SUPPLEMENTARY DATA

Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles

Belén Gómez-González[1,#], María García-Rubio[1,2,#], Rodrigo Bermejo[3], Hélène Gaillard[1,2], Katsuhiko Shirahige[4], Antonio Marín[2], Marco Foiani[3], and Andrés Aguilera[1,2,*]

[1] *Centro Andaluz de Biología Molecular y Medicina Regenerativa CABIMER, Universidad de Sevilla-CSIC, Av. Américo Vespucio s/n, 41092 Sevilla, Spain.*

[2] *Departamento de Genética, Universidad de Sevilla, 41012 Sevilla, Spain.*

[3] *Fondazione Istituto FIRC di Oncologia Molecolare and University of Milan, 20139 Milan, Italy*

[4] *Research Center for Epigenetic Disease, Institute of Molecular and Cellular Biosciences, Tokyo University, Bunkyo-ku, Tokyo 113-0032, Japan*

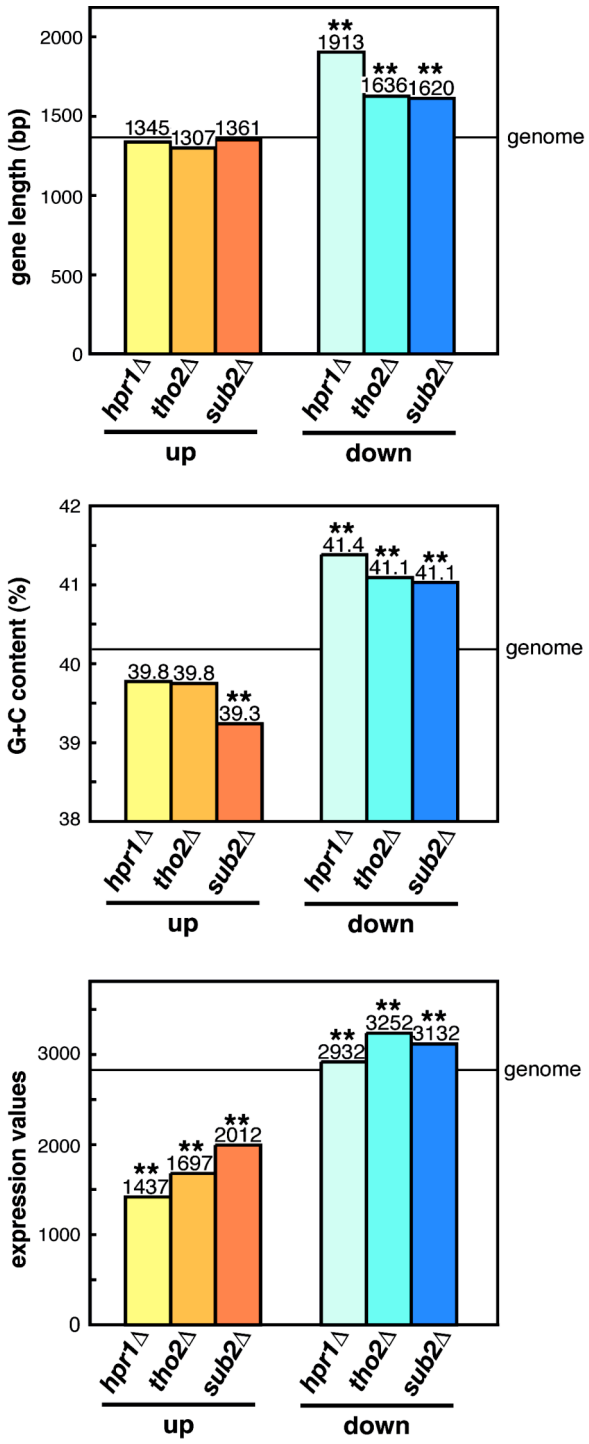
[#] These authors contributed equally to this work

* Corresponding author

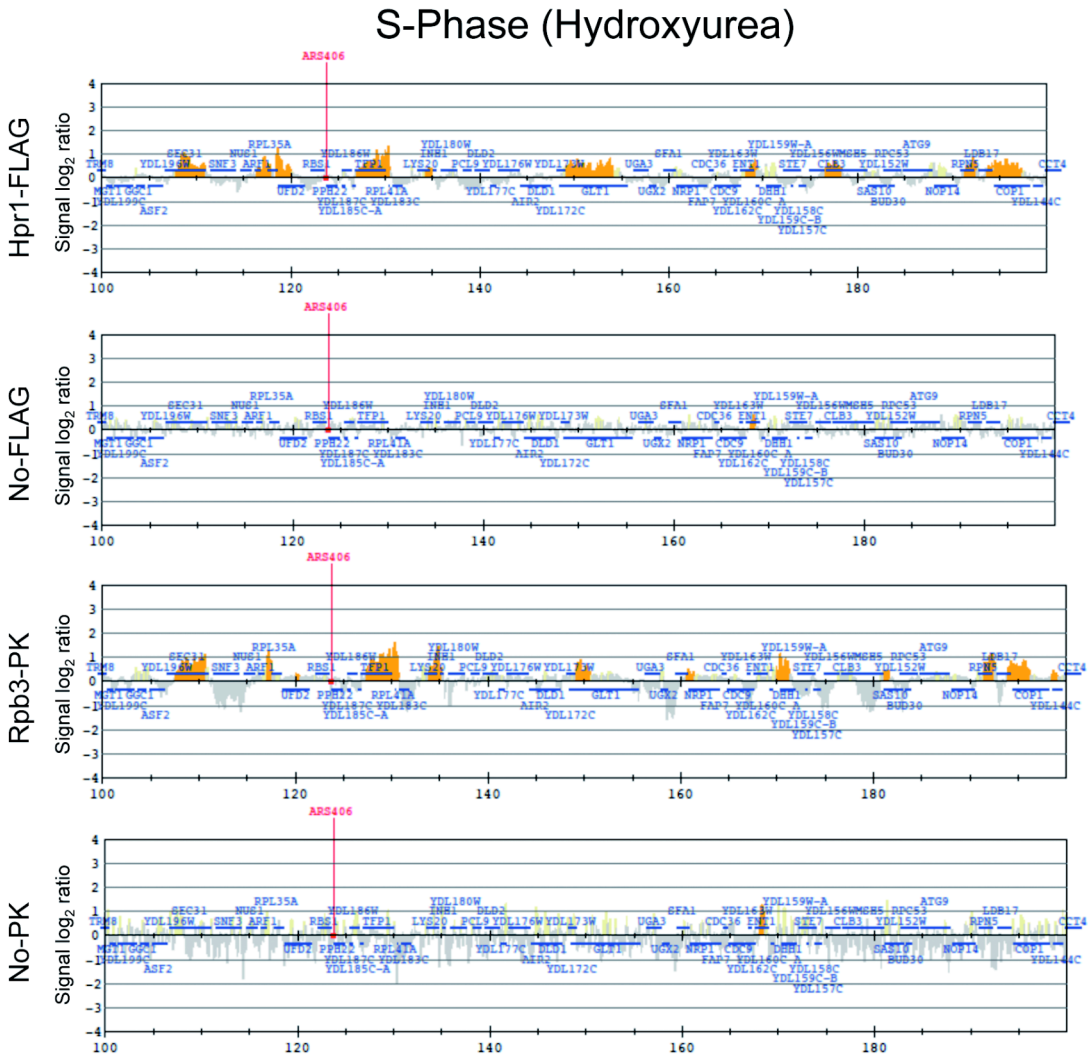
Inventory: **11 Supplementary Figures**

2 Supplementary Tables

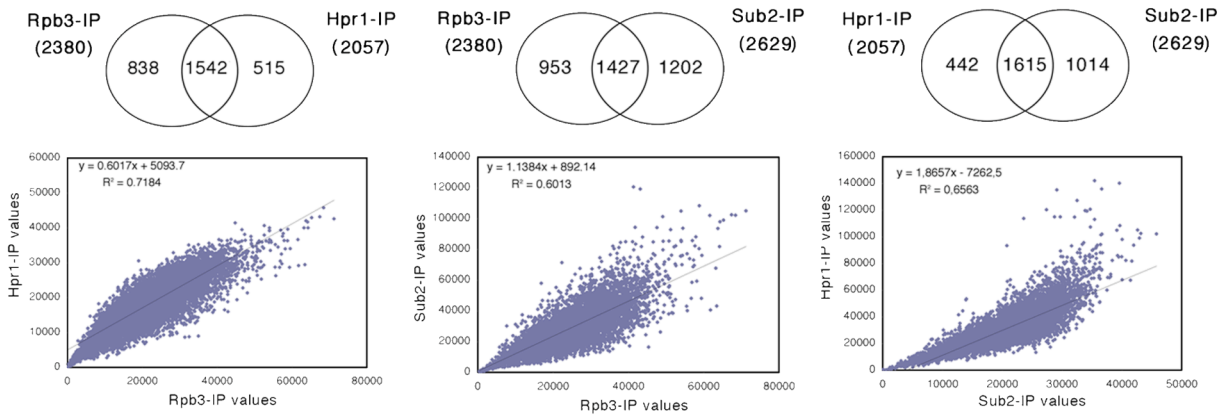
4 Supplementary Datasets



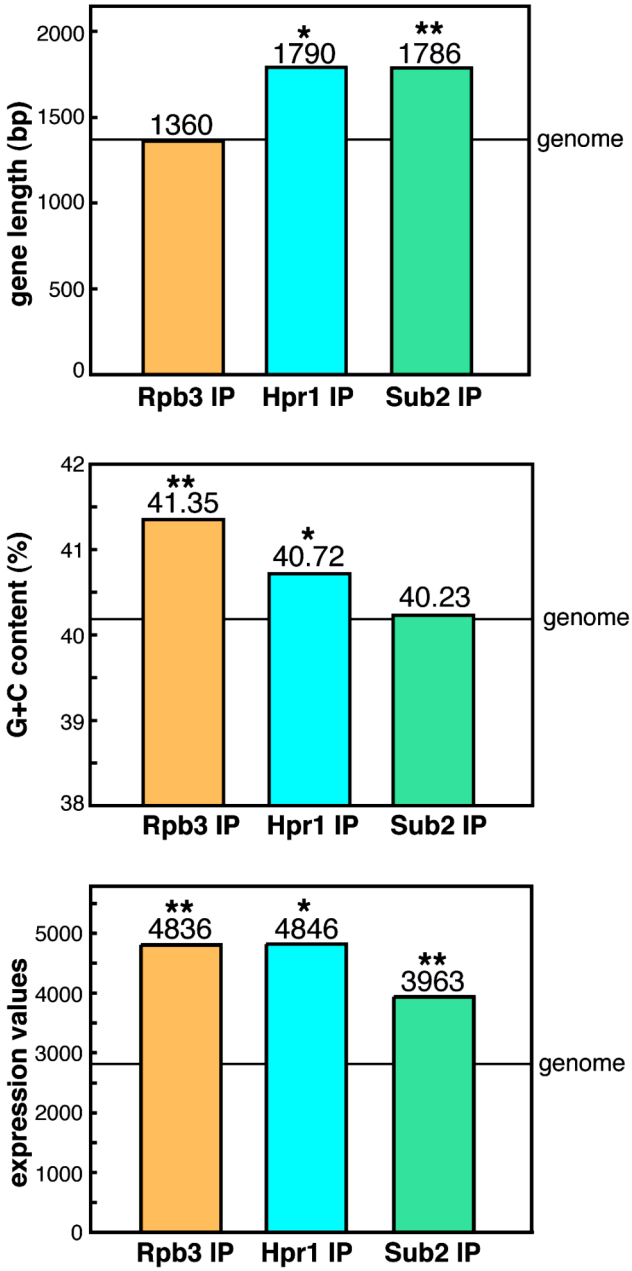
Supplementary Figure S1. Statistic analysis of length, G+C content, and model-based expression level of genes whose expression is altered in THO/TREX null mutants. Average length, G+C content and model-based expression level with respect to wild type of the genes that are either up- or down-regulated more than 1.5 fold in distinct THO/TREX mutants. Black line, genome average. **, $P < 0.0001$ (Mann-Whitney U test).



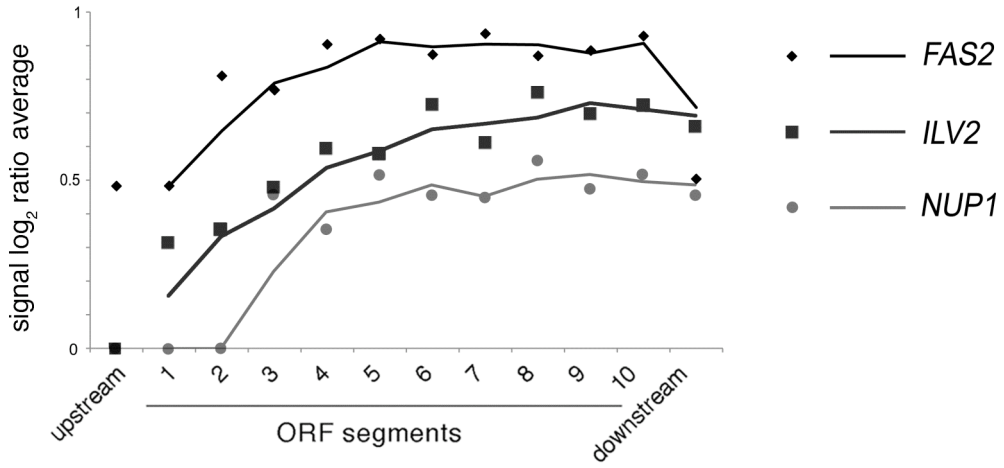
Supplementary Figure S2. ChIP-chip experiments in No-tag strains. Recruitment of an Hpr1-FLAG tagged and Rpb3-PK tagged strains versus No-tag strains. Other details as described in Fig. 2.



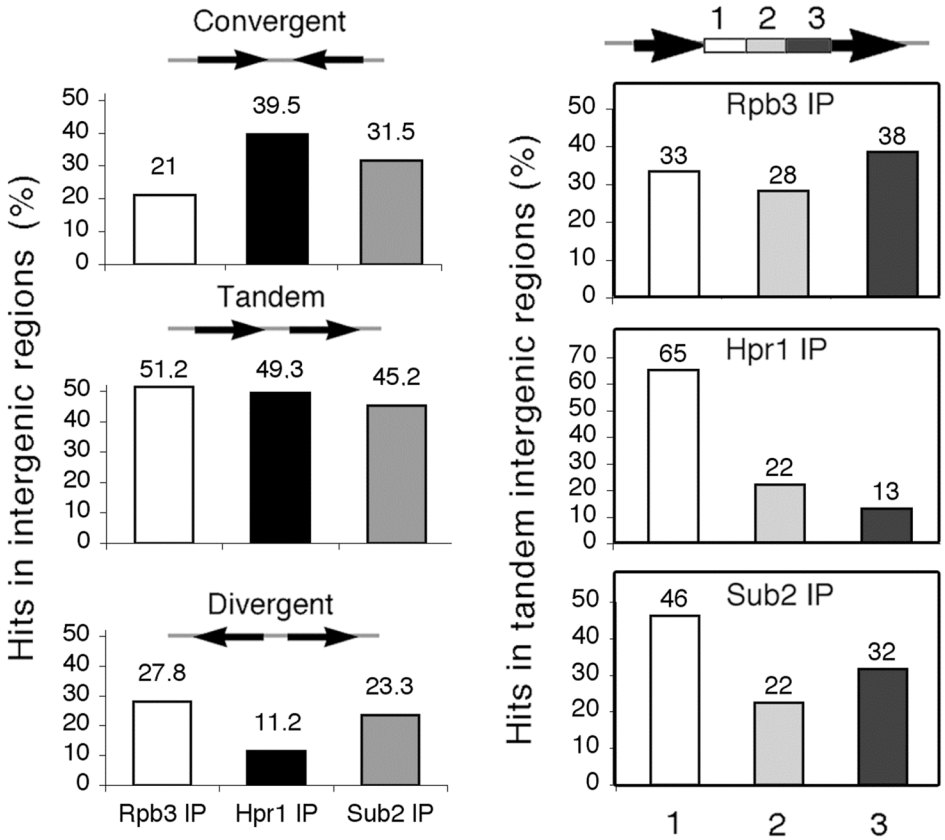
Supplementary Figure S3. Overlap analysis of genes with significant binding of Rpb3, Hpr1, and Sub2 clusters. Venn diagrams representing the overlap between genes with significant binding of Rpb3, Hpr1 and Sub2 clusters (above). Plot of Rpb3, Hpr1 and Sub2 IP values for the coincident hits (below)**.** A linear regression line and its equation are shown for each pair of data sets.



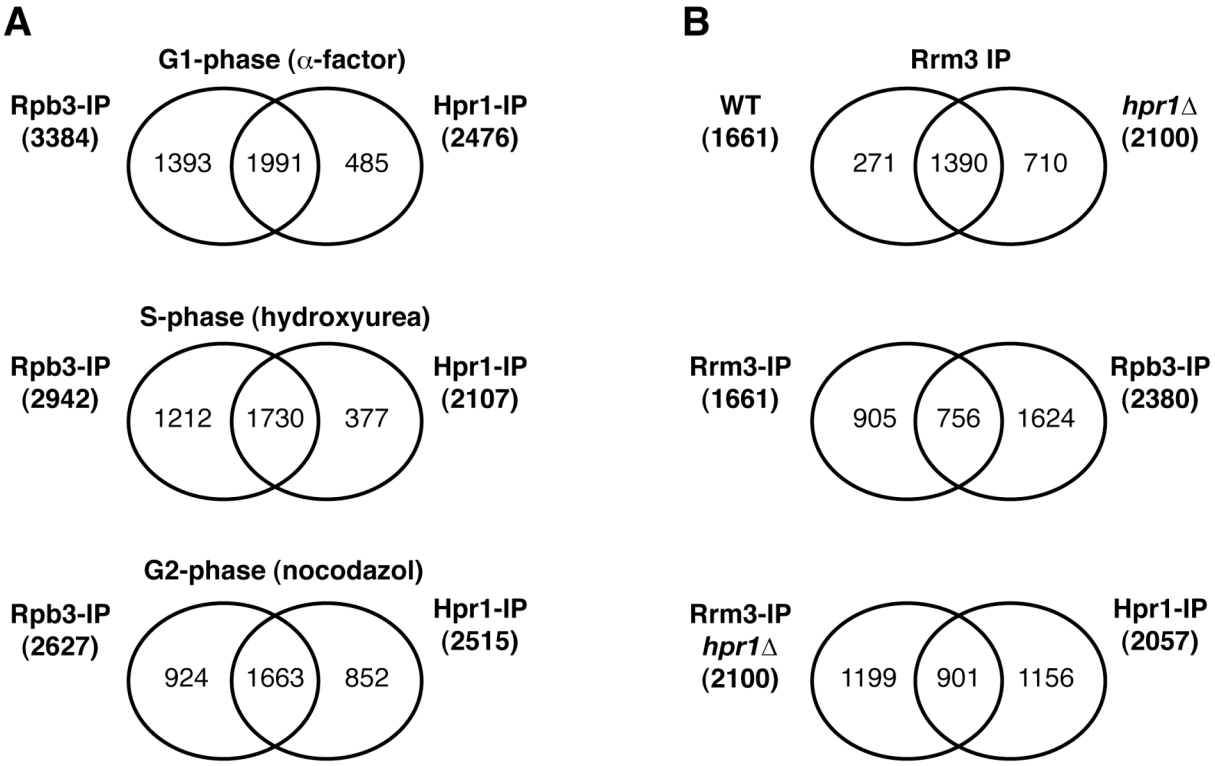
Supplementary Figure S4. Statistic analysis of length, G+C content, and wild-type model-based expression level of genes recruiting Rpb3, Hpr1 or Sub2. Average length, G+C content and model-based expression level of the genes that showed a positive recruitment of Rpb3, Hpr1 or Sub2. Black lines indicate the genome average values. *, $P < 0.001$; **, $P < 0.0001$ (Mann-Whitney U test).



Supplementary Figure S5. Binding profile of Hpr1 along the *FAS2*, *ILV2* and *NUP1* genes. Each gene was divided into 10 equivalent segments from the start and end coordinates. Two additional segments of the same size were considered upstream (5') and downstream (3') each ORF. The average signal $\log_2$ ratio for Hpr1 hits mapping on each segment was plotted. The trending line shows the moving average (period 2). The enrichment of Hpr1 in the three genes occurs further away from the promoter. The amounts of Hpr1 bound correlate with the respective gene expression levels (10630, 9279, and 2326; model-based expression units).



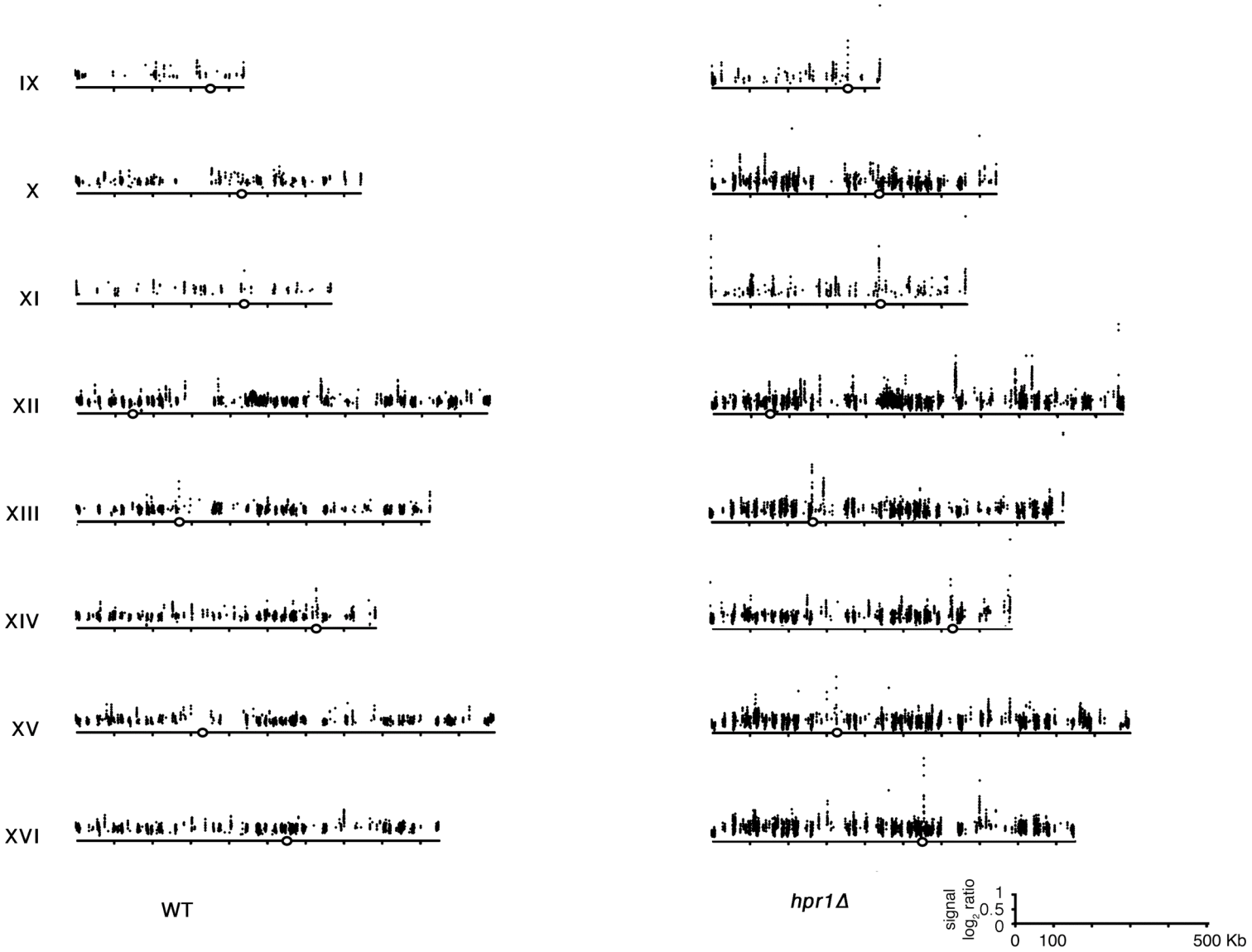
Supplementary Figure S6. THO-Sub2 recruitment increases toward the end of ORFs and at terminator regions. Percentage of Rpb3, Hpr1 or Sub2 ChIP hits in the different types (divergent, tandem and convergent) intergenic regions (left). Percentages of Rpb3, Hpr1 or Sub2 ChIP hits inside the different segments (1, 2 and 3) of tandem intergenic regions (right). Intergenic spacers accumulate 17%, 6% or 6% of total Rpb3, Hpr1 or Sub2 ChIP hits, respectively. The percentage distribution of Rpb3 ChIP hits among convergent, tandem and divergent intergenic spacers is close to 25:50:25%, which agrees with the number of the three kinds of spacers in the yeast genome. Instead, there is a clear bias toward convergent regions for Hpr1 and, to a lesser extent, for Sub2. If tandem intergenic spacers are divided in three segments (terminator, medial and promoter) we see that whereas Rpb3 is homogeneously distributed along all three, Hpr1 and, to a lesser extent, Sub2 are enriched toward the termination regions. Therefore, in contrast to Rpb3, which is recruited throughout all the ORF length, both Hpr1 and Sub2 are recruited following an increasing pattern toward the end of ORFs, being absent from promoters.



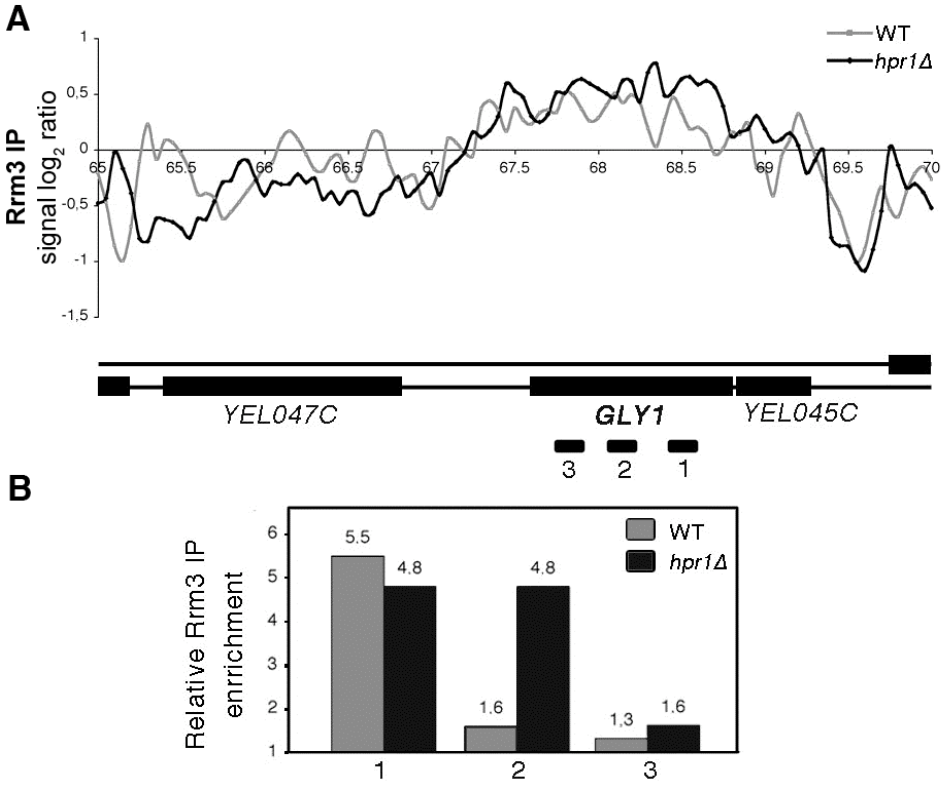
Supplementary Figure S7. Gene overlap between different ChIP-chip experiments. A. Venn diagrams showing the overlap between genes with significant binding of Rpb3 and Hpr1 in each cell cycle phase are shown. **B.** Venn diagrams showing the overlap between genes with significant binding of Rrm3 in wild-type or *hpr1Δ* and the binding of Rpb3 or Hpr1 in asynchronous cultures are shown.

.

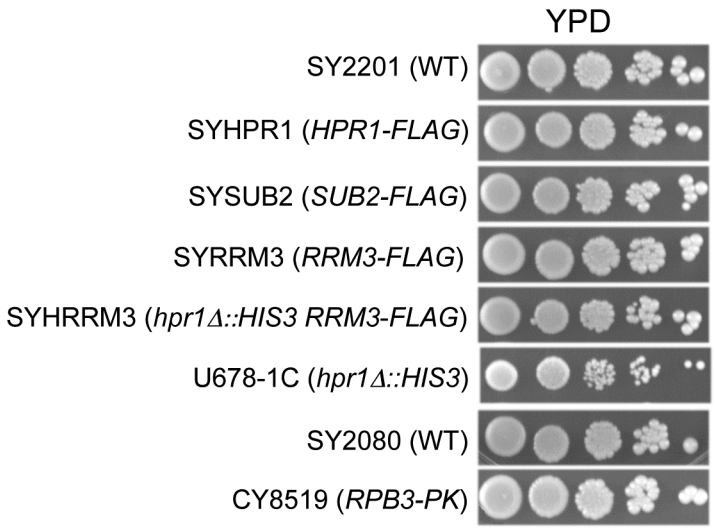

I
II
III
IV
V
VI
VII
VIII
WT
hpr1Δ
signal log2 ratio
1
0.5
0
0
100
500 Kb


IX
X
XI
XII
XIII
XIV
XV
XVI
WT
hpr1Δ
signal log2 ratio
1
0.5
0
0
100
500 Kb

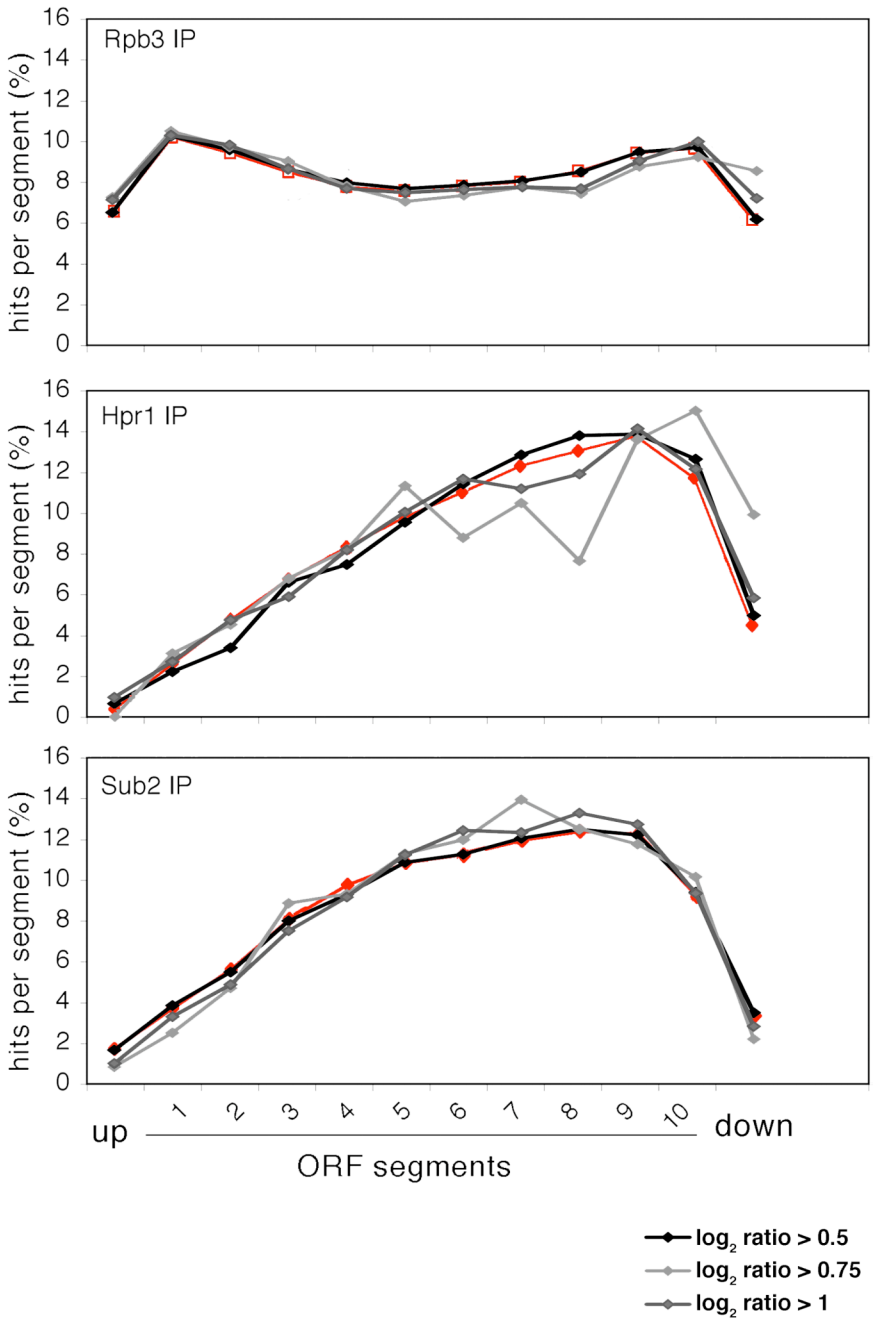
Supplementary Figure S8. Genomic view of Rrm3 recruitment in wild-type and *hpr1Δ* cells. A representation of each chromosome with the signal $\log_2$ ratio values for the significant ChIP hits is plotted. The X axis shows chromosomal coordinates in kb. Positions of centromeres are indicated as open circles.



Supplementary Figure S9. Rrm3 recruitment at the *GLY1* gene in wild-type and *hpr1Δ* cells. A. Detailed analysis of ChIP-chip data of Rrm3-FLAG at the *GLY1* region. **B.** Specific ChIP analysis of Rrm3-FLAG recruitment using RT-qPCR of three regions of the *GLY1* gene in wild-type and *hpr1Δ* cells.



Supplementary Figure S10. Analysis of growth of the different tagged strains used in this study. Serial dilutions (10-fold) from exponentially growing cultures are shown.



Supplementary Figure S11. THO-Sub2 recruitment increases toward the end of ORFs and at terminator regions. Composite profile of Rpb3, Hpr1, and Sub2 occupancy (detected by ChIP) across the average ORF plotted as Rpb3, Hpr1, and Sub2 percentage of ChIP hits per segment when considering a detection and change p values < 0.01 and different signal $\log_2$ ratio cut offs (> 0.5, > 0.75, or > 1) for at least three contiguous loci. Red profiles correspond to the data depicted in Figure 3A (p values < 0.025 and signal $\log_2$ ratio > 0).

Supplementary Table I: Results of the placement simulation for Rpb3, Hpr1, and Sub2 on various genomic areas.

Comparisson			Simulation on number of bases overlapping				
Fixed track	**Randomized track**	**Hits**	**Expected**	**Change**	**Ratio**	**p Value**	**Rnd**
Rpb3 asyn	Rpb3 asyn	1934479	374314+/-17672	increase	5.17	0	0/1000
Rpb3 asyn	Hpr1 asyn	926961	291019+/-16114	increase	3.19	0	0/1000
Rpb3 asyn	Sub2 asyn	299000	126344+/-10929	increase	2.37	1.60E-56	0/1000
Hpr1 asyn	Rpb3 asyn	926961	290494+/-16153	increase	3.19	0	0/1000
Hpr1 asyn	Hpr1 asyn	1654620	241845+/-14787	increase	6.84	0	0/1000
Hpr1 asyn	Sub2 asyn	488667	117854+/-10563	increase	4.15	2.80E-270	0/1000
Sub2 asyn	Rpb3 asyn	299000	124271+/-11165	increase	2.41	1.70E-55	0/1000
Sub2 asyn	Hpr1 asyn	488667	115542+/-10705	increase	4.23	1.80E-266	0/1000
Sub2 asyn	Sub2 asyn	858109	68805+/-8286	increase	12.47	0	0/1000

Comparisson			Simulation on number of bases overlapping				
Fixed track	**Randomized track**	**Hits**	**Expected**	**Change**	**Ratio**	**p Value**	**Rnd**
Rpb3 G1	Rpb3 G1	4318296	1587718+/-27359	increase	2.72	0	0/1000
Rpb3 G1	Hpr1 G1	2056674	923314+/-24638	increase	2.23	0	0/1000
Hpr1 G1	Rpb3 G1	2056674	921236+/-23385	increase	2.23	0	0/1000
Hpr1 G1	Hpr1 G1	2587119	576633+/-21848	increase	4.49	0	0/1000

Comparisson			Simulation on number of bases overlapping				
Fixed track	**Randomized track**	**Hits**	**Expected**	**Change**	**Ratio**	**p Value**	**Rnd**
Rpb3 S	Rpb3 S	3159303	839562+/-23144	increase	3.76	0	0/1000
Rpb3 S	Hpr1 S	1764291	586694+/-20625	increase	3.01	0	0/1000
Hpr1 S	Rpb3 S	1764291	584778+/-21014	increase	3.02	0	0/1000
Hpr1 S	Hpr1 S	2200993	411822+/-18459	increase	5.34	0	0/1000

Comparisson			Simulation on number of bases overlapping				
Fixed track	**Randomized track**	**Hits**	**Expected**	**Change**	**Ratio**	**p Value**	**Rnd**
Rpb3 G2	Rpb3 G2	3100360	823347+/-25757	increase	3.77	0	0/1000
Rpb3 G2	Hpr1 G2	1373746	598492+/-21017	increase	2.3	3.70E-298	0/1000
Hpr1 G2	Rpb3 G2	1373746	597195+/-21655	increase	2.3	6.30E-282	0/1000
Hpr1 G2	Hpr1 G2	2308966	452993+/-17571	increase	5.1	0	0/1000

Comparisson			Simulation on number of bases overlapping				
Fixed track	**Randomized track**	**Hits**	**Expected**	**Change**	**Ratio**	**p Value**	**Rnd**
Rrm3 WT	Rrm3 WT	1549270	202148+/-12605	increase	7.66	0	0/1000
Rrm3 WT	Rrm3 *hpr1Δ*	793816	193659+/-12790	increase	4.1	0	0/1000
Rrm3 *hpr1Δ*	Rrm3 WT	793816	195306+/-12810	increase	4.06	0	0/1000
Rrm3 *hpr1Δ*	Rrm3 *hpr1Δ*	1495990	191518+/-12965	increase	7.81	0	0/1000

Hits: number of peaks of the clusters actually falling into the indicated genomic area.
Expected: number of hits achieved by random repositioning of the genomic areas (average +/- standard deviation).
Change: direction of the change of Hits respect to the expected (significant changes are indicated in bold, see p Value and Rnd).
Ratio: ratio between the actual number of hits and the random average.
p Value: indicates the probability that this ratio is casual.
Rnd: number of randomized cases that achieved a ratio greater than or equal to the actual one.

Supplementary Table II. Yeast Strains used in this study.

Strain	Genotype	Source
SY2080	*MAT***a** *ade2-1 ura3 leu2-3,112 his3-11,15 can1-100 GAL+ PSI+*	H. Klein
CY8519	*MAT***a** *ade2-1 ura3 leu2-3,112 his3-11,15 can1-100 RPB3::rpb3-6PKTRP1 GAL+ PSI+*	M. Foiani
SY2201	*MAT***a** *ade2-1 ura3 leu2-3,112 his3-11,15 can1-100 ura3::URA3/GPD-TK(7X) GAL+ PSI+*	E. Schwob
SYHPR1	SY2201 *HPR1-FLAG*	This study
SYSUB2	SY2201 *SUB2-FLAG*	This study
SYRRM3	SY2201 *RRM3-FLAG*	This study
SYHRRM3	SY2201 *RRM3-FLAG hpr1Δ::HIS3*	This study
W303-1A	*MAT***a** *ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1*	R. Rothstein
U678-1C	*MAT***a** *ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 hpr1Δ::HIS3*	R. Rothstein
SChY58a	*MAT***a** *ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 hpr1Δ::kanMX4*	García-Rubio et al., 2008
RK2-6C	*MAT***a** *ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 tho2Δ::KanMX4*	Piruat and Aguilera, 1998
DLY23 sub2Δ	*MAT***a** *ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 sub2Δ::HIS3*	Libri *et al.*, 2001
WHRB-2C	*MAT***a** *ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 bar1Δ rrm3Δ::KanMX6 hpr1Δ::HIS3*	This study

Supplementary References

Garcia-Rubio M, Chavez S, Huertas P, Tous C, Jimeno S, Luna R, Aguilera A (2008) Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. *Mol Genet Genomics* **279:** 123-132

Piruat JI, Aguilera A (1998) A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *EMBO J* **17:** 4859-4872

Libri D, Graziani N, Saguez C, Boulay J (2001) Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev* **15:** 36-41

Supplementary Dataset I: Excel Spreadsheet containing the list of genes that are either down- or up-regulated in the *tho2Δ*, *hpr1Δ*, and *sub2Δ* mutants, the Gene Ontology results for the genes with altered expression levels in the *tho2Δ*, *hpr1Δ*, and *sub2Δ* mutants, and the Gene Ontology results for the Rrm3-bound genes. A. The genes for which the expression levels are either down- or up-regulated in at least one of the *tho2Δ*, *hpr1Δ*, or *sub2Δ* mutant are shown. Significant changes in expression are shown as 1, no changes in expression as 0. **B.** The GO results obtained using the 51 up-regulated (upper table) and the 221 down-regulated (lower table) genes common to the *tho2Δ*, *hpr1Δ*, and *sub2Δ* mutants (see Figure 1) are shown. The GO Term Finder application (Version 0.83) provided by the *Saccharomyces Genome Database* was used. **C.** The GO results obtained using the 52 top Rrm3-bound genes in wild-type (upper table) and the 137 top Rrm3-bound genes in *hpr1Δ* (lower table) are shown. The GO Term Finder application (Version 0.83) provided by the *Saccharomyces Genome Database* was used.

Dataset can be found in file: Supplementary Dataset I.pdf

Supplementary Dataset II. Whole chromosome data for Rpb3-IP, Hpr1-IP, and Sub2-IP ChIP-chips performed in this study. Rpb3-IP, Hpr1-IP, and Sub2-IP patterns as histogram bars in the y axis show the average signal ratio of loci significantly enriched in the immunoprecipitated fraction along the indicated regions in $\log_2$ scale. The x axis shows chromosomal coordinates. Positions of ARS elements are indicated. The horizontal bars mark the positions of the indicated ORFs.

Dataset can be found in file: Supplementary Dataset II.pdf

Supplementary Dataset III. Whole chromosome data for Hpr1-IP ChIP-chips from different cell cycle stages. Hpr1-IP in G1, G2 and S phase patterns as histogram bars in the y axis show the average signal ratio of loci significantly enriched in the immunoprecipitated fraction along the indicated regions in log2 scale. The x axis shows chromosomal coordinates. Positions of ARS elements are indicated. The horizontal bars mark the positions of the indicated ORFs.

Dataset can be found in file: Supplementary Dataset III.pdf

Supplementary Dataset IV. Whole chromosome data for Rrm3-IP ChIP-chips performed in this study. Rrm3-IP in both wild-type and *hpr1Δ* cells patterns as histogram bars in the y axis show the average signal ratio of loci significantly enriched in the immunoprecipitated fraction along the indicated regions in $\log_2$ scale. The x axis shows chromosomal coordinates. Positions of ARS elements are indicated. The horizontal bars mark the positions of the indicated ORFs

Dataset can be found in file: Supplementary Dataset IV.pdf