

Supplemental Information Inventory

Supplementary Figures

Figure S1. Purification of Red1 and Identification of Its Associated proteins, Related to Figure 1

Figure S2. Red1- and Mtl1-associated Factors Differentially Affect Heterochromatin Domains, Related to Figure 2

Figure S3. ncRNA Regulates Gene Expression in Response to Environmental Changes, Related to Figure 4

Figure S4. Mtl1 Forms a Red1-Independent Complex with Nrl1 and Ctr1, Related to Figure 5 and 7

Figure S5. Unannotated Introns are Present at HOODs and Pericentromeric Regions, Related to Figure 6

Figure S6. Nrl1 and Cwf10 Regulate Formation of HOODs, Related to Figure 6

Figure S7. Detection of Cryptic Introns in Non-coding RNAs and Other Parts of the Genome. Related to Figure 6

Supplementary Tables

Table S1. Rmn1, Nrl1 and Cwf10 facilitate small RNA production and H3K9me at genes and retrotransposons, Related to Figures 2, 5, and 6

Table S2. List of Genes and ncRNA Upregulated in *red1* Δ , *mtl1-1*, and *rrp6* Δ Mutants (log₂ ratio), Related to Figure 3

Table S3. List of Previously Unannotated Introns Identified in Wild-type and Different Mutants, Related to Figures 6 and S7

Table S4. List of Strains Used in This Study, Related to Experimental Procedures

Supplementary Methods

Extended Experimental Procedures, Related to Experimental Procedures

Supplementary References

Supplementary Information

Supplementary Figure Legends

Figure S1. Purification of Red1 and Identification of Its Associated proteins, Related to Figure 1 (A) Proteins associated with Red1, Mtl1, Pir1, and Rmn1 that were identified from the indicated purifications. (B) Predicted protein domains of Mtl1, Ctr1 and their human orthologs. (C) Co-IP analysis of the interaction between Red1 and Cwf10. (D) Immunofluorescence analysis of Red1 and Mtl1 localization. (E) Venn diagram depicting the overlap in Mtl1 and Red1 binding sites across the genome, as determined by ChIP-chip analysis.

Figure S2. Red1- and Mtl1-associated Factors Differentially Affect Heterochromatin Domains, Related to Figure 2 (A) ChIP analysis of Red1-MYC, Mtl1-MYC, and Pir1-MYC enrichment at the *mei4* island. Numbers shown below ChIP lanes represent fold enrichments. (B) ChIP analysis of H3K9me2 enrichment at *mei4* in indicated strains. (C) Average distribution of Red1 and Mtl1 across Red1-dependent and –independent islands, as determined by ChIP-chip analysis. Also shown is average H3K9me2 enrichment in wild-type and indicated mutants. (D) Table summarizing results of ChIP-chip analysis for Red1 and Mtl1 and Pir1 localizations in wild-type cells, and H3K9me2 enrichment in wild-type and mutant strains at Red1-dependent and –independent islands. (E) H3K9me2 enrichment at Red1-independent islands and a pericentromeric region in indicated strains. (F) Pir1-MYC distribution at heterochromatin islands (left). RT-PCR analysis of *ssm4* and *mei4* transcript levels in wild-type and indicated mutants (right). The *act1* locus was used as a control. +RT and –RT indicate the presence or absence of reverse transcriptase. (G) Quantitative analysis of small RNAs at HOODs in *rrp6Δ* and *rrp6Δ rmn1Δ* strains. Staples indicate 5 percentile (lower) and 95 percentile (upper). The lower and upper bounds of the boxes denote 25 and 75 percentile, respectively.

Figure S3. ncRNA Regulates Gene Expression in Response to Environmental Changes, Related to Figure 4 (A) ChIP-chip analysis of Red1-MYC, Mtl1-MYC, and Rrp6-MYC enrichment at an unannotated ncRNA upstream of the *SPBC1271.09* gene (top). RNA-Seq analysis is shown for wild type, *red1Δ*, *mtl1-1*, and *rrp6Δ* (bottom). (B) ChIP-chip analysis of Rrp6 enrichment

upstream of the *pho1* locus (top). RNA-Seq analysis is shown for wild-type, *red1Δ*, *mtl1-1*, and *rrp6Δ* (bottom). (C) ChIP-chip analysis of Red1-MYC, Mtl1-MYC, and Rrp6-MYC enrichment at an unannotated ncRNA upstream of the *SPCC11E10.01* gene (top). RNA-Seq analysis is shown for wild-type, *red1Δ*, *mtl1-1*, and *rrp6Δ* (bottom). (D) Additional examples of genes that contain upstream ncRNA with Red1 and Mtl1 enrichment.

Figure S4. Mtl1 Forms a Red1-Independent Complex with Nrl1 and Ctr1, Related to Figure 5 and 7 (A) Proteins associated with Nrl1 and Ctr1 that were identified from the indicated purifications. Components of the Sm protein complex identified in the Ctr1 purification are underlined. (B) Normalized small RNA reads in *rrp6Δ* and *rrp6Δ nrl1Δ* plotted in alignment with the loci map of HOOD 3 and HOOD 4 (top). The signals above and below the line represent small RNAs that map to the top and bottom DNA strands, respectively. ChIP-chip analysis of H3K9me2 in *rrp6Δ* and *rrp6Δ nrl1Δ* (bottom). (C) ChIP-chip analysis of H3K9me2 enrichment at heterochromatin islands in wild-type, *nrl1Δ* and *ctr1Δ*.

Figure S5. Unannotated Introns are Present at HOODs and Pericentromeric Regions, Related to Figure 6 (A) Schematic of RNA-Seq reads that map to *Tf2* and *SPCC1442.04c* in *rrp6Δ*, which contain cryptic introns. Genomic location of the introns and sequences at intron junctions are shown below. 3' and 5' splice sites are indicated with bold blue lettering. (B) Schematic of cryptic introns detected by RNA-Seq in the pericentromeric region of *ago1Δ* (top). The arcs above (red) and below (blue) the line represent intron junction reads that map to the top and bottom DNA strands, respectively. The thickness of the arc corresponds to the number of reads. Small RNA reads at the pericentromeric region in wild-type are plotted (middle). The signals above and below the line represent small RNAs that map to the top and bottom DNA strands, respectively. Schematic of RNA-Seq reads detected at the pericentromeric region that contain cryptic introns (bottom). Genomic location of the introns and sequences at intron junctions are shown. 3' and 5' splice sites are indicated with bold blue lettering. The coordinates shown correspond to the February 2013 genome assembly.

Figure S6. Nrl1 and Cwf10 Regulate Formation of HOODs, Related to Figure 6 (A) Normalized number of small RNA reads at HOODs in indicated strains. The signals above and below the line

represent small RNAs that map to the top and bottom DNA strands, respectively. Small RNA-Seq data for two independent *rrp6Δ* samples and corresponding double mutants are shown. **(B)** Quantitative analysis of small RNAs at HOODs in indicated samples. Staples indicate 5 percentile (lower) and 95 percentile (upper). The lower and upper bounds of the boxes denote 25 and 75 percentile, respectively. **(C)** The normalized number of small RNA reads at the *mcp3* HOOD 12 locus in *rrp6Δ* and *rrp6Δ cwf10-1* are plotted (top). The signals above and below the line represent small RNAs that map to the top and bottom DNA strands, respectively. ChIP-chip analysis of H3K9me2 is also plotted (bottom). **(D)** Splicing of annotated introns in wild-type and *cwf10-1*. The arcs above (red) and below (blue) the line represent intron junction reads that map to the top and bottom DNA strands, respectively. The thickness of the arc corresponds to the number of reads.

Figure S7. Detection of Cryptic Introns in Non-coding RNAs and Other Parts of the Genome. Related to Figure 6 **(A)** Graphs show the comparison of the intron length distribution for annotated introns (red) to the length distribution for previously unannotated cryptic introns (black) in wild-type, *red1Δ*, *mtl1-1* and *nrl1Δ* strains. The table shows splice site utilization for annotated introns compared to cryptic introns. **(B)** Cryptic introns detected by RNA-Seq at ncRNA loci in indicated strains. The arcs above (red) and below (blue) the line represent intron junction reads that map to the top and bottom DNA strands, respectively. The thickness of the arc corresponds to the number of reads. **(C)** Splicing of annotated introns in meiotic genes in indicated strains during vegetative growth. **(D)** Cryptic introns detected by RNA-Seq at the extended 3' end of loci with convergent genes in indicated strains (top). Normalized RNA-Seq reads are shown for wild type, *mtl1-1* and *rrp6Δ* (bottom). **(E)** Example of a cryptic intron detected in the UTR region. **(F)** Alternative splicing of annotated introns detected by RNA-Seq in *nrl1Δ* (top) and *red1Δ* (bottom).

Table S1. Rmn1, Nrl1 and Cwf10 Facilitate Small RNA Production and H3K9me at Genes and Retrotransposons, Related to Figures 2, 5, and 6 (see attached Excel file).

Table S2. List of Genes and ncRNA Upregulated in *red1Δ*, *mtl1-1*, and *rrp6Δ* Mutants (log₂ ratio), Related to Figure 3 (see attached Excel file).

Table S3. List of Previously Unannotated Introns Identified in Wild-type and Different Mutants, Related to Figures 6 and S7. Introns with at least five reads are highlighted (see attached Excel file).

Table S4. List of Strains Used in This Study, Related to Experimental Procedures (see attached separate Excel file).

Extended Experimental Procedures

Strains. Standard cell culture methods and media were used. Deletion strains were made using a standard PCR based gene targeting strategy (Bahler et al., 1998). To construct a temperature-sensitive mutation in *mtl1* (*mtl1-1*), error prone PCR was used to amplify a cassette containing a FLAG-tagged *mtl1* and a Kanamycin selection marker. The PCR product was integrated into the *mtl1* locus of a wild type strain, and transformants from G418 selection plates were tested for temperature sensitivity (ts). Candidate ts colonies were confirmed. Candidates were crossed with wild type, analyzed for marker co-segregation by tetrad analysis, and sequenced for confirmation. To construct a strain with a deletion of intron 2 at *SPCC1442.04c*, the *ura4* gene was inserted and then replaced by a fragment lacking intron 2. The *mtl1-1* strain was grown at 26°C on plates, and liquid cultures were grown overnight at 30°C or 33°C to log phase for all experiments. The *rrp6Δ cwf10-1* strain was grown overnight at 26°C. For +/- phosphate media, EMM was prepared with or without 15.5 mM sodium phosphate and 20 mM potassium phosphate.

Purification and Co-Immunoprecipitation. Cells were harvested from overnight 2L cultures grown to OD₅₉₅ 1.5, washed and flash frozen in liquid nitrogen. Thawed cell pellets were ground with glass beads in a Pulverisette 6 system (Labsynergy) in buffer containing complete protease inhibitors (Roche) and 1mM PMSF. Lysate was cleared by centrifugation at 27,000 × g for 1h, and the supernatant was incubated with anti-FLAG M2 affinity gel (Sigma) or anti-c-MYC agarose affinity gel (Sigma) for 2h. Beads were washed extensively, and precipitated proteins were eluted with 200µl 1mg/ml FLAG or MYC peptides (Sigma). Proteins were precipitated

with 10% TCA and resuspended in sample buffer. Samples were separated in 4-12% Bis-Tris Gel (Invitrogen) for either Western blot analysis or mass spectrometry analysis.

Mass Spectrometry Analysis. The sample preparation and mass spectrometry analysis were performed as described (Zofall et al., 2009). In brief, the gel bands stained with Coomassie Blue were subjected to in-gel tryptic digestion to extract the peptides. After desalting, each sample was loaded on an Agilent 1200 nano-capillary HPLC system (Agilent Technologies) with a 10 cm integrated μ RPLC-electrospray ionization (ESI) emitter columns, coupled online with an Orbitrap velos mass spectrometer (Thermo Fisher Scientific) for LC-MS analysis. Peptides were eluted using a linear gradient of 2% mobile phase B (acetonitrile with 0.1% formic acid) to 42% mobile phase B within 40 min at a constant flow rate of 250 nL/min. The fourteen most intense molecular ions in the MS scan were sequentially selected for collision-induced dissociation (CID) using a normalized collision energy of 35%. The mass spectra were acquired at the mass range of m/z 380-2000, with ion source capillary voltage setting at 1.7 kV, S-Lens RF level set at 69% and ion source temperature setting at 250 °C. The MS/MS data were searched against UniProt *Schizosaccharomyces pombe* database from the European Bioinformatics Institute (<http://www.ebi.ac.uk/integr8>) using BioWorks interfaced SEQUEST (Thermo Fisher Scientific). Up to two missed cleavage sites was allowed during the database search. The cut-off for legitimate identifications were: charge state dependent cross correlation ($X_{\text{corr}} \geq 2.0$ for $[M+H]^{1+}$, ≥ 2.5 for $[M+2H]^{2+}$ and ≥ 3.0 for $[M+3H]^{3+}$ with delta correlation ($\Delta C_n \geq 0.10$).

Immunofluorescence. Paraformaldehyde fixed cells were incubated with the primary antibodies mouse anti-MYC (9E10, Covance) or mouse anti-FLAG (F7425, Sigma), at 1:1000 and 1:500 dilution, respectively. Cells were subsequently incubated with the appropriate secondary antibody, Alexa Fluor 488 anti-mouse IgG (Molecular Probes, Invitrogen) or Alexa Fluor 647 anti-rabbit IgG (Molecular Probes, Invitrogen) at 1:2000 dilutions. Immunostained cells were analyzed using a Zeiss Axioplan 2 fluorescence microscope with oil immersion objective lens (Plan Aplanachromat, 100 \times , NA 1.4, Zeiss).

Northern Analysis. Total RNA samples were treated with RNase H in the presence or absence of target-specific oligonucleotides to generate shorter fragments for better resolution and

analysis. For analysis of snR99, snR3, Rpl30-2, and TER1 RNAs, 10 µg of RNA samples were resolved on a 6% polyacrylamide TBE-Urea gel (Invitrogen). Samples were transferred to Hybond N+ nylon membrane (GE Healthcare) in 0.5X TBE buffer. For analysis of *pho1* mRNA and its upstream ncRNA, RNA samples were separated in a 1% agarose gel. RNA was transferred to BrightStar-Plus membrane (Ambion) using Northern Max transfer buffer (Ambion). For both polyacrylamide and agarose gel Northern blots, RNA was UV cross-linked using a Stratalinker (Stratagene). The ³²P-labelled single-stranded RNA probes were generated using the MAXIscript T7 kit (Invitrogen). The membrane was hybridized overnight with the probe in ULTRAhyp buffer at 65°C (Ambion).

Southern Analysis. Genomic DNA was extracted from cells that were propagated for 90 generations. The DNA was digested with EcoRI for analysis of telomere length. Southern blot analysis was done following standard procedure. The probe used is same as described previously (Hall et al., 2003).

ChIP and ChIP-chip. ChIP and ChIP-chip experiments were performed as previously described (Cam et al., 2005). Briefly, cells from exponentially growing cultures were fixed with 3% paraformaldehyde, and for epitope-tagged protein ChIP chromatin was cross-linked with 10mM dimethyl adipimidate. Cells were washed in PBS buffer, resuspended in lysis buffer (50 mM HEPES/KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% DOC), and disrupted by glass bead for lysis. The cell lysate was sonicated to shear chromatin into fragments of 500–1,000 bp, and the resulting homogenate was precleared with Protein A/G beads prior to immunoprecipitation with antibody against dimethylated H3K9 (Abcam), anti-FLAG M2 affinity gel (Sigma), or anti-MYC antibodies (3µl 9E10, Covance and 3µl A14, Santa Cruz). Antibodies were recovered with Protein A or Protein G bead slurry. Beads were washed extensively, and cross-linking was reversed by incubation at 65 °C. Immunoprecipitated DNA and DNA from whole cell crude extract (WCE) were analyzed by multiplex PCR or labeled with Cy5/Cy3 and used for microarray-based ChIP-chip analysis by hybridization to a custom 4X44K oligonucleotide array according to Agilent's recommended procedure. Custom python scripts were used to call peaks in ChIP-chip data. Briefly, the stretches of probes as arranged on the genome were identified such that a minimum 5 consecutive probes had a mean protein

enrichment of 85 percentile or above as compared to all the enrichment values in the ChIP-chip array.

RT-PCR. Total RNA was isolated using the MasterPure™ Yeast RNA Purification Kit (Epicentre) according to the manufacturer's instructions. Reverse transcription was performed using the One-step RT-PCR kit (Qiagen) with 100 ng of DNase treated total RNA.

Small RNA and RNA-Seq Library Preparation. RNA was purified for small RNA and RNA-Seq libraries from 4 OD⁵⁹⁵ units of log phase cells (0.5) using the MasterPure™ Yeast RNA Purification Kit (Epicentre) according to the manufacturer's instructions. For RNA-Seq libraries, ribosomal RNA was first removed using the Ribo-Zero™ rRNA Removal Magnetic Kit (human/mouse/rat) (Epicentre), and the library was constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre). For small RNA library preparation, small RNAs (21-25nt) were excised following denaturing electrophoresis (Urea-PAGE, 17.5%), by using SybrGold nucleic acid stain (Life Technologies) and UV transillumination. Gel-eluted RNAs were recovered with a cellulose acetate Corning® Costar® Spin-X® centrifuge tube filter (Sigma Aldrich), ethanol precipitated overnight, and resuspended in DEPC treated water. For library construction, 3' and 5' adaptor ligation and PCR amplification were performed using the NEBNext® Small RNA Library Prep Set for Illumina® (NEB) according to the manufacturer's instructions. Amplified DNA was purified using the Qiagen MinElute PCR Purification Kit (Qiagen). DNA fragments of 140-150bp were excised from a 6% PAGE gel, eluted overnight, and recovered using a cellulose acetate Corning® Costar® Spin-X® centrifuge tube filter (Sigma Aldrich). Library DNA was ethanol precipitated overnight and resuspended in TE buffer. The final library preparation was analyzed using an Agilent 2100 BioAnalyzer (Agilent).

Small-RNA Sequencing. The libraries were sequenced on the Illumina MiSeq platform. The sequencer-generated files were aligned using the commercial aligner Novoalign. The aligned files were further processed with custom Python scripts to generate files for visualization. The data was first normalized to account for multiply mapping reads, and was further normalized to million mapped reads.

RNA-Seq Analysis. The libraries were sequenced on the Illumina MiSeq Platform. The libraries generated were stranded and were sequenced as single end and paired end. The sequenced files were aligned using the TopHat (Kim et al., 2013) aligning program. The transcripts were assembled using the Cufflinks program and the Ensemble Feb 2013 version of the genome was used for aligning the reads. Annotation for the same version of the genome was used as reference annotation for Cufflinks to assemble the transcripts and to calculate the abundance of transcripts. Cuffmerge was used to pool all the assembled transcripts. The junctions.bed files generated during the Tophat alignment of the RNA-Seq data were used to obtain the locations and abundances of the introns detected in each sequencing run. Custom Python scripts were used to analyze, collate and tabulate the junction data.

Supplemental References

- Bahler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., 3rd, Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* *14*, 943-951.
- Hall, I.M., Noma, K., and Grewal, S.I. (2003). RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl. Acad. Sci. USA* *100*, 193-198.
- 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 Biol.* *14*, R36.

