Genes with internal repeats require the THO complex for transcription

Vladimir Voynov, Kevin J. Verstrepen, An Jansen, Vanessa M. Runner, Stephen Buratowski, and Gerald R. Fink

PNAS 2006;103;14423-14428; originally published online Sep 18, 2006;
doi:10.1073/pnas.0606546103

Genes with internal repeats require the THO complex for transcription

Vladimir Voynov*†, Kevin J. Verstrepen‡§, An Jansen*, Vanessa M. Runner¶, Stephen Buratowski¶, and Gerald R. Fink*†‖

*Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142; †Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; ‡Bauer Center for Genomics Research, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138; §Department of Microbial and Molecular Systems, Faculty of Bioscience Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 22, B-3001 Leuven, Belgium; ¶Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115

Contributed by Gerald R. Fink, August 1, 2006

The evolutionarily conserved multisubunit THO complex, which is recruited to actively transcribed genes, is required for the efficient expression of *FLO11* and other yeast genes that have long internal tandem repeats. *FLO11* transcription elongation in Tho⁻ mutants is hindered in the region of the tandem repeats, resulting in a loss of function. Moreover, the repeats become genetically unstable in Tho⁻ mutants. A *FLO11* gene without the tandem repeats is transcribed equally well in Tho⁺ or Tho⁻ strains. The Tho⁻ defect in transcription is suppressed by overexpression of topoisomerase I, suggesting that the THO complex functions to rectify aberrant structures that arise during transcription.

adhesion | Hpr1 | Thp2 | topoisomerase I

Transcription involves a highly orchestrated series of events in which the core polymerase is joined by many additional proteins that promote initiation, elongation, and termination (1–3). Efficient transcription also depends on the configuration of the DNA template because transcription creates negative supