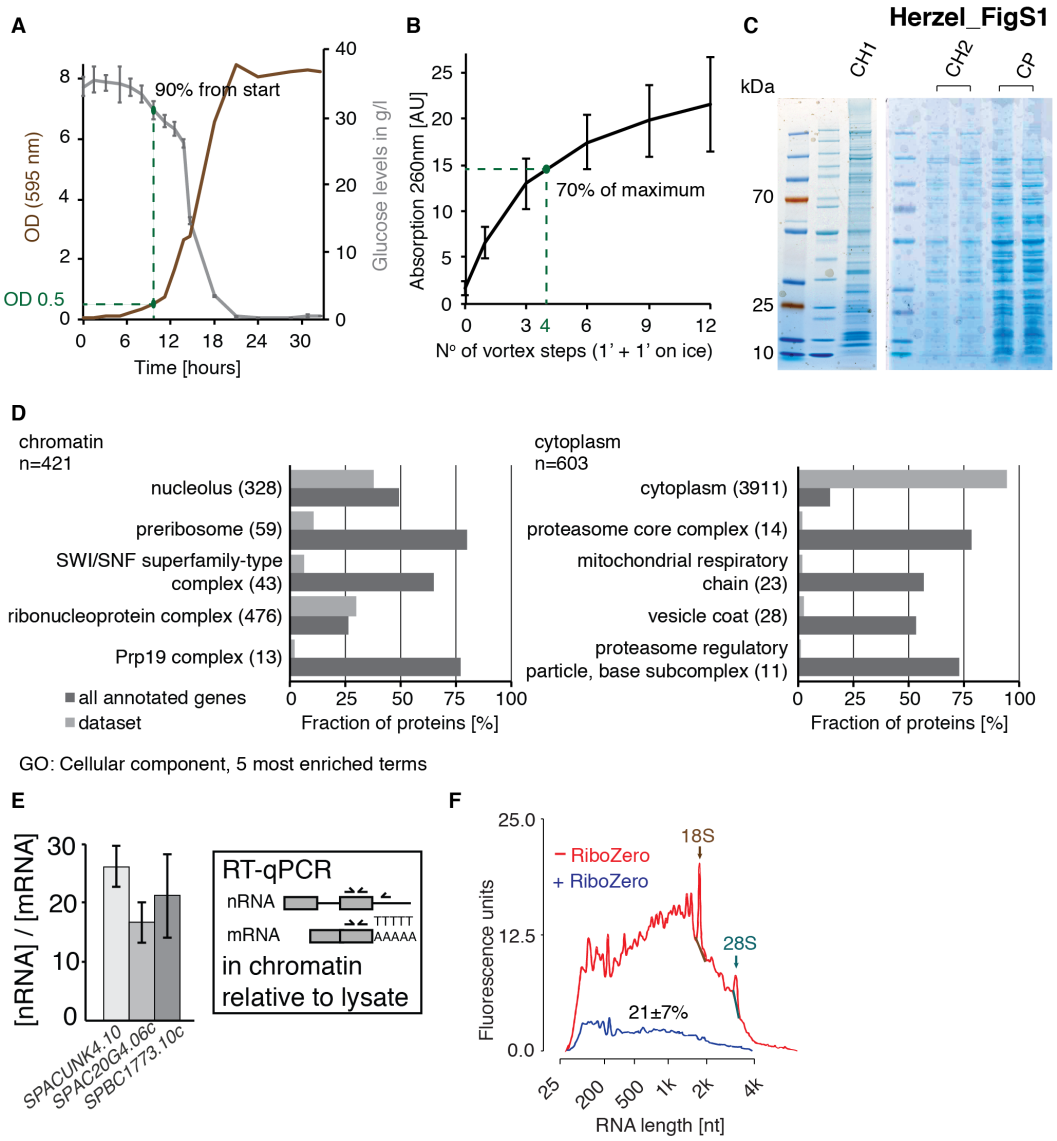


Figure S1: Characterization of the *S. pombe* chromatin fraction from exponentially growing cell cultures. (A) *S. pombe* growth curve in YES medium and associated Glucose consumption profile (SD of technical replicates (n=3) shown). *S. pombe* cells are generally harvested at OD 0.5 after at least one cell number duplication. (B) Nucleotide content of cell lysate after different vortex intervals. After 4 vortex intervals ~70 % of cells are lysed compared to lysis in 16 vortex intervals (SD is shown, n=5). (C) Coomassie-stained SDS-PAGE of 2 chromatin samples (CH1, CH2) and 1 cytoplasmic sample (CP) submitted to mass spectrometry analysis. (D) GO analysis of proteins identified by mass spectrometry in the chromatin and cytoplasmic fraction. GO analysis was performed with topGO package in R. The five most enriched terms for "Cellular component" are shown. The number of proteins annotated for each term is given in brackets. (E) RT-qPCR to measure nascent RNA enrichment over mRNA for 3 genes after poly(A)$^+$ mRNA depletion of the chromatin fraction. RT primers downstream of the polyA cleavage sites target nascent RNA, and oligo(dT) RT primer targets mRNA, respectively (SEM is shown, n=3-6). (F) Bioanalyzer traces of non-rRNA depleted nRNA sample (-RiboZero) and rRNA depleted nascent RNA sample (+RiboZero). 18S and 28S rRNA peaks are indicated, as well as the mass fraction of nRNA, which is kept after rRNA depletion (mean and SD, n=7).

Herzel_FigS2

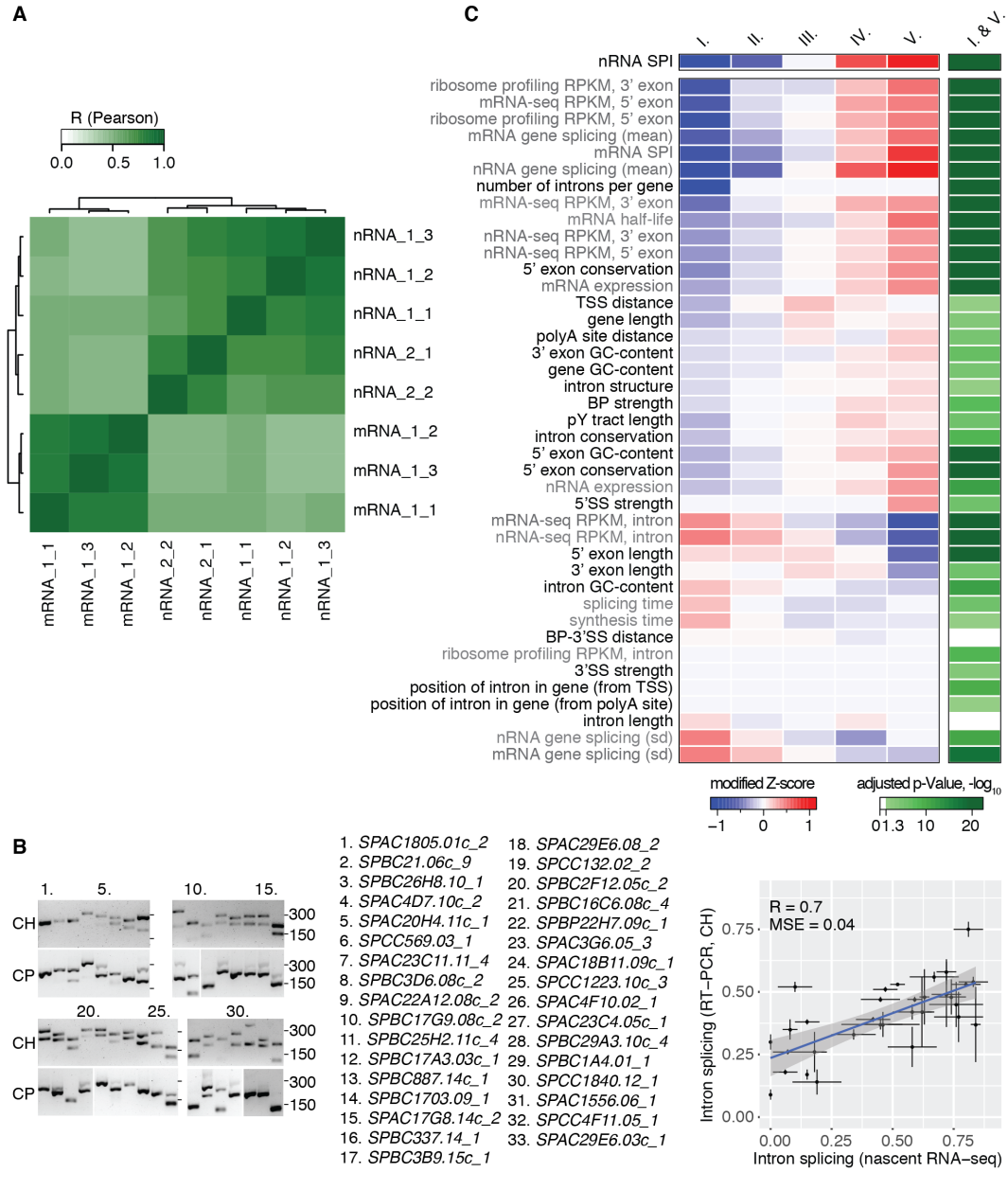


Figure S2: Intron splicing quantification from nascent RNA-seq data. (A) Heatmap of Pearson correlation coefficients of the splicing per intron values (SPI) derived from 3 mRNA-seq and 5 nRNA-seq replicates (nRNA-seq data from 2 different WT strains). 2,129 introns were present in all datasets and used for the correlation. (B) Validation of the co-transcriptional intron splicing quantification by RT-PCR. Left panel, agarose gel electrophoresis after RT-PCR of 33 randomly selected introns quantified by nRNA-seq (sorted from lowest to highest nRNA-seq intron splicing value). CH – chromatin-associated, non-polyadenylated RNA. CP – cytoplasmic RNA. Right panel, correlation of nascent RNA-seq splicing estimates (SPI) and RT-PCR splicing estimate. Black crosses reflect standard deviations. Linear regression (blue line and grey shadow) and Pearson correlation (R) were performed. Values of the RT-PCR quantification of 3 biological replicates are given in Table S7. (C) 38/41 intron-specific features correlated significantly between lowest and highest spliced introns, highlighting aspects of gene architecture, sequence, expression and transcript dynamics associated with pre-mRNA splicing. Introns were grouped into quintiles according their nRNA SPI: I. SPI [0.0-0.40], II. (0.40-0.54], III. (0.54-0.63], IV. (0.63-0.73], V. (0.73-1.00]. Each row of the heatmap shows the median modified Z-score of one intron feature for each quintile (sequence-based feature in black font and RNA-seq feature derived in grey font; blue – smaller than global median, red – greater than global median). The adjusted p-value of significant differences between group I and V are plotted in green (Wilcoxon-rank sum test, Bonferroni correction). Compare Fig. S3 for boxplots of all features and Supplemental Methods and Table S6, S9 for more detailed information.

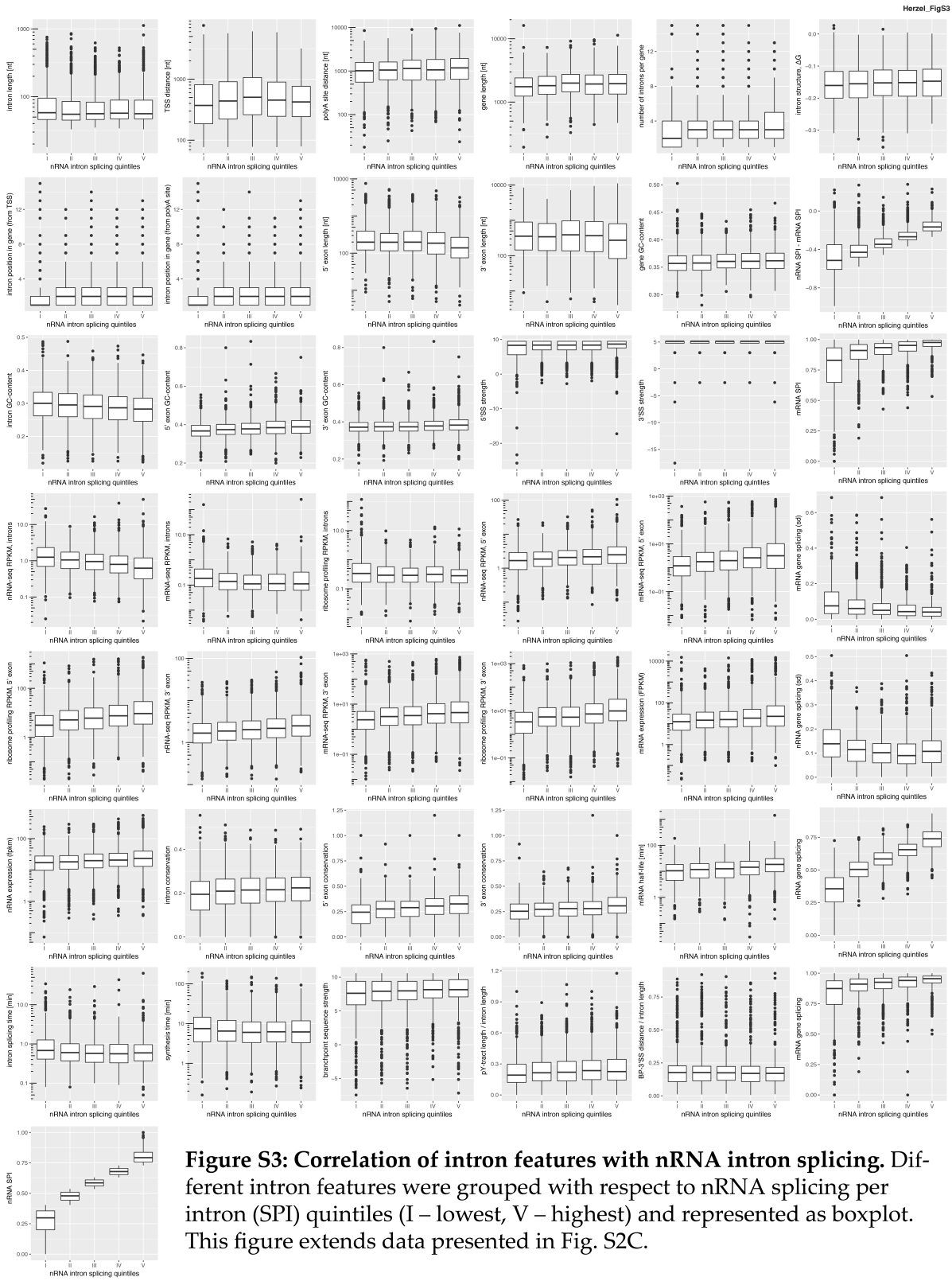
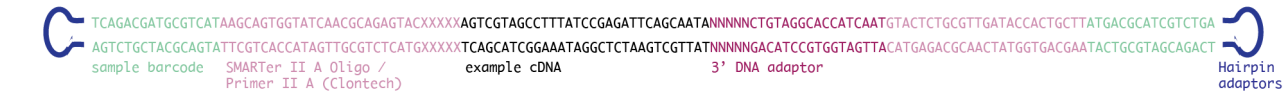
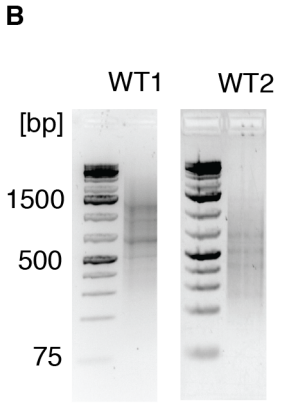
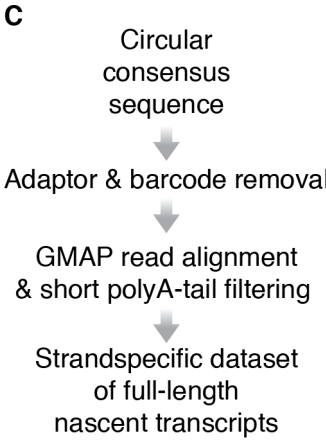
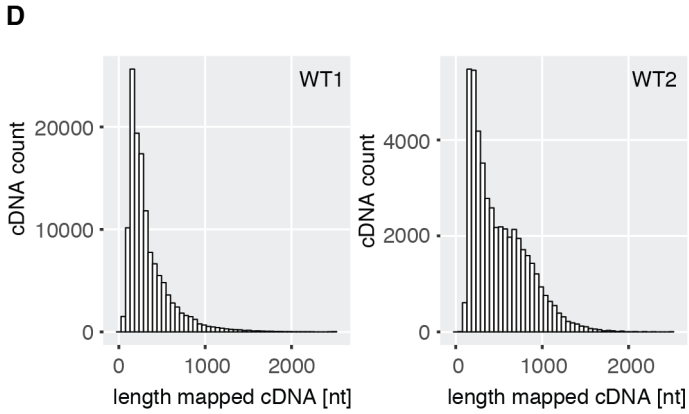


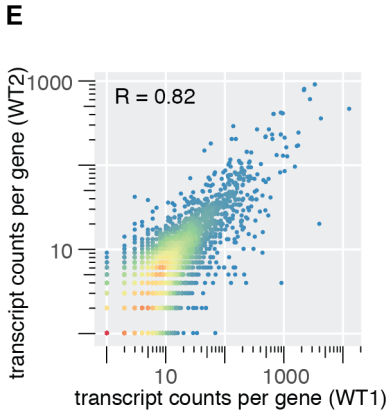
Figure S3: Correlation of intron features with nRNA intron splicing. Different intron features were grouped with respect to nRNA splicing per intron (SPI) quintiles (I – lowest, V – highest) and represented as boxplot. This figure extends data presented in Fig. S2C.

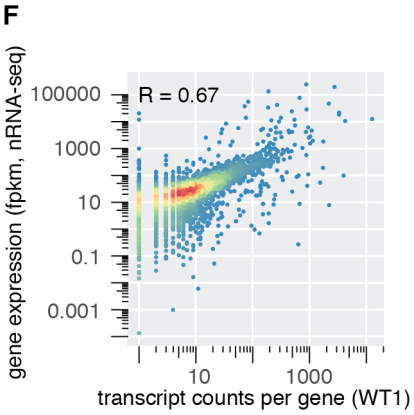


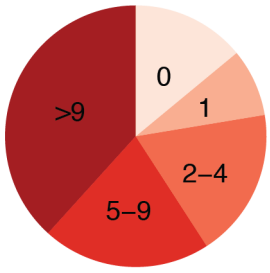




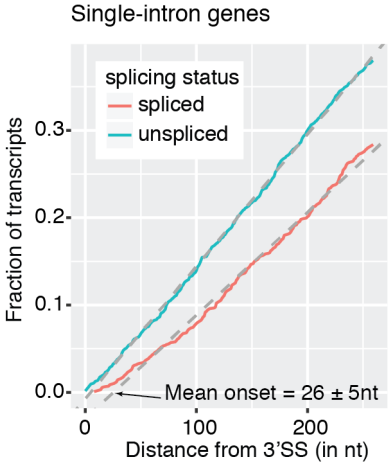








Fraction of genes with n reads



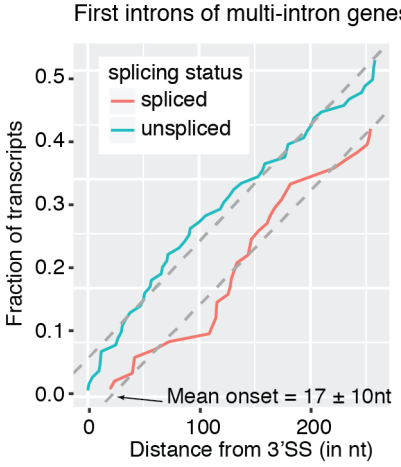


Figure S4: Long read sequencing library preparation, data processing and characterization. (A) Adaptor design and sequence details for complex, strand-specific nascent RNA Pacific Biosciences sequencing library (B) Two *S. pombe* double-stranded cDNA libraries are shown (1.5% agarose gels, final double-stranded cDNA). (C) Schematic of post-sequencing processing steps to remove adaptors, ensure strandedness and map transcripts back to the genome. (D) Sequenced and mapped transcript length distribution of the samples shown in B. Six SMRT cells with diffusion-loading were run for sample WT1 and 2 SMRT cells (1 diffusion- and 1 magbead-loading) for WT2, which explains the difference in observed read length. (E) Correlation between transcript counts per gene for WT1 and WT2 LRS samples. Pearson correlation coefficient (R) is given. (F) Correlation between the gene expression values from short read Illumina nascent RNA-seq data and the read counts per gene of the full-length nascent RNA LRS data. (G) Pie chart illustrating read counts per gene (WT1 and WT2 replicate combined). (H) Cumulative distribution of Pol II positions of spliced and unspliced long read observations for single intron genes and first introns in multi-intron genes. A model assuming uniform sampling is indicated (grey dashed line). Splicing onset inferred by the model (x-intercept: mean ± 95% confidence interval) is indicated and not significantly different between single intron genes and first introns in multi-intron genes.

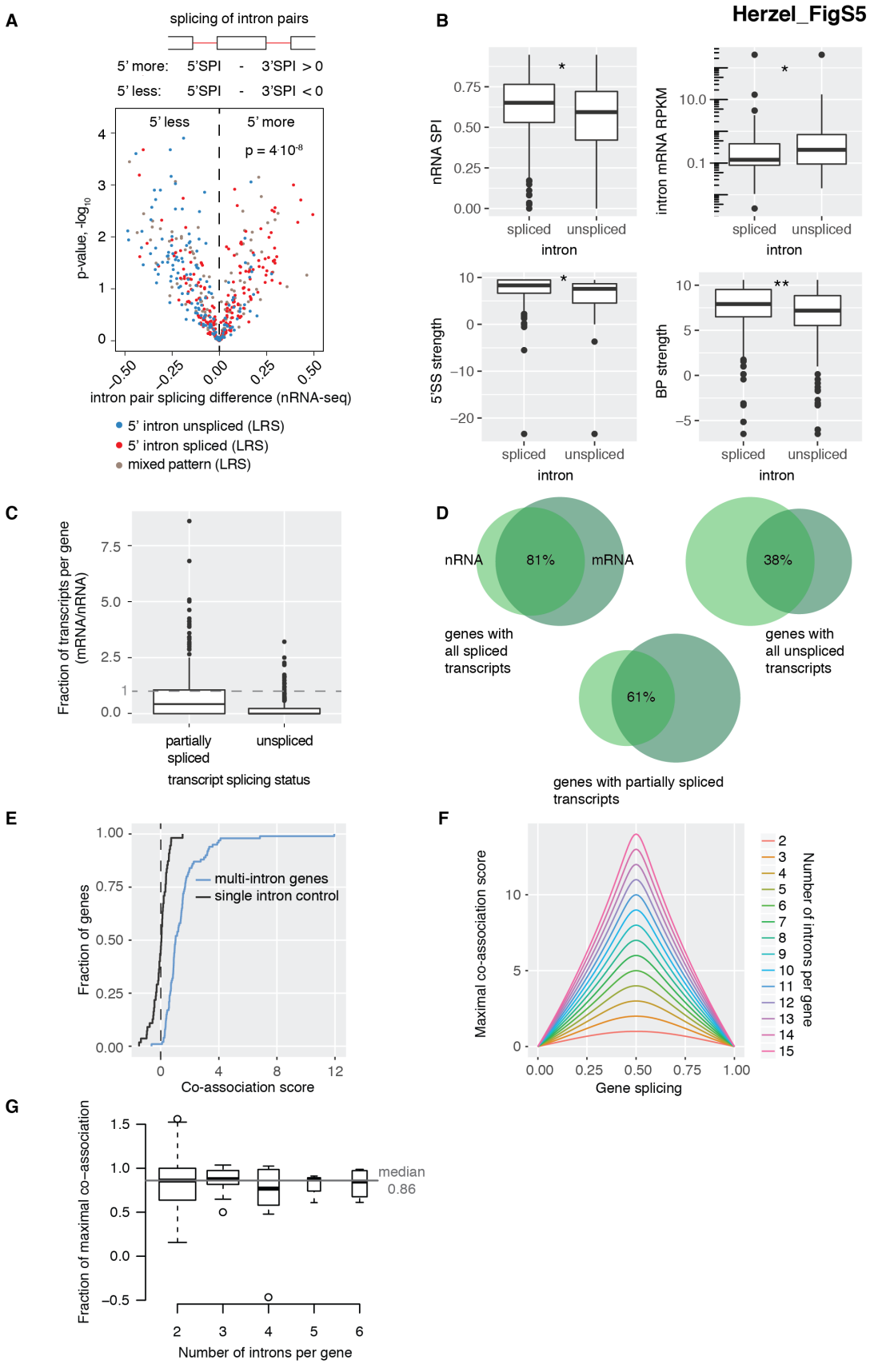

Herzel_FigS5
A
splicing of intron pairs
5' more: 5'SPI - 3'SPI > 0
5' less: 5'SPI - 3'SPI < 0
5' less
5' more
p = 4·10-8
p-value, -log10
intron pair splicing difference (nRNA-seq)
5' intron unspliced (LRS)
5' intron spliced (LRS)
mixed pattern (LRS)
B
nRNA SPI
intron mRNA RPKM
5'SS strength
BP strength
spliced
unspliced
intron
C
Fraction of transcripts per gene (mRNA/nRNA)
partially spliced
unspliced
transcript splicing status
D
nRNA
81%
mRNA
genes with all spliced transcripts
38%
genes with all unspliced transcripts
61%
genes with partially spliced transcripts
E
Fraction of genes
multi-intron genes
single intron control
Co-association score
F
Maximal co-association score
Gene splicing
Number of introns per gene
G
Fraction of maximal co−association
median
0.86
Number of introns per gene

Figure S5: Characteristics of (un)spliced introns in partially spliced transcripts. (A) Overlap of the intron splicing status in partially spliced transcripts with the differences in co-transcriptional splicing of adjacent introns (nRNA-seq). Volcano plot shows splicing differences between intron pairs (x-axis, nRNA-seq) and the associated negative $\log_{10}$ of the p-value (n=3, Student's t-test) for intron pairs also detected by LRS. Analogous to Fig. 2C. The color code reflects whether the 5' intron of the intron pair is spliced or unspliced in the LRS data. 'Mixed pattern' refers to intron pairs, where the 5' intron is spliced in some and unspliced in other partially spliced in LRS reads. (B) 4/42 intron features correlate significantly between the spliced and unspliced introns detected in partially spliced transcripts (LRS data). Asterisks indicate significance according to the Wilcoxon-rank sum test and Bonferroni-correction ($p < 0.05$ *, $p < 0.01$ **). (C) Relative abundance change from nRNA to mRNA of partially or completely unspliced transcripts. Boxplot depicting the fold-change of mRNA to nRNA of partially or completely unspliced transcript fraction per gene. Partially spliced transcripts decrease in abundance to 43% (median 0.43) and completely unspliced transcripts are not detected in most cases (median 0.00; Wilcoxon-rank sum test p-value<2e-16). The mRNA LRS data were reanalyzed from (Kuang et al. 2017). (D) Overlap of nRNA and mRNA LRS data with respect to multi-intron genes associated with the three transcript classes 'all spliced', 'all unspliced' and 'partially spliced'. The percent of overlap is given for genes identified in the nRNA LRS data. (E) Splicing of introns in individual genes is co-associated. For the 100 genes in 4B a co-association score was calculated as the $\log_2$-fold change of the observed to predicted fraction of "all or none" splicing. As a negative control, co-association estimates were calculated for 55 pairs of single intron genes with matched read counts in the LRS data. As the paired single intron genes reside in different locations in the genome, their splicing should not be co-associated as for multi-intron genes. The variation around 0 reflects experimental noise in the splicing quantification by LRS and the nRNA SPI. (F) The co-association score increases with the number of introns per gene, as the difference to the prediction becomes more pronounced. Maximal possible co-association scores $\log_2$(observed/predicted "all or none" splicing) are plotted with respect to possible gene splicing values ranging from 0-1 for genes with different number of introns. For an observed "all or none" splicing fraction of 1 the maximal score is reached. Here, the predicted fraction of "all or none" splicing is approximated assuming identical SPIs within a gene ($spi_1=spi_2=\ldots-spi_n$=gene splicing), which is conceivable for most intron-containing genes (see Fig. 2C). (G) Co-association does not depend on the number of introns per gene. The distribution of the fraction of maximal co-association is presented as boxplot with respect to the number of introns per gene. The fraction of maximal co-association represents the ratio of the co-association score for a given gene to the maximal possible co-association value at the average gene splicing value (see Fig. 4D for 2-,3-, and 4-intron gene values). The median for the 100 analyzed genes is 0.86 (grey line).

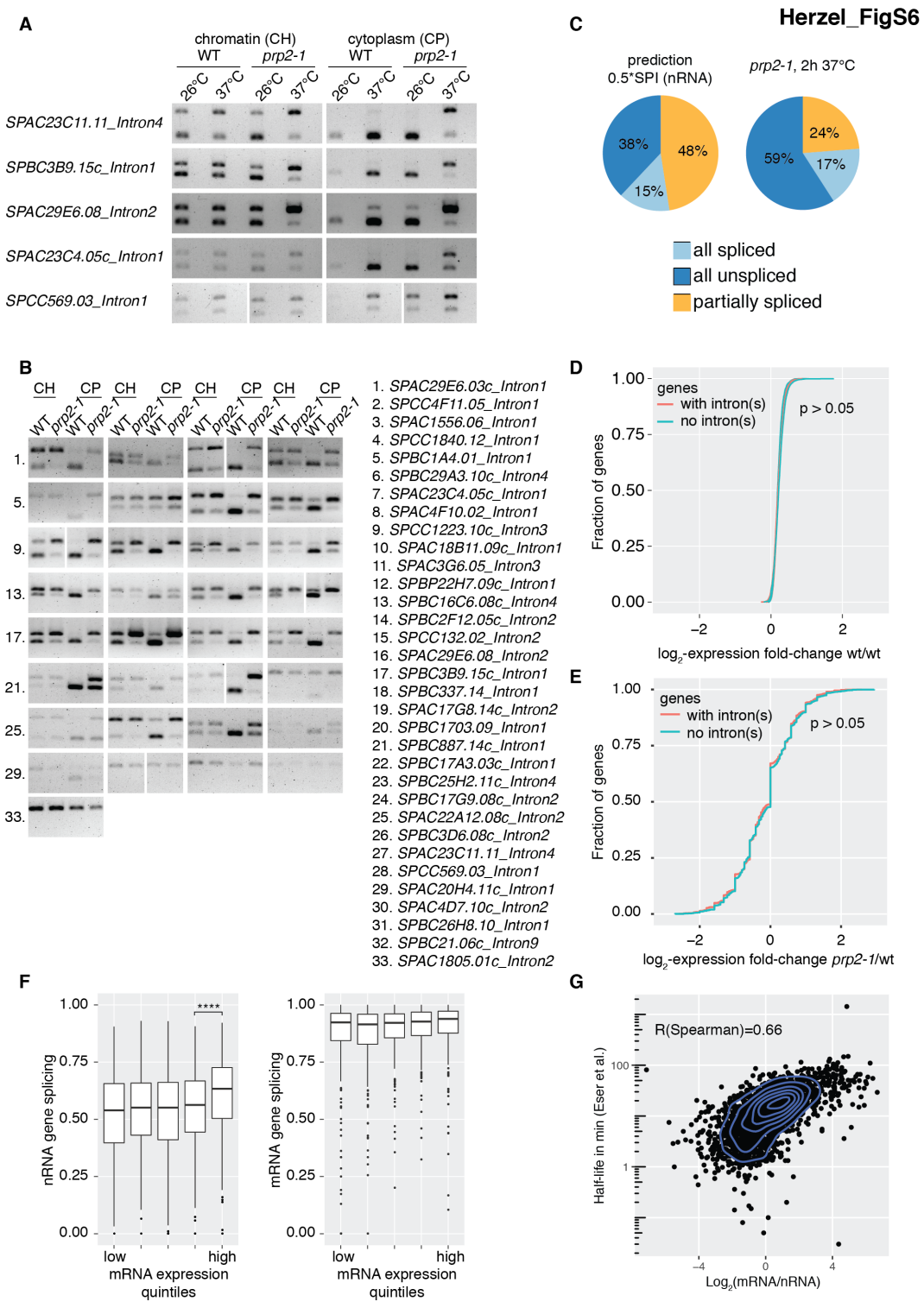

Herzel_FigS6
A
chromatin (CH)
cytoplasm (CP)
WT
prp2-1
26°C
37°C
SPAC23C11.11_Intron4
SPBC3B9.15c_Intron1
SPAC29E6.08_Intron2
SPAC23C4.05c_Intron1
SPCC569.03_Intron1
C
prediction
0.5*SPI (nRNA)
prp2-1, 2h 37°C
38%
48%
15%
24%
59%
17%
all spliced
all unspliced
partially spliced
B
CH
CP
WT
prp2-1
1.
5.
9.
13.
17.
21.
25.
29.
33.
1. SPAC29E6.03c_Intron1
2. SPCC4F11.05_Intron1
3. SPAC1556.06_Intron1
4. SPCC1840.12_Intron1
5. SPBC1A4.01_Intron1
6. SPBC29A3.10c_Intron4
7. SPAC23C4.05c_Intron1
8. SPAC4F10.02_Intron1
9. SPCC1223.10c_Intron3
10. SPAC18B11.09c_Intron1
11. SPAC3G6.05_Intron3
12. SPBP22H7.09c_Intron1
13. SPBC16C6.08c_Intron4
14. SPBC2F12.05c_Intron2
15. SPCC132.02_Intron2
16. SPAC29E6.08_Intron2
17. SPBC3B9.15c_Intron1
18. SPBC337.14_Intron1
19. SPAC17G8.14c_Intron2
20. SPBC1703.09_Intron1
21. SPBC887.14c_Intron1
22. SPBC17A3.03c_Intron1
23. SPBC25H2.11c_Intron4
24. SPBC17G9.08c_Intron2
25. SPAC22A12.08c_Intron2
26. SPBC3D6.08c_Intron2
27. SPAC23C11.11_Intron4
28. SPCC569.03_Intron1
29. SPAC20H4.11c_Intron1
30. SPAC4D7.10c_Intron2
31. SPBC26H8.10_Intron1
32. SPBC21.06c_Intron9
33. SPAC1805.01c_Intron2
D
genes
with intron(s)
no intron(s)
p > 0.05
Fraction of genes
1.00
0.75
0.50
0.25
0.00
−2
0
2
log2-expression fold-change wt/wt
E
genes
with intron(s)
no intron(s)
p > 0.05
Fraction of genes
log2-expression fold-change prp2-1/wt
F
mRNA gene splicing
****
low
high
mRNA expression
quintiles
G
R(Spearman)=0.66
Half-life in min (Eser et al.)
100
1
−4
0
4
Log2(mRNA/nRNA)

Figure S6: Global reduction of pre-mRNA splicing levels in the temperature-sensitive mutant of U2AF65 (*prp2-1*). (A) Splicing inhibition in *prp2-1* is detected in cytoplasmic RNA and chromatin-associated RNA by RT-PCR. Empty lanes reflect -RT enzyme controls. Less RNA was used as RT input in the CP 26°C WT sample, hence the RT-PCR signal is lower in this fraction. (B) RT-PCRs of the same 33 genes as in Fig. S2 (sorted by decreasing splicing levels according to nRNA-seq data). RNA was extracted from WT and *prp2-1* cells grown for 2 hours at 37°C. Cytoplasmic and chromatin-associated RNA were reverse transcribed using random hexamers and PCR was done with gene-specific primers (Table S2). (C) Pie chart extending data shown in Fig. 4C. Prediction of the three transcript classes was done as in Fig. 3B, but with nRNA SPI reduced by 50% to mimic splicing global splicing inhibition in *prp2-1*. The *prp2-1* data are the same as in Fig. 4C for comparison. (D) mRNA expression changes between intron-containing and intronless genes and two wild-type replicates are not significantly different (data from (Lipp et al. 2015), Kolmogorov-Smirnov test). Cumulative distribution of expression changes between WT replicates for intron-containing and intronless genes. Plot dimensions are as shown in Fig. 4E. (E) nRNA expression, i.e. transcription, changes between intron-containing and intronless genes are not significantly different (Kolmogorov-Smirnov test). Cumulative distribution of differences in transcript count changes between WT and *prp2-1* nascent RNA LRS samples for intron-containing and intronless genes. (F) Highest expressed genes show highest co-transcriptional gene splicing. Average gene splicing values were grouped into quintiles according their mRNA expression value (quantified with Cufflinks). Strong difference in splicing (Wilcoxon-rank sum test and Bonferroni-correction $p<0.0001$) is observed between the 4^{th} and 5^{th} quintile in nRNA gene splicing. No significant correlation between mRNA gene splicing and expression are detected. (G) Positive correlation of mRNA half-life (data from (Eser et al. 2016)) with $\log_2$-fold change between mRNA and nRNA levels.

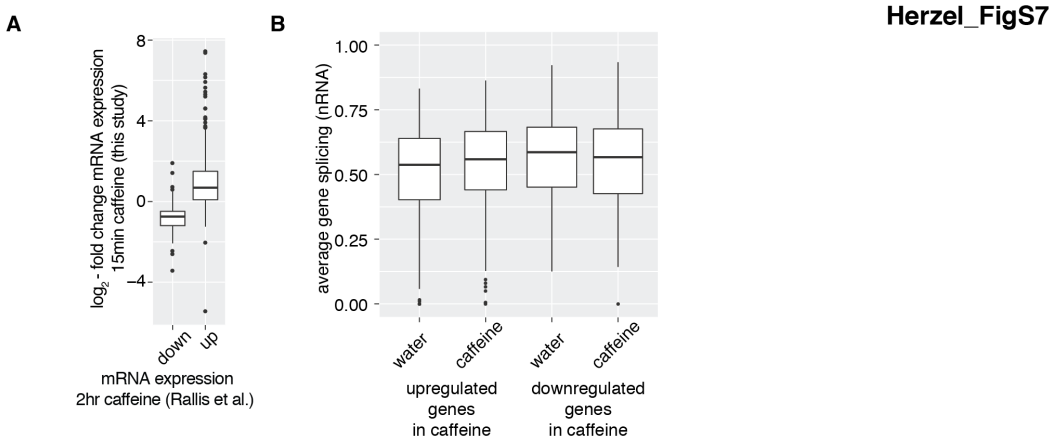


Figure S7: Gene splicing and expression changes in *S. pombe* cells upon caffeine treatment. (A) Comparison of gene expression changes with published data. Genes were grouped according to the classification of down- and upregulated genes after 2 hour treatment with caffeine (Rallis and Bahler 2013) and the respective distribution of mRNA expression changes after 10mM caffeine treatment for 15 minutes is plotted. (B) Co-transcriptional gene splicing distributions in up- and downregulated genes upon caffeine treatment. No significant splicing differences are detected (Cuffdiff quantification of expression differences, Wilcoxon-rank sum test for splicing differences).

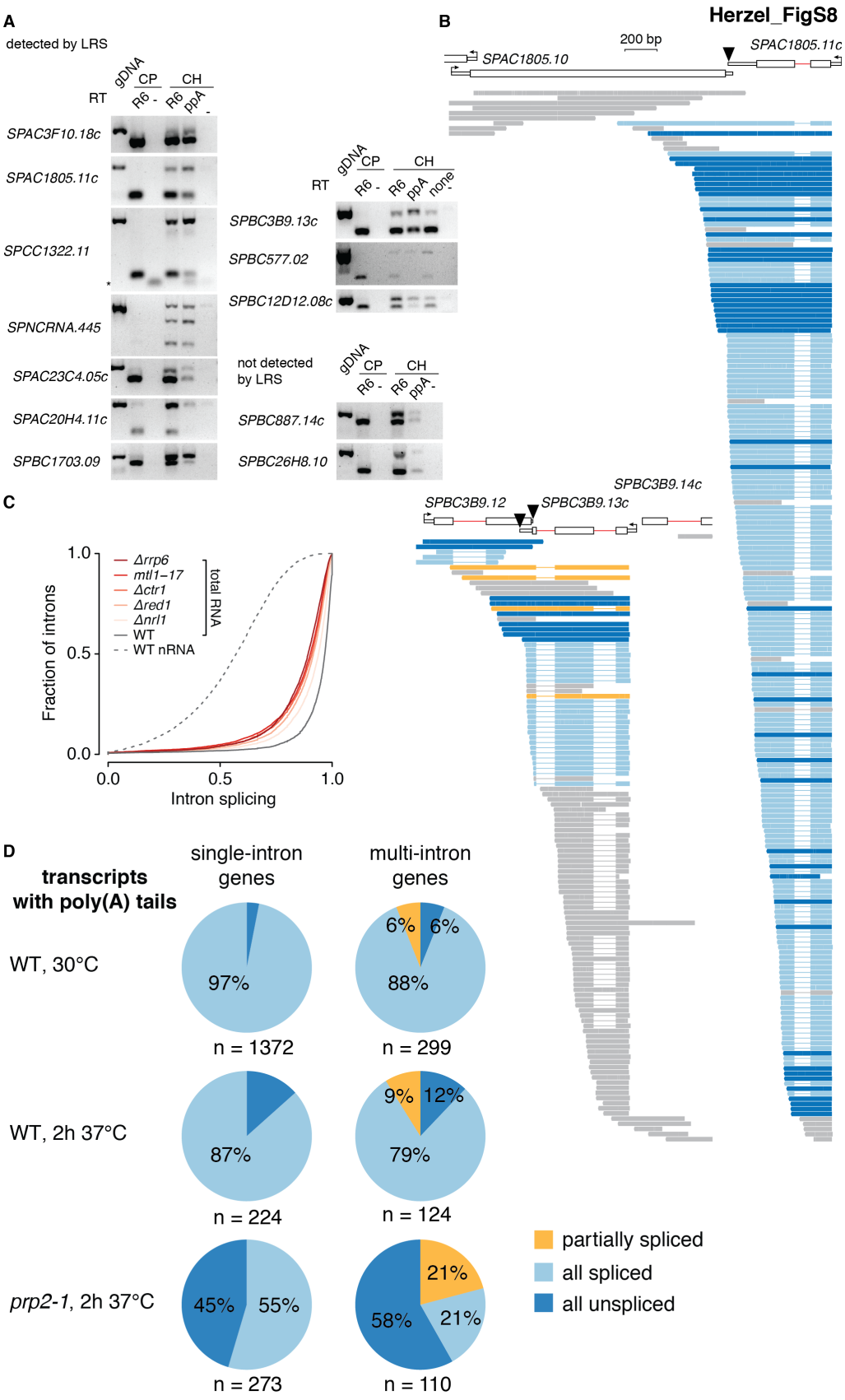

Herzel_FigS8
A
detected by LRS
gDNA
CP
CH
RT
R6 -
R6
ppA
SPAC3F10.18c
SPAC1805.11c
SPCC1322.11
SPNCRNA.445
SPAC23C4.05c
SPAC20H4.11c
SPBC1703.09
none
SPBC3B9.13c
SPBC577.02
SPBC12D12.08c
not detected by LRS
SPBC887.14c
SPBC26H8.10
B
200 bp
SPAC1805.10
SPAC1805.11c
SPBC3B9.12
SPBC3B9.13c
SPBC3B9.14c
C
Δrrp6
mtl1-17
Δctr1
Δred1
Δnrl1
WT
WT nRNA
total RNA
Fraction of introns
Intron splicing
1.0
0.5
0.0
D
transcripts with poly(A) tails
single-intron genes
multi-intron genes
WT, 30°C
97%
6%
6%
88%
n = 1372
n = 299
WT, 2h 37°C
87%
9%
12%
79%
n = 224
n = 124
prp2-1, 2h 37°C
45%
55%
21%
58%
21%
n = 273
n = 110
partially spliced
all spliced
all unspliced

Figure S8: Validation of polyA site cleavage defect in unspliced RNAs by RT-PCR. (A) RT-PCR data validating and extending LRS data. Intron-spanning PCRs for genes identified by LRS having uncleaved unspliced transcripts and genes without sequenced uncleaved unspliced transcripts. A genomic DNA (gDNA) sample and cytoplasmic RNA (CP) were used to define the spliced and unspliced signal. -RT controls were included for cytoplasmic and chromatin-associated RNA samples (-). cDNA was either generated in a random hexamer primed RT reaction (R6, all RNAs reverse transcribed) or using a post-polyA RT primer (ppA, only uncleaved transcripts reverse transcribed). 'none' reflects a PCR sample, where no RT primer but enzyme was added (residual cDNA synthesis due to very short gDNA fragments, signal similar to R6). Higher unspliced PCR signal is detected for 9/10 genes with respective LRS data. 2 genes were assayed, which did not show unspliced, uncleaved transcript in LRS. * unspecific band. (B) Same LRS examples of intron-containing genes with unspliced transcripts extending over the polyA site as in Fig. 7A. Black triangles mark the polyA cleavage sites. (C) Pre-mRNA splicing levels are globally reduced in MTREC and exosome mutants/ deletions. Cumulative distribution of intron splicing levels (reanalyzed data from total RNA-seq data (Zhou et al. 2015)). (D) Pie charts reflecting the fraction of spliced and unspliced polyadenylated transcripts derived from single or multi-intron genes. These transcripts were excluded from other analyses as they could reflect potential mRNA contaminants. However, they also provide information on the splicing status of cleaved and polyadenylated nuclear transcripts and show almost complete pre-mRNA splicing, except for the *prp2-1* strain. They were identified by filtering reads ending within +/- 100 nt of an annotated polyA site and short poly(A) tails (> 4 nt).

2. Supplemental methods

S. pombe strains

For strain information please refer to Table S1.

Genome version and annotation

For all experiments described here, the *S. pombe* genome version EF2 was used (http://goo.gl/PuaBni). For accurate representation of protein-coding gene and ncRNA (except snoRNAs, annotation ASM294v2.31) start and ends we implemented 5′ and 3′ end information from transcription units defined previously (Eser et al. 2016) (Table S5).

Growth conditions and harvest

S. pombe cells were handled according to the Fission yeast handbook (http://research.stowers.org/baumannlab/documents/Nurselab_fissionyeasthandbook.pdf). *S. pombe* cell cultures were grown in complete liquid media (YES - 5 g Bacto Yeast extract, 225 mg Adenine, 225 mg Uracil, 225 mg L-Histidine, 225 mg L-Leucine, 225 mg L-Lysine, 3% D-Glucose in 1 L) at 30°C and 250 rpm or on YES agar plates at 30°C. The *Δrrp6* strain was grown under G418 selection on YES agar plates. 3-4 liquid cultures were inoculated from individual single colonies per experiment and strain. For cell fractionation experiments cells were grown in 50 ml overnight to high densities and then diluted to an OD (595nm) of 0.2 in 1l. Cells were harvested by centrifugation in exponential growth at an OD (595nm) around 0.5. The cell pellets were washed with ice-cold PBS, quick-frozen in liquid nitrogen in 6 aliquots and kept at -80°C.

For *prp2-1* splicing analysis YES cultures originating from single *prp2-1 S. pombe* and 972h- S. pombe cell colonies were grown at 26°C to an OD (595nm) of 0.4, then shifted to 37°C for 2 hours and harvested subsequently.

For total RNA extraction dense 5-10 mL cell cultures were diluted to OD (595nm) of 0.2 in 50 mL and harvested at an OD (595nm) of about 0.5.

RNA purification

Nascent RNA was prepared as described (Carrillo Oesterreich et al. 2016) and used for nascent RNA sequencing and long-read sequencing. All steps were done at 4°C and on ice. Yeast cell pellets harvested by centrifugation were used for this preparation and thawed on ice at 4°C. Each cell aliquot was resuspended in 1 ml B1 buffer (HEPES, pH 8.0 20mM, KCl 60mM, NaCl 15mM, MgCl2 5mM, CaCl2 1mM, Triton X-100 0.8%, Sucrose 0.25M, Spermidine 2.5mM, Spermine 0.5mM, DTT 1mM, PMSF 0.2mM) and the cell suspension was transferred into a fresh 2 mL eppendorf tube containing 1 mL Zirconia beads. Cells were vortexed in 5x 1 min pulses at maximum speed. In between, the cell-bead-suspension was kept on ice for 1 min. One 15 ml Falcon tube per sample was punctured and placed into a 50 ml Falcon tube carrying the tube lit with a circle cut with the approximate diameter of the smaller tube. The 15 ml Falcon tube was kept in place by wrapping Parafilm around at the 12 ml mark. The cell-bead-suspension was transferred to the punctured 15 ml Falcon tube and the 2 ml eppendorf tube was washed 3x with B1 buffer and the buffer was transferred to the cell suspension. Tubes were spun at 400 g for 5 min at 4°C. Without touching the cell pellet, 4x750 µl were transferred into 2 fresh 1.5 ml eppendorf tubes per sample and spun again at 400 g for 5 min at 4°C to eliminate unlysed cells. Cell lysate was transferred to fresh eppendorf tubes once more and then spun at 2,000 g for 15 min at 4°C. The supernatant was kept as cytoplasmic fraction. The RNA was immediately extracted with Phenol:Chloroform:IAA, pH 6.6 and the addition of 1% SDS. For protein analysis the fraction was snap frozen in liquid nitrogen.

The nuclear pellet was resuspended in 800 µl B1 buffer and the centrifugation was repeated. The nuclear pellet was resuspended in B2 buffer (HEPES, pH 7.6 20mM, NaCl 450mM, MgCl2 7.5mM, EDTA 20mM, Glycerol 10%, NP-40 1%, Urea 2M, Sucrose 0.5M, DTT 1mM, PMSF 0.2mM), vortexed for 5 sec and the sample was once more centrifuged at 20,000 g for 15 min and 4°C. After one 800 µl B2 buffer wash (partial resuspension, spin) the brownish dense pellet was resuspended in 250-350 µl buffer P (Sodium acetate 50mM, NaCl 50mM, SDS 1%). One volume of Phenol:Chloroform:IAA, pH 6.6 was added and the sample was incubated at 37°C and 1,150 rpm for 30 minutes. To separate aqueous and organic phase from each other samples were spun at room temperature at maximum speed. ~80% of the aqueous phase were transferred to a 1.5 ml eppendorf tube and precipitated with ethanol. 1/10th

of the sample volume of 3M Sodiumacetate, pH 5.3 and 0.5 µl Glycoblue (ThermoScientific), were added. At least 2.5 volumes of ice-cold 100% Ethanol were added to precipitate the DNA and RNA in the sample. Samples were placed at -80°C for at least 30 min. Precipitates were collected by centrifugation at 20,000 g for at least 30 min and 4°C. The supernatant was discarded and the pellet was washed once with ~300 µl 80% ice-cold Ethanol. Centrifugation at 20,000 g for at least 5 min and 4°C followed. All supernatant was discarded by pipetting and the pellet was dried for ~6 min at 37°C. Samples were resuspended for further use in DEPC-water. All RNA samples were treated twice with 20U Turbo DNase for 30 minutes (Life technologies) and purified with the RNA Clean & Concentrator-5 kit (Zymoresearch).

Protein analysis

Western blots of different samples taken during *S. pombe* cell fractionation were performed to assay enrichment of proteins characteristic for chromatin. The Bradford Protein Assay (Bio-Rad, Assay protocol5) was used to determine protein concentrations. BSA standard curves were prepared in triplicates. Samples were adjusted to neutral pH for protein measurement. ~3 µg of protein were loaded per lane in western blot analysis. Antibodies against two nuclear proteins, the largest Pol II subunit Rpb1 (8WG16), Histone H3 (ab1791, Abcam), and two cytoplasmic proteins, GAPDH (Novus Biologicals, NB300-221) and ribosomal protein L5 (Santa Cruz, sc-103865), were used. Coomassie gel staining and mass spectrometry were performed as described elsewhere (Shevchenko et al. 2006) at the mass spectrometry facility of the MPI-CBG Dresden. Two independent chromatin preparations and one cytoplasmic fractionation were assayed. Proteins were classified as specific to the chromatin fraction, if the unique peptide count for both chromatin samples was twofold or higher relative to the chromatin preparation (191 proteins). 437 proteins were found to be enriched in at least one of two chromatin samples compared to cytoplasm (2nd sample also enriched or equal to cytoplasm). Analogous to that proteins were classified as cytoplasmic, if the unique peptide count in the cytoplasmic sample was twofold or higher compared to both chromatin preparations (383 proteins). A total of 608 proteins were enriched in the cytoplasm sample relative to at least one chromatin sample. 108 proteins could not be assigned to any cellular fraction in particular. Many splicing factors were detected in the chromatin

fraction, likely reflecting the greater prevalence of intron-containing genes and their co-transcriptional splicing compared to *S. cerevisiae* (Table 1), where splicing factors were not detected on chromatin (Carrillo Oesterreich et al. 2010). The full overview of unique peptide counts per gene and sample and classification can be found in Table S3. Gene Ontology (GO) analysis was performed using the R package topGO version 2.18.0 (Alexa et al. 2006). Enrichment of GO-terms was tested with the "weight" algorithm. The p-value was determined with Fisher's exact test. GO term annotations (8/20/2013) were retrieved from PomBase. The fraction of genes/proteins with the associated GO term within the data sets and of the total number of genes/proteins with this GO term was determined. The five most enriched 'cellular component' terms are depicted in the Figure S1D. 421/437 chromatin-associated and 603/608 cytoplasmic proteins were considered in this analysis.

Removal of poly(A)$^+$ RNA and ribosomal RNA

As polyadenylated transcripts can be retained on chromatin (Bhatt et al. 2012) and/or contaminate the chromatin fraction, the nascent RNA signal would possibly be masked in sequencing experiments without poly(A)$^+$ RNA removal. PolyA- RNA was obtained using oligo-dT coated magnetic beads binding to poly(A)$^+$ RNA (Dynabeads mRNA DIRECT Micro Purification Kit, Life technologies). rRNA was removed from ~5 µg chromatin-associated poly(A)$^-$ RNA using the Ribo-Zero Gold rRNA Removal Kit (Yeast) from Epicentre/Illumina.

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 NanoDrop 2000 (ThermoScientific) or fluorometric measurements with the Qubit dsDNA BR Assay or the RNA BR Assay (Life technologies). Single-end short-read and long-read sequencing was done after Bioanalyzer, Qubit dsDNA BR assay and Kappa library quantification.

RT-(q)PCR

Reverse transcription of RNA was done using SuperScript III reverse transcriptase (Invitrogen). The

protocol recommended for the enzyme was used in either 10 or 20 µl reactions. 60 ng of RNA were used as input. Reverse transcription primers were used in the following final concentrations: Random hexamers (Roche, 1 µl /20 µl RT reaction), 0.1 µM gene-specific primers, e.g. post polyA site primer, and 5 µM oligo-dT primer. In some experiments a 'no RT primer' control was also included to assess random priming from residual short gDNA oligonucleotides. RNA, dNTPs and RT primer were denatured at 65°C for 5 min and immediately cooled on ice for at least 1 min. Afterwards RT buffer, 0.1M DTT and RNaseOUT were added in the respective amounts. Superscript III enzyme was added last (or water for -RT controls). Samples with random hexamer priming were incubated at room temperature for 5 min. RT samples were incubated at 55°C for 30 min and the enzyme was inactivated with a 70°C incubation for 15 min. cDNA samples were diluted 1:10 for further use.

For subsequent qPCRs, Sybrgreen (Life technologies) qPCR reaction mix was used. In each well of a 96-well plate 5 µl Sybrgreen qPCR reaction mix, 3 µl primer solution and 2 µl 1:10 diluted cDNA solution were added. Each sample was assayed in 2-3 technical triplicates and at least two no template controls were performed. Each run was done as recommended by the manufacturer. Optimal primer concentrations were determined in a test run for 3 primer concentrations (250 nM (final 83 nM), 500 nM (final 167 nM), 1 µM (final 333 nM)) and 4 different cDNA dilutions for 1:10, 1:100, 1:500 and 1:1000. Primer efficiencies in a range of 95-105% were accepted for further use and the determined cDNA levels were adjusted accordingly.

For conventional PCRs, except for long-read sequencing samples, Phusion DNA polymerase (NEB) was used with recommended reaction settings and reagent concentrations. 25 cycle PCR reactions were carried out in 10 or 20 µl reaction volumes with 1-3 µl 1:10 diluted cDNA template. 1 µM PCR primer solutions were used (final 100 nM). For the validation of the splicing quantification from nRNA-seq, 41 introns were randomly selected. Primers were designed within the flanking exons of an intron of interest using default settings in Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/). This resulted in 39 possible primer pairs and RT-PCR evaluation by agarose gel electrophoresis showed a signal and a distinct unspliced/spliced band and nothing else for 33 introns. Oligonucleotides and PCR results are given in Table S2 and Fig. S2B. A strong positive correlation to our nascent RNA-seq data was observed (R=0.7, Fig. S2B, Table S7), despite the smaller dynamic range of the quantification by RT-

PCR.

Short-read RNA sequencing

For RNA-seq of different cellular fractions RNA samples were submitted to the Yale Center for Genome Analysis (YCGA). Poly(A)$^+$ RNA depleted, rRNA depleted nascent RNA and cytoplasmic poly(A)$^+$ RNA were analyzed by RNA-seq. Random hexamer primed libraries were prepared with standard Illumina library protocols. Single-end sequencing with 76 bp read length was done. Samples were sequenced in biological triplicates.

Mapping short-read RNA-seq data to the *S. pombe* genome

Fastq files were filtered for quality with the FASTQC toolkit (FASTX Toolkit version 0.0.13) and mapped to the genome with TopHat2 (version 2.0.12) (Kim et al. 2013) using the following settings: fastq_quality_filter -Q 33 -q 20 -p 90; tophat2 -p 5 -i 30 -I 900 –g 1 -N 2 -G <Spombe_EF2> –segment-length 25 –library-type fr-firststrand –minanchor-length 8 –splice-mismatches 0 –min-coverage-intron 30 –max-coverageintron 900 –min-segment-intron 30 –max-segment-intron 900. For indexing, file conversion and analysis samtools version 1.1 (Li et al., 2009) and BEDtools version 2.20.1 (Quinlan and Hall 2010) were used. Sequencing data were visualized using the IGV genome browser (Robinson et al. 2011; Thorvaldsdottir et al. 2013). Coverage data (bedgraph-format) were obtained using the samtools view and depth function to first extract mapped reads per DNA strand and then converting them into coverage data (wig-format). Wig-files were converted with awk into the bedgraph-format. Raw and mapped read counts are given in Table S4.

Normalization of RNA-seq coverage data

RNA-seq coverage was adjusted for library size, excluding reads mapped to rDNA regions. rDNA region coordinates are: III: 1-24713, III: 2439224-2452883. Non-rDNA counts per position were summed and divided by the read length to determine the mapped library size. Library sizes were weighted using the R-package edgeR (calcNormFactors without additional parameters) and multiplied by 10^{-6}. Counts per position were divided by this scaling factor to yield RPM (Reads Per Million). For read

coverage analysis over gene features, such as exons and introns, RPMs were normalized by feature length, resulting in RPKM (Reads Per Kilobase of transcript per Million mapped reads).

Intron splicing quantification

For intron splicing calculation junction reads originating from spliced (split, cigar contains "N" and unspliced (unsplit, cigar contains "M", but no "N") transcripts were extracted from all mapped reads. Overlaps with 8 nucleotide windows around annotated 5' SS and 3' SS junctions were searched for split and unsplit reads separately with BEDtools intersect. An overlap of at least 4 nucleotides on each side of the junction was required. The sum of all split reads per 5'SS (identical with 3'SS split read count) was divided by the sum of the 5'SS split read count and the average of unsplit reads over 5' and 3'SS per intron. This resulted in a splicing score (splicing per intron, SPI) ranging from 0 to 1 with 1 being 100% spliced. A cutoff of at least 10 reads per junction to report an SPI was applied.

Accurate splicing quantification of first introns might be confounded by biases due to lower RNA-seq coverage in short first exons. This arises from an underrepresentation of 5' transcript ends due to size selection of RNA fragments in the range or longer than the read length and sequencing 76 nt from the 3' end of fragmented RNA in single-end sequencing. We accounted for this by including only introns with a distance to the transcription start site greater than the read length. Although absolute sequencing depth was similar between mRNA and nRNA sequencing samples, mRNA data contained a higher fraction of rRNA reads (~50%) and thus reduced sequencing depth for protein coding genes than nRNA samples, which were treated specifically to deplete rRNA (~25% of data). Thus, the number of introns with sufficient read coverage to estimate intron splicing levels differs between nRNA samples with 4,481 and mRNA with 2,181 introns.

Gene expression analysis

To determine gene expression values between replicates and different samples Cufflinks version 2.2.1 and Cuffdiff were used with the following settings:

cufflinks -p 20 -G <Spombe_EF2> -b <Spombe_EF2.fasta>; cuffdiff –frag-bias-correct <Spombe_EF2.fasta> –num-threads 20 –library-type fr-firststrand –library-normmethod geometric

<Spombe_EF2.gtf>

FPKM (Fragments Per Kilobase Of Exon Per Million Fragments Mapped) values were required to be greater than the difference between high and low confidence boundary (Nagaraj et al. 2011) and assigned with a Cufflinks flag "OK". The Cuffdiff results taking replicates into account were used for differential expression analysis and for the correlation to SPIs.

Gene architecture analysis

Feature coordinates, length and distances

Intron, exon and gene sequences and coordinates were extracted from the *S.pombe* EF2 genome sequence and annotation ASM294v2.31 and from (Eser et al. 2016) (Table S5). Table S6 and S9 include all feature data used in this study.

Feature IDs

In genes annotated according to (Eser et al. 2016), the gene ID (id.gene) includes the name assigned by (Eser et al. 2016) followed by the ASM294v2.31 id. The same applies for intron and exon ids. Introns and exons are numbered with respect to genome position, not position within the gene. Thus, intron 1 in a plus strand gene refers to the first intron, and intron 1 in a minus strand gene refers to the last intron. "exon.minus1" refers to the upstream/ 5' exon and "exon.plus1" to the downstream/ 3'exon.

Splice site motif score calculation

The position weight matrix was derived for all hexanucleotides at 5' end of intron (5'SS), trinucleotides at 3' end of introns (3'SS) and the heptanucleotide branchpoint sequences obtained from (Eser et al. 2016) (all sequences given in Table S6 (motif.5SS, motif.3SS, motif.BP). This 4x6, 4x3 or 4x7 matrix contains the fraction of A, C, G or U bases at each position of the motif in all *S. pombe* introns. The LogOdds matrix was calculated using a background nucleotide composition considering average frequency of A, C, G and U bases in *S. pombe* introns. For each individual sequence motif the score (score.5SS, score.3SS, score.BP) was calculated by summing up LogOdds matrix entries at the respective motif position (column) and base identity (row).

Modified Z-score calculation for visualization of feature differences among intron groups

The modified Z-score (Ramachandran and Tsokos 2015) was calculated to transform intron feature data onto a uniform scale for visualization in Figure 2D and S2C, using the following formula:

(1) $z_i = \frac{x_i - \bar{x}}{MAD}$, with x_i being the median of the intron feature in group i, $\bar{x}$ being the intron feature median for all introns and MAD the median of absolute deviation.

(2) $MAD = median(|x_i - m|)$, with m being all intron feature values.

Glossary for Table S6

- id.intron – intron id (numbering according to chromosome position, not considering strand information)
- chr – chromosome
- start.intron – intron start with respect to chromosome position (5'SS on plus strand, 3'SS on minus strand; 0-based)
- end.intron – intron end with respect to chromosome position (3'SS on plus strand, 5'SS on minus strand; 1-based)
- strand – chromosome strand, plus or minus
- length.intron – nucleotide distance from intron start to end
- distance.TSS - nucleotide distance from intron 5'SS to 5' end of the gene
- distance.polyA - nucleotide distance from intron 3'SS to 3' end of the gene
- distance.CDS.start - nucleotide distance from intron 5'SS to the 1st nucleotide of the start codon
- distance.CDS.end - nucleotide distance from intron 3'SS to the last nucleotide of the stop codon
- length.gene - nucleotide distance from the 5' end of the gene to 3' end of the gene (including introns)
- length.ORF - nucleotide distance from the 5' end of the gene to 3' end of the gene (excluding introns)
- length.CDS - nucleotide distance from the 1st nucleotide of the start codon to last nucleotide of the stop codon (excluding introns)
- count.introns – number of introns per gene
- order.introns.fwd – relative position of intron in gene starting from the 5' end (1 – first intron)
- order.introns.rev – relative position of intron in gene starting from the 3' end (1 – last intron)
- id.exon.minus1 - upstream/5'exon id (numbering according to chromosome position, not considering strand information)
- length.exon.minus1 – nucleotide distance from exon start to end
- id.exon.plus1 - downstream/3'exon id (numbering according to chromosome position, not considering strand information)

- length.exon.plus1 – nucleotide distance from exon start to end
- gc_fract.genes – fraction of either G or C nucleotides per gene (including introns) gc_fract.intron – fraction of either G or C nucleotides per intron
- gc_fract.exon.minus1 – fraction of either G or C nucleotides per 5′exon
- gc_fract.exon.plus1 – fraction of either G or C nucleotides per 3′exon
- motif.5SS – first 6 nucleotides of the intron 5′end
- position.end.5SS – distance of 5′SS motif end to intron 5′end (always 6 nucleotides) score.5SS – motif score estimating similarity of the hexanucleotide similarity to 5′SS consensus sequence
- motif.3SS – last 3 nucleotides of the intron
- position.end.3SS – distance of 3′SS motif end to intron 5′end (corresponds to length.intron)
- score.3SS – motif score estimating similarity of the trinucleotide similarity to 3′SS consensus sequence
- coverage.sum.Sp_CH_nRNA_introns – length normalized nascent RNA-seq read coverage over introns
- coverage.sum.Sp_CP_mRNA_introns – length normalized mRNA-seq read coverage over introns
- coverage.sum.Sp_RP_mRNA_introns – length normalized Ribo-seq read coverage over introns (data from (Duncan and Mata 2014))
- coverage.sum.Sp_CH_nRNA_exon.minus1 – length normalized nascent RNA-seq read coverage over 5′exons
- coverage.sum.Sp_CP_mRNA_exon.minus1– length normalized mRNA-seq read coverage over 5′exons
- coverage.sum.Sp_RP_mRNA_exon.minus1 – length normalized Ribo-seq read coverage over 5′exons (data from (Duncan and Mata 2014))
- coverage.sum.Sp_CH_nRNA_exon.plus1 – length normalized nascent RNA-seq read coverage over 3′exons
- coverage.sum.Sp_CP_mRNA_exon.plus1 – length normalized mRNA-seq read coverage over 3′exons

- coverage.sum.Sp_RP_mRNA_exon.plus1 – length normalized Ribo-seq read coverage over 3'exons (data from (Duncan and Mata 2014))
- fpkm.log2.fold_change_mw.mc – gene expression difference quantified with Cuffdiff between cytoplasmic mRNA (water-control, mw) and cytoplasmic mRNA (caffeine treated, mc), $\log_2(fpkm_{mc}/fpkm_{mw})$
- fpkm.p.value_mw.mc – p-value associated with gene expression difference quantified with Cuffdiff between cytoplasmic mRNA (water-control, nw) and cytoplasmic mRNA (caffeine treated, mc)
- fpkm.mc – gene expression value quantified with Cufflinks/Cuffdiff for cytoplasmic mRNA (caffeine treated, mc)
- fpkm.mw – gene expression value quantified with Cufflinks/Cuffdiff for cytoplasmic mRNA (water control, mw)
- fpkm.log2.fold_change_nw.mw – gene expression difference quantified with Cuffdiff between nascent RNA (water control, mw) and cytoplasmic mRNA (water control, mw), $\log_2(fpkm_{mw}/fpkm_{nw})$
- fpkm.p.value_nw.mw – p-value associated with gene expression difference quantified with Cuffdiff between nascent RNA (water control, nw) and cytoplasmic mRNA (water control, mw)
- fpkm.nw – gene expression value quantified with Cufflinks/Cuffdiff for nascent RNA (water control, nw)
- fpkm.nc – gene expression value quantified with Cufflinks/Cuffdiff for nascent RNA (caffeine treated, nc)
- fpkm.log2.fold_change_nw.nc – gene expression difference quantified with Cuffdiff between nascent RNA (water-control, nw) and nascent RNA (caffeine treated, nc), $\log_2(fpkm_{nc}/fpkm_{nw})$
- fpkm.p.value_nw.nc – p-value associated with gene expression difference quantified with Cuffdiff between nascent RNA (water control, nw) and nascent RNA (caffeine treated, nc)
- conservation.sum_intron - intron length adjusted Schizosaccharomyces clade multiple sequence alignment (MSA) score per intron (MSA data from (Rhind et al. 2011))
- conservation.sum_exon.minus1 – 5'exon length adjusted Schizosaccharomyces clade multiple

sequence alignment (MSA) score per 5'exon (MSA data from (Rhind et al. 2011))

- conservation.sum_exon.plus1 – 3'exon length adjusted Schizosaccharomyces clade multiple sequence alignment (MSA) score per 3'exon (MSA data from (Rhind et al. 2011))
- half.life – mRNA half-life estimates from (Eser et al. 2016) (time needed to degrade half of the mature RNAs)
- splicing.time – estimates from (Eser et al. 2016) (time to process half of the precursor RNA junction/ intron)
- motif.BP – generated using FELINES (Drabenstot et al. 2003), branchpoint prediction in (Eser et al. 2016)
- position.BP.3SS – nucleotide distance between last branchpoint nucleotide and last 3'SS nucleotide (intron 3'end)
- score.BP – motif score estimating similarity of the heptanucleotide similarity to branchpoint consensus sequence
- length.PY – comma-separated list of any intronic nucleotide stretch length of 4 or more U or C bases, of which are more than 50% U bases
- position.PY.3SS – comma-separated list of any intronic nucleotide stretch position relative to the 3'SS of 4 or more U or C bases, of which are more than 50% U bases
- position.PY.BP – comma-separated list of any intronic nucleotide stretch position relative to the branchpoint of 4 or more U or C bases, of which are more than 50% U bases (negative values indicate that the PY-tract is upstream of the branchpoint)
- length.allPY – summed length of all intronic tracts with 4 or more U or C bases, of which are more than 50% U bases
- position.BP.3SS.relative – nucleotide distance between last branchpoint nucleotide and last 3'SS nucleotide (intron 3'end) divided by the intron length
- deltaG_intron_norm – free energy prediction of RNA secondary structure with Quikfold (http://unafold.rna.albany.edu/?q=DINAMelt/Quickfold) at 30°C and default parameters; the resulting free energy in kcal/mol is divided by the intron length
- mc1_spliced – splicing per intron (SPI) value cytoplasmic mRNA, caffeine treated, replicate 1

- mc2_spliced – splicing per intron (SPI) value cytoplasmic mRNA, caffeine treated, replicate 2
- mc3_spliced – splicing per intron (SPI) value cytoplasmic mRNA, caffeine treated, replicate 3
- mw1_spliced – splicing per intron (SPI) value cytoplasmic mRNA, water control, replicate 1
- mw2_spliced – splicing per intron (SPI) value cytoplasmic mRNA, water control, replicate 2
- mw3_spliced – splicing per intron (SPI) value cytoplasmic mRNA, water control, replicate 3
- nc1_spliced – splicing per intron (SPI) value nascent RNA, caffeine treated, replicate 1
- nc2_spliced – splicing per intron (SPI) value nascent RNA, caffeine treated, replicate 2
- nc3_spliced – splicing per intron (SPI) value nascent RNA, caffeine treated, replicate 3
- nprp5WT1_spliced – splicing per intron (SPI) value nascent RNA, wild-type strain 2, replicate 1
- nprp5WT2_spliced – splicing per intron (SPI) value nascent RNA, wild-type strain 2, replicate 2
- nw1_spliced – splicing per intron (SPI) value nascent RNA, water control, replicate 1
- nw2_spliced – splicing per intron (SPI) value nascent RNA, water control, replicate 2
- nw3_spliced – splicing per intron (SPI) value nascent RNA, water control, replicate 3
- spl_diff.adjacent_mw – cytoplasmic mRNA SPI difference between upstream (5′) intron and this intron (smaller 0: 5′ intron less well spliced than this intron)
- p.value.adjacent_mw – p-value from Student's t-test associated with mRNA splicing difference between upstream (5′) intron and this intron, 3 replicates each
- spl_diff.adjacent_nprp5WT – SPI difference between upstream (5′) intron and this intron (smaller 0: 5′ intron less well spliced than this intron)
- p.value.adjacent_nprp5WT – p-value from Student's t-test associated with splicing difference between upstream (5′) intron and this intron, 2 replicates each
- spl_diff.adjacent_nw – nascent RNA SPI difference between upstream (5′) intron and this intron (smaller 0: 5′ intron less well spliced than this intron)
- p.value.adjacent_nw – p-value from Student's t-test associated with nascent RNA splicing difference between upstream (5′) intron and this intron, 3 replicates each
- spl_diff.condition_nw_nc – nascent RNA SPI difference between the two conditions, water control and caffeine treated (smaller 0: water control less well spliced)

- p.value.condition_nw_nc – p-value from Student's t-test associated with nascent RNA splicing difference between water control and caffeine treatment
- spl_diff.condition_mw_mc – cytoplasmic mRNA SPI difference between the two conditions, water control and caffeine treated (smaller 0: water control less well spliced)
- p.value.condition_mw_mc – p-value from Student's t-test associated with mRNA splicing difference between water control and caffeine treatment
- spl_diff.condition_nw_nprp5WT – nascent RNA SPI difference between the two wild-type strains (smaller 0: water control/ WT1 less well spliced)
- p.value.condition_nw_nprp5WT – p-value from Student's t-test associated with nascent RNA splicing difference between the two wild-type strains
- spl_mean_mc – average cytoplasmic mRNA SPI, caffeine treated
- spl_mean_mw – average cytoplasmic mRNA SPI, water control
- spl_mean_nc – average nascent RNA SPI, caffeine treated
- spl_mean_nprp5WT – average nascent RNA SPI, wild-type strain 2
- spl_mean_nw – average nascent RNA SPI, water control/ wild-type strain 1
- spl.gene_mean_mc – average SPI for all introns of a gene in cytoplasmic mRNA (caffeine treated)
- spl.gene_mean_mw – average SPI for all introns of a gene in cytoplasmic mRNA (water control)
- spl.gene_mean_nc – average SPI for all introns of a gene in nascent RNA (caffeine treated)
- spl.gene_mean_nprp5WT – average SPI for all introns of a gene in nascent RNA (wild-type 2)
- spl.gene_mean_nw – average SPI for all introns of a gene in nascent RNA (water control)
- spl.gene_sd_mc – standard deviation between SPIs of all introns in a gene, cytoplasmic mRNA (caffeine treated)
- spl.gene_sd_mw – standard deviation between SPIs of all introns in a gene, cytoplasmic mRNA (water control)
- spl.gene_sd_nc – standard deviation between SPIs of all introns in a gene, nascent RNA (caffeine treated)

- spl.gene_sd_nprp5WT – standard deviation between SPIs of all introns in a gene, nascent RNA (wild-type 2)
- spl.gene_sd_nw – standard deviation between SPIs of all introns in a gene, nascent RNA (water control)
- spl.gene_median_mc – median SPI for all introns of a gene in cytoplasmic mRNA (caffeine treated)
- spl.gene_median_mw – median SPI for all introns of a gene in cytoplasmic mRNA (water control)
- spl.gene_median_nc – median SPI for all introns of a gene in nascent RNA (caffeine treated)
- spl.gene_median_nprp5WT – median SPI for all introns of a gene in nascent RNA (wild-type 2)
- spl.gene_median_nw – median SPI for all introns of a gene in nascent RNA (water control)
- dsGene3end.category – closest gene end relative to 3′ end of this gene on same or different strand and what end it is (5′ or 3′)
- distance.dsGene3end – distance to the closest end relative to the 3′ end

3′ end ligation and long-read sequencing library preparation

3′ end ligation and long-read sequencing library preparation of poly(A)$^+$ and rRNA-depleted, chromatin-associated RNA was done as described previously (Carrillo Oesterreich et al. 2016). Column purification with the RNA Clean & Concentrator-5 kit (Zymoresearch) and ethanol precipitation (see above) followed rRNA removal. All rRNA-depleted RNA and 50 pmol 5N-barcoded 3′end linker (/5rApp/NNNNNCTGTAGGCACCATCAAT/3ddC/, Integrated DNA Technologies) were combined to a final volume of 6 µl. After denaturation (65°C 5 min, >1 min on ice), the remaining components for 3′end ligation were added (final 50 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, pH 7.5, 25% PEG 8000, 40 U RNaseOUT, 200 U T4 RNA ligase II [truncated K227Q, NEB]), and samples were incubated for 10 hours at 16°C. RNA column purification (RNA Clean & Concentrator-5 kit, Zymoresearch) was performed to remove unligated adaptor and enzyme. A 'minus Ligase' control was included to ensure that RT and PCR are specific to 3′end ligated RNAs.

Nascent RNA was reversed transcribed (SMARTer PCR cDNA Synthesis Kit, Clontech). The included

3'SMART CDS Primer II A was substituted with a custom primer (Table S2). RNA (~1 μg) was used per reaction. Double-stranded DNA was generated by a low cycle PCR (Advantage 2 PCR Kit, Clontech, around 13 cycles) according to manual instructions. Sample-specific barcodes were included in the PCR oligonucleotides (Table S2) for the 3 *prp2-1* cDNA libraries and the corresponding 972h- cDNA libraries. These 6 cDNA libraries were pooled at equal amounts after Qubit quantification and prior submission for Pacific Biosciences library preparation. Double-stranded cDNA (>1 μg) was submitted to the Yale Center for Genomic Analysis (YCGA) for Pacific Biosciences library preparation and sequencing with standard protocols (SMRTbell Template Prep Kit 1.0). Sequencing was done with either diffusion- or magbead-loading. The latter allows preferential sequencing of transcripts longer than 500 nt, which helps quantifying splicing profiles of multi-intron transcripts for longer genes, but does not report on transcripts <500 nt. Diffusion-loading includes those, but slightly over represents spliced transcripts due to their shorter length than respective unspliced transcripts. To account for this, both strategies were applied: 6 diffusion-loaded SMRT cells were sequenced for WT1, 1 diffusion-loaded and 1 magbead-loaded SMRT cell for WT2 and 7 magbead-loaded SMRT cells for all *prp2-1* 37°C and WT 37°C samples.

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 respective genome using GMAP (Wu and Watanabe 2005). To remove potential mRNA contaminants reads ending within +/-100 nt of an annotated polyA site and short poly(A) tails (> 4 nt) were removed from the dataset. Raw and mapped read counts are given in Table S4.

Classification of multi-intron transcripts

Mapped data (bam-format) were converted into 12 column bed-files using the bamToBed function available in samtools. Position overlaps with annotated *S. pombe* introns were identified with BEDtools

intersect and a custom intron annotation file generated using the gene annotation described above. Reads overlapping completely with multiple introns were considered multi-intron transcripts. If the block count (number of exons) of a read was identical to the number of overlapping introns +1, a read was considered to be "all spliced". If the block count was 1 irrespective of the number of overlapping introns, a read was considered to be "all unspliced". Reads, which had a block count greater 1, but fewer block counts than the number of overlapping introns +1, were classified as "partially spliced". Those can also contain rare cases of exon skipping events. Further, rare cases of alternative 5'SS and 3'SS usage are not included in this classification and could be present for reads classified as "all spliced" or "partially spliced". We compared partially spliced transcripts identified by LRS to our nRNA-seq data and found a highly significant overlap between datasets (p-value=$4*10^{-08}$, Fisher's exact test). 5' introns classified as 'unspliced' or 'spliced' in partially spliced transcripts distributed according to the difference in nRNA-seq splicing values between adjacent intron pairs and thus recapitulated the differences in SPI by nRNA-seq (Fig. S5A).

Splicing onset calculation for single intron genes and first introns in multi-intron transcripts

The splicing onset position was calculated as described in (Carrillo Oesterreich et al. 2016). In brief, Pol II positions (3' ends of LRS reads) for all spliced and unspliced nascent RNA molecules were determined relative to 3'SSs. The total number of observations is low and results in a sparse coverage of single Pol II positions for each gene (Fig. S4G). To determine the average onset of splicing, we consider Pol II positions of spliced reads over all genes and the corresponding cumulative frequency distribution (Fig. S4H). A linear phase is valid for all positions <300 nt, which includes the minimal exon length requirement (259 nt, 10% quantile of internal and terminal exon lengths) (Kolmogorov-Smirnov test with respect to uniform distribution in the window 0-259 nt, p-value > 0.05). To infer the onset of co-transcriptional we fit a linear model to the observed cumulative distribution within the window of minimal exon length. The x-intercept of the linear model corresponds to the average onset of splicing ± 95% confidence interval. The confidence interval on the intercept was derived by bootstrapping.

Prediction of transcript splicing status from nRNA-seq

To determine the expectation of the abundance of partially spliced transcripts, in the case that individual intron splicing is independent of each other, we assume that the SPI, calculated from short read RNA-seq, reflects the probability of splicing of a particular intron in the chromatin fraction at steady state. Further, we assume that the gene expression values from nRNA-seq reflect an estimate of relative transcript number per gene in the chromatin fraction at steady state. However, for individual genes it is expected that transcripts vary in length and number of introns due to their ongoing synthesis. Hence, we only include first and second introns in this analysis to estimate the three populations of transcripts: (1) all spliced, (2) all unspliced and (3) partially spliced. Further, this restriction matches the LRS data characteristics, as the LRS data report mainly on first and second intron splicing status. For this global analysis only genes are considered, which have sufficient read counts to calculate a SPI for each intron. For each intron 'spliced' or 'unspliced' sampling with replacement was performed according to its SPI. Sampling was performed as many times as there were molecules. Sampling results were combined into a matrix with columns representing each intron and rows representing each transcript. If a row contained both 'spliced' and 'unspliced' column entries, it was classified as 'partially spliced'. If a row contained only 'spliced' or only 'unspliced' column entries, it was classified as 'all spliced' or 'all unspliced', respectively. The absolute counts and transcript class were combined for all genes and represented in a pie chart illustration as shown in Figure 4A, S6C.

In a second analysis, we considered genes with 10 or more multi-intron LRS reads (100 in pooled dataset of WT1 and WT2) and the proportion of "all or none" transcripts (all introns spliced out or all introns still present) was determined. This corresponds to the observed fraction of "all or none" splicing. The predicted fraction of "all or none" splicing was calculated from nRNA-seq SPI data. Again we assume that the SPI provides an estimate for the probability of splicing of a particular intron. In cases, where no SPI could be calculated, the average gene splicing values (mean SPI of all other introns) was used. The number of introns per gene was limited to the number of introns covered in the LRS data. The proportion of "all or none" transcripts was calculated as the sum of the proportion of "all spliced" transcripts and "all unspliced" transcripts, which themselves are the product of the different intron splicing probabilities, SPIs, or the product of 1 –the splicing probability, SPI, of the considered

introns, respectively:

$p(all\ or\ none) = \prod_1^n SPI_n + \prod_1^n (1 - SPI_n)$, n - intron index, p – proportion of "all or none" transcripts

From the observed and predicted fraction of "all or none" splicing we calculated a co-association score, which we define as the $\log_2$-ratio between observed and predicted fraction of "all or none" splicing. It provides an estimate of the deviation from independence of splicing (completely independent splicing of adjacent introns reflects in a co-association score of 0). To evaluate, if splicing independence can be recovered using this metric, we randomly paired LRS data from single intron genes with the same number of LRS reads (>9 intron-spanning reads per gene required) to form mock two-intron genes. Our dataset contains 65 such pairs. For 55 pairs SPIs from nRNA-seq were available and could be used to calculate the predicted fraction of "all or none" splicing. As the paired single intron genes reside in different locations in the genome, their splicing should not be co-associated as for multi-intron genes. As expected for independence of splicing between those randomly paired introns, the co-association score is centered around 0, providing an estimate for background variation of the co-association score (Fig. S5E). The variation around 0 reflects experimental noise in the splicing quantification by LRS and the nRNA SPI. The co-association score is a measure for the difference between observation and prediction, assuming independence of splicing, and increases with the number of introns per gene, as the difference to the expectation is more pronounced. For example, the predicted fraction of "all or none" splicing for a 2-intron gene with a co-transcriptional splicing value of $spi_1=spi_2=0.6$ is $0.6^2+(1-0.6)^2=0.52$, which is greater than for a 3-intron gene with the same splicing values ($0.6^3+(1-0.6)^3=0.28$). Thus the co-association scores correlate with the number of introns per gene (Fig. S5F). The co-association itself does not depend on the number of introns per gene, which we determined by correlating the fraction of maximal co-association with the number of introns per gene (Fig. S5G). The fraction of maximal co-association represents the ratio of the co-association score for a given gene to the maximal possible co-association value at the average gene splicing value (see Fig. 4D for 2,3, and 4-intron gene values). The median for the 100 analyzed genes is 0.86.

Microarray analysis of global gene expression pattern in *prp2-1*

To analyze global gene expression changes in the splicing mutant *prp2-1* published microarray data

(GSM1532958, GSM1532959, GSM1532960) from (Lipp et al. 2015) were used. The microarray study was carried out with total RNA from mid-log phase grown cells in rich YES5 medium. The $\log_2$-ratio between perturbation (*prp2-1* shifted to 37°C, Cy5 signal at 635nm on array) and control (Cy3 signal at 532nm on array) for exon probes was used for comparison of intron-containing (2,549) and intronless (2,743) genes. To assess significance of differences in gene expression the Kolmogorov-Smirnov test was performed and the Bonferroni-corrected p-value is reported in Figure 5E. Changes in gene expression of intronless genes between *prp2-1* and WT samples may be indirect effects of altered expression of the large number of intron-containing genes (Fig. 5E). The analogous comparison between intron-containing and intronless genes for two wild-type control samples (GSM1532924, GSM1532925) is presented in Figure S6D.

Mapping and analysis of mRNA LRS data

mRNA long-read sequencing data (meiosis hour 0; GSM2103021 (Batch 1_time point 1) and GSM2103027 (Batch 2_time point 1)) from (Kuang et al. 2017) (GSE79802) were mapped using GMAP (Wu and Watanabe 2005) and the following parameters:

--min-intronlength=20 --max-intronlength-middle=1000 --max-intronlength-ends=1000 --trim-end-exons=0 --localsplicedist=1000 --no-chimeras --allow-close-indels=2 --microexon-spliceprob=0.5 --npaths=1 --nofails --failed-input=no_mapping --mapboth -A --format=samse

The same classification of multi-intron transcripts into “all spliced”, “all unspliced” and “partially spliced” groups as for the nRNA LRS data was performed and used for comparison to the nRNA data in Fig. 3 and S5.

Data processing, mapping and analysis

Data from the following studies were integrated into the analysis described above:

Fission yeast genome sequence conservation data (Rhind et al. 2011)

Ribo-seq data from vegetatively growing fission yeast cells

(E-MTAB-2176) (Duncan and Mata 2014)

Exosome deletion total RNA-seq data (Zhou et al. 2015)

Long-read sequencing data of mRNA (Kuang et al. 2017)

RNA metabolism kinetics & annotation (Eser et al. 2016)

Caffeine mRNA expression data Rallis (Rallis and Bahler 2013)

prp2-1 microarray data (Lipp et al. 2015)

3. List of supplemental tables

Supplemental Table 1: *S. pombe* strains used in this study.
Supplemental Table 2: Oligonucleotides used in this study.
Supplemental Table 3: Mass spectrometry of *S. pombe* chromatin.
Supplemental Table 4: RNA-seq and LRS read counts.
Supplemental Table 5: *S. pombe* gene annotation used in this study.
Supplemental Table 6: nRNA-seq splicing quantification and intron features.
Supplemental Table 7: Validation of nRNA-seq intron splicing values by RT-PCR.
Supplemental Table 8: List of genes with evidence for unspliced uncleaved transcripts.
Supplemental Table 9: List of the 42 intron features.

4. Supplemental references

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**: 1600-1607.

Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**: 279-290.

Carrillo Oesterreich F, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. 2016. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* **165**: 372-381.

Carrillo Oesterreich F, Preibisch S, Neugebauer KM. 2010. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40**: 571-581.

Drabenstot SD, Kupfer DM, White JD, Dyer DW, Roe BA, Buchanan KL, Murphy JW. 2003. FELINES: a utility for extracting and examining EST-defined introns and exons. *Nucleic Acids Res* **31**: e141.

Duncan CD, Mata J. 2014. The translational landscape of fission-yeast meiosis and sporulation. *Nat Struct Mol Biol* **21**: 641-647.

Eser P, Wachutka L, Maier KC, Demel C, Boroni M, Iyer S, Cramer P, Gagneur J. 2016. Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. *Mol Syst Biol* **12**: 857.

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

Kuang Z, Boeke JD, Canzar S. 2017. The dynamic landscape of fission yeast meiosis alternative-splice isoforms. *Genome Res* **27**: 145-156.

Lipp JJ, Marvin MC, Shokat KM, Guthrie C. 2015. SR protein kinases promote splicing of nonconsensus introns. *Nat Struct Mol Biol* **22**: 611-617.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**.

Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Paabo S, Mann M. 2011. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**: 548.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Rallis C, Bahler J. 2013. Inhibition of TORC1 signaling and increased lifespan: gained in translation? *Aging (Albany NY)* **5**: 335-336.

Ramachandran KM, Tsokos CP. 2015. Chapter 14 - Some Issues in Statistical Applications: An Overview. In *Mathematical Statistics with Applications in R (Second Edition)*, doi:https://doi.org/10.1016/B978-0-12-417113-8.00014-X, pp. 687-731. Academic Press, Boston.

Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF, Heiman DI et al. 2011. Comparative functional genomics of the fission yeasts. *Science* **332**: 930-936.

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

Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**: 2856-2860.

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859-1875.
Zhou Y, Zhu J, Schermann G, Ohle C, Bendrin K, Sugioka-Sugiyama R, Sugiyama T, Fischer T. 2015. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun* **6**: 7050.