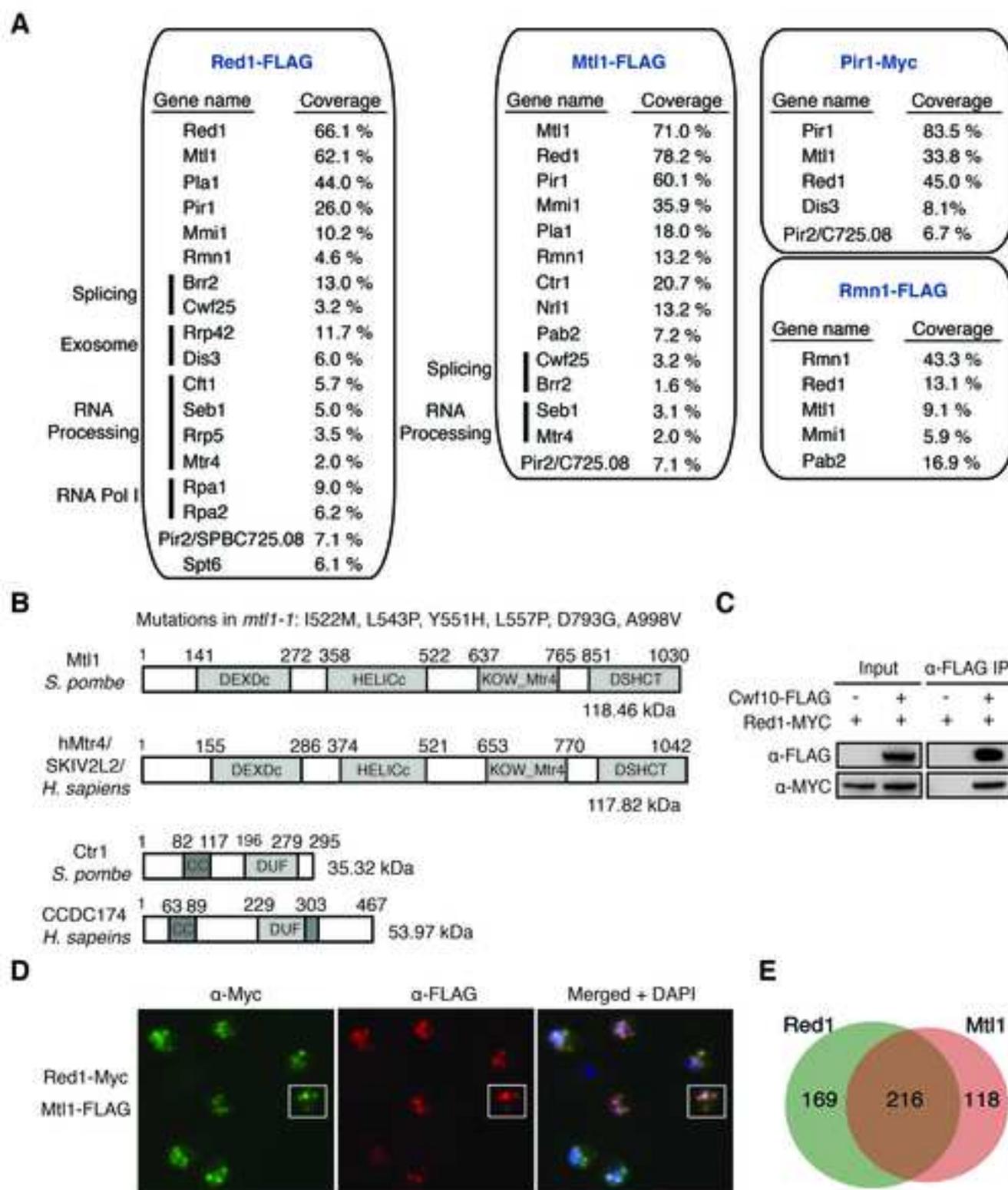


Fig. S1

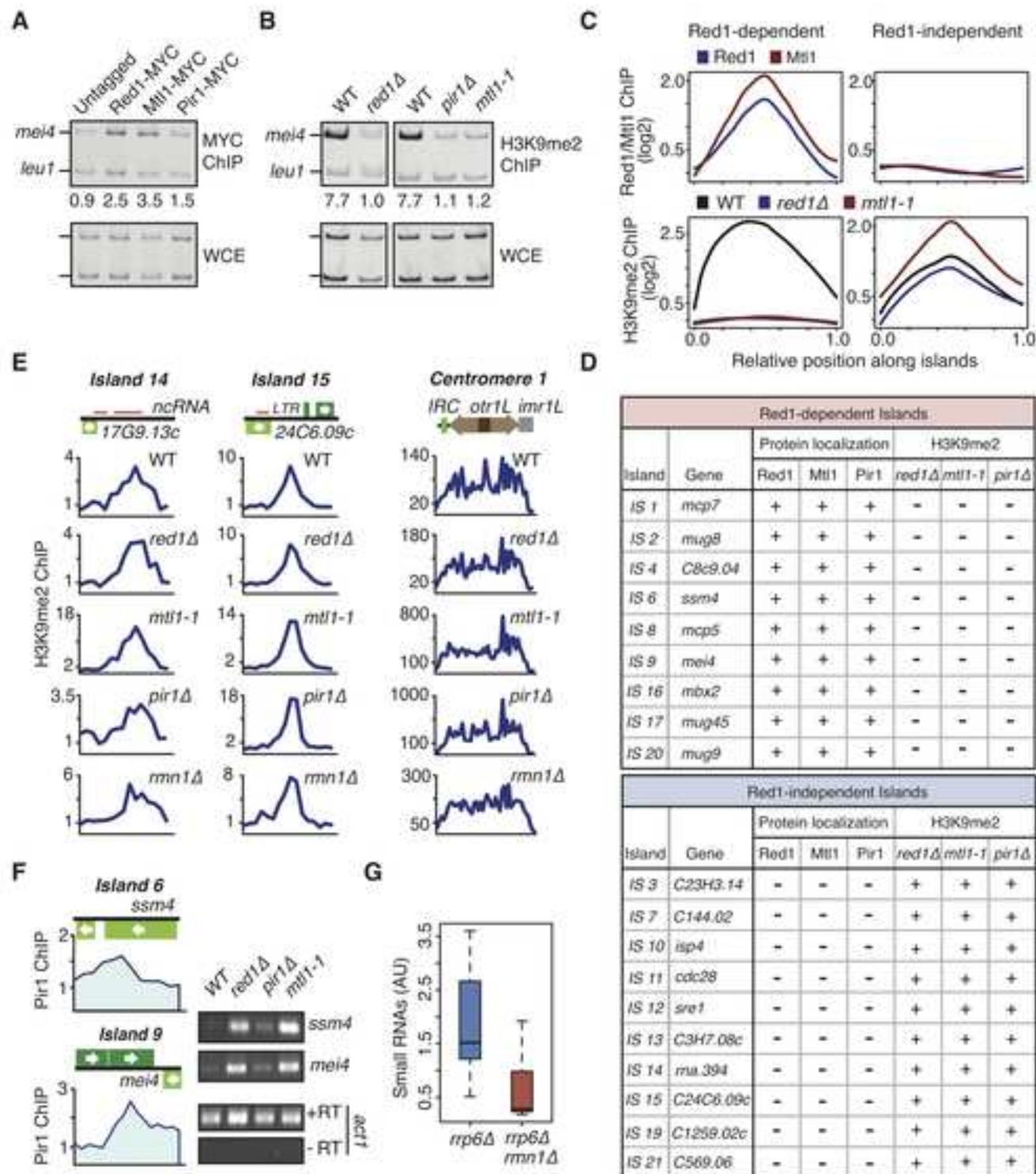


Fig. S2

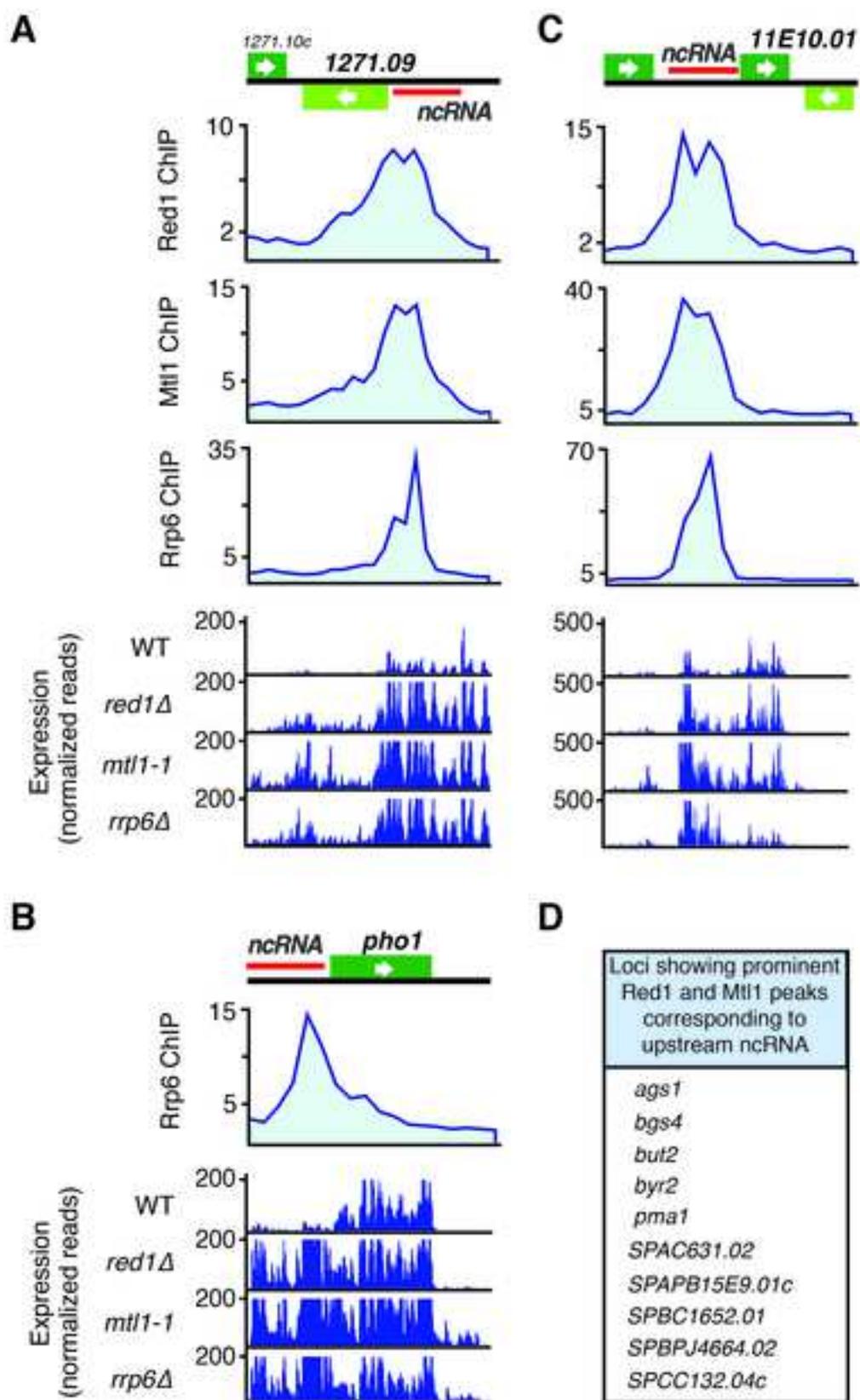


Fig. S3

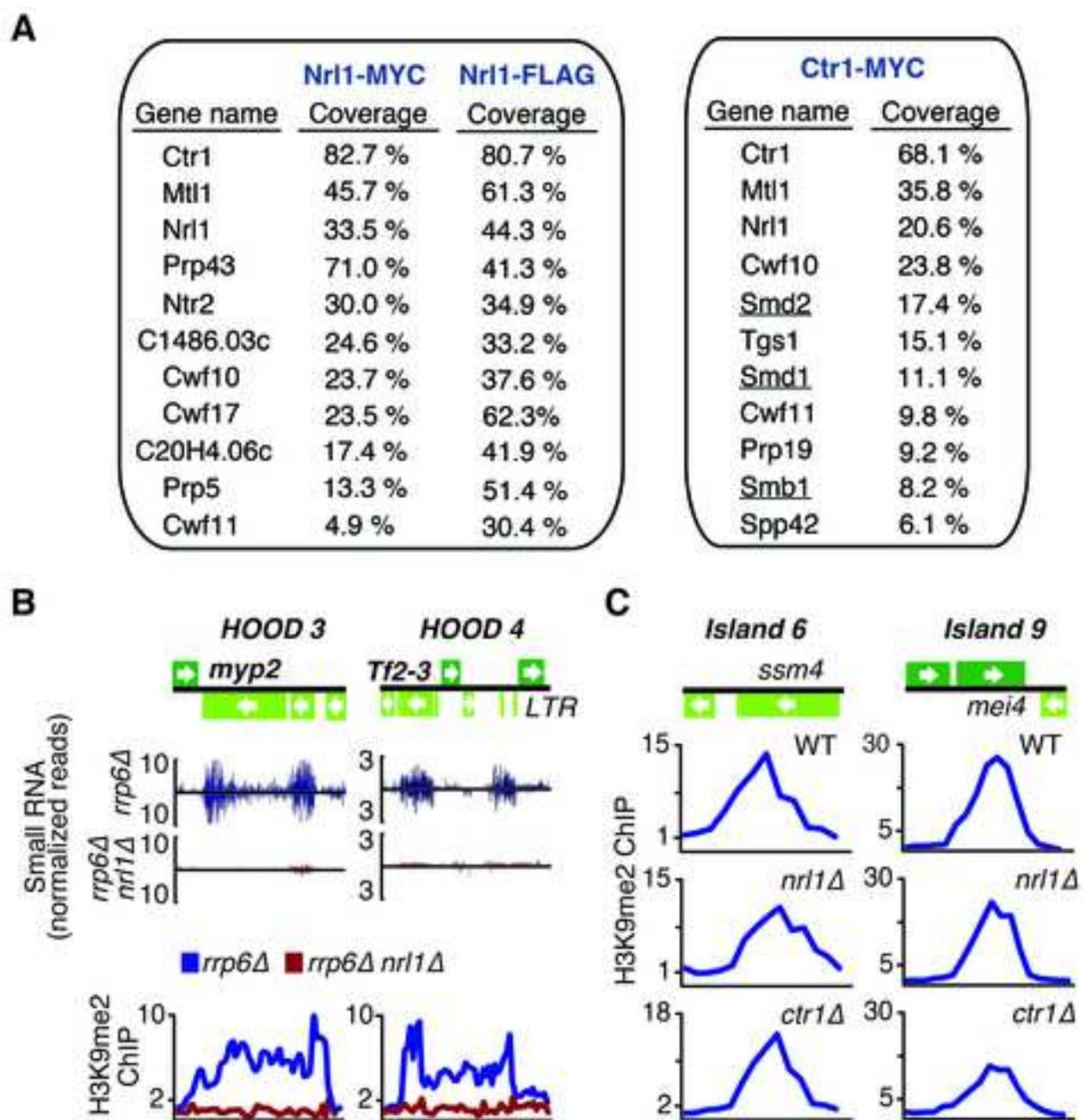


Fig. S4

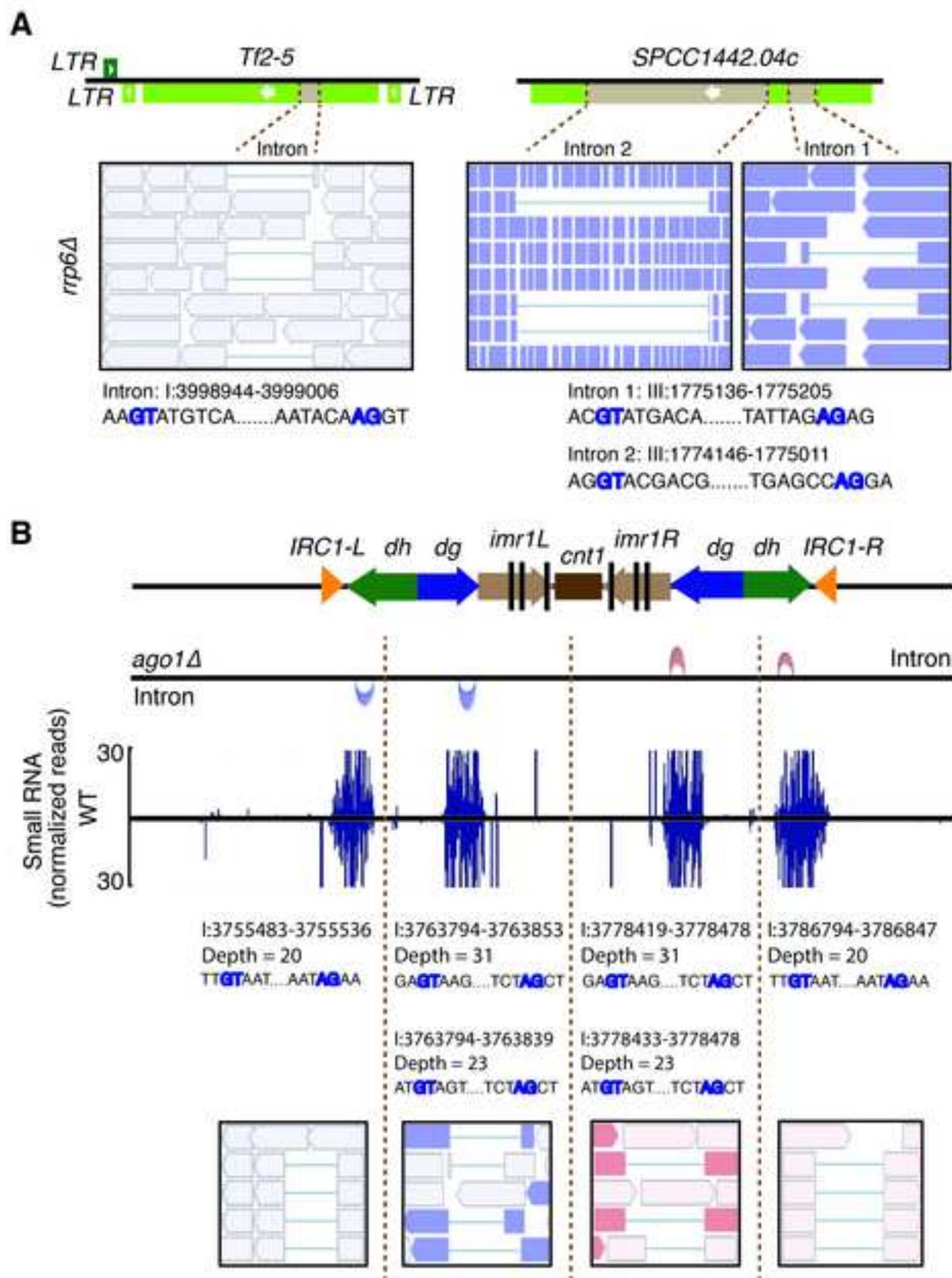


Fig. S5

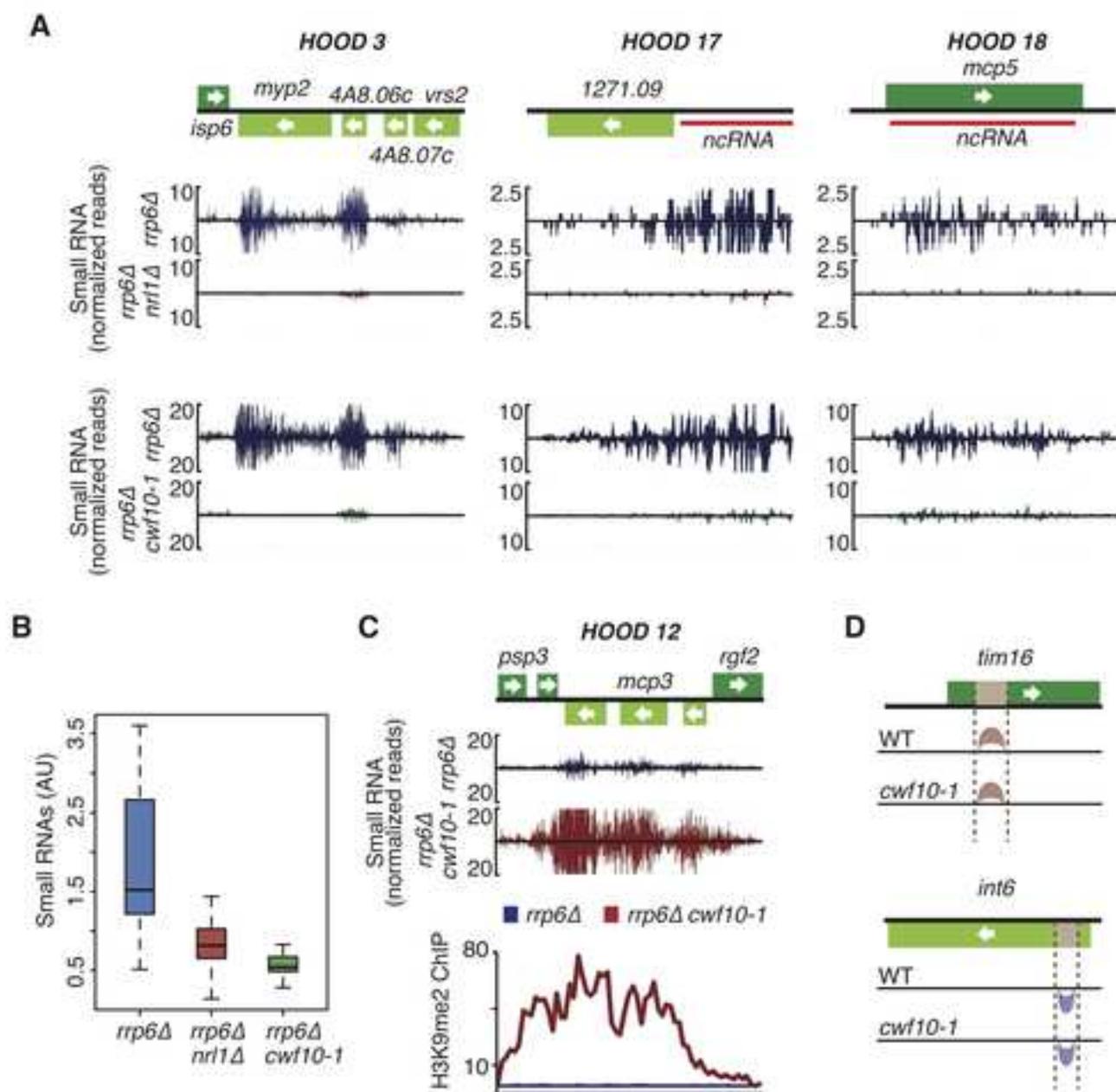


Fig. S6

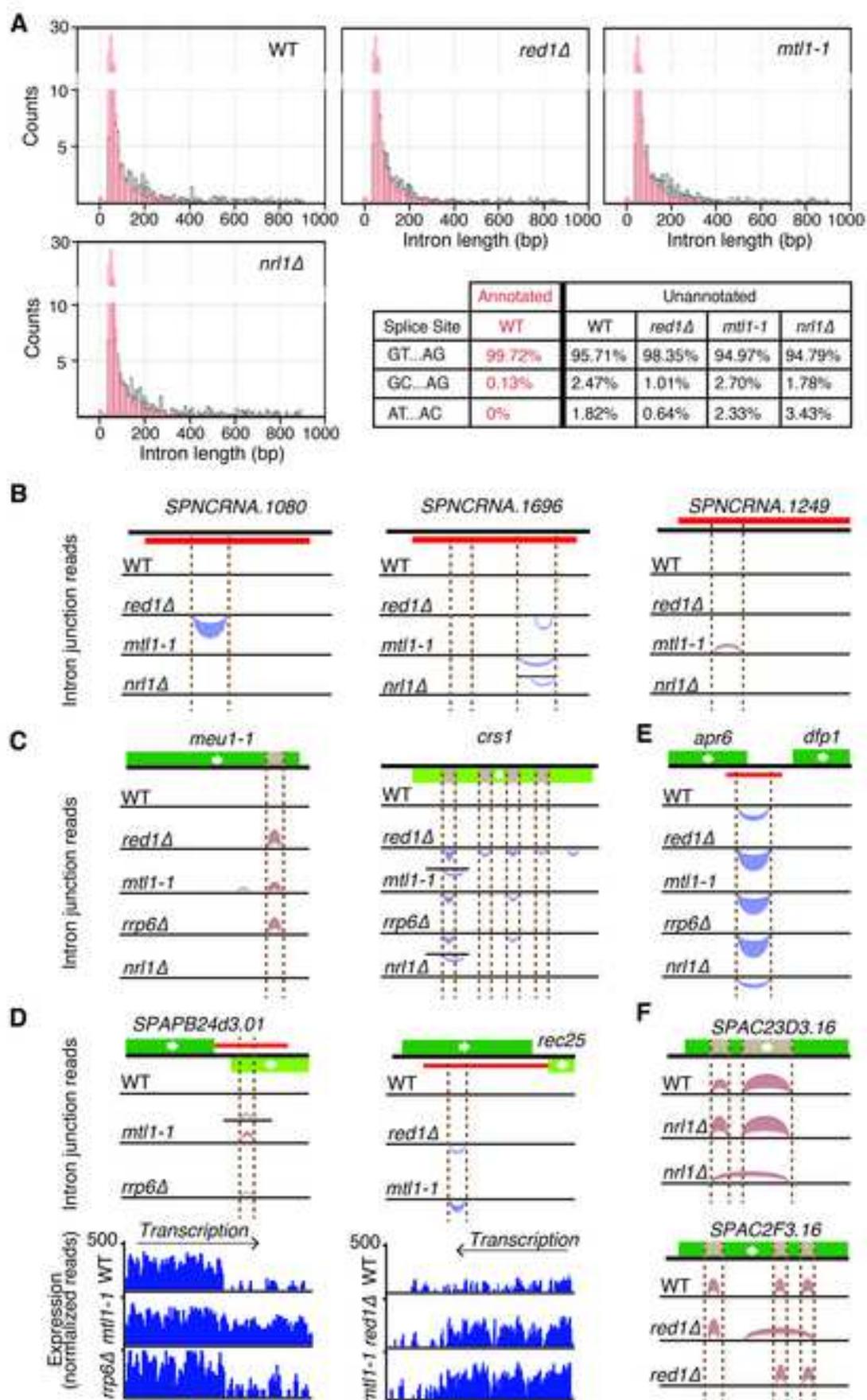


Fig. S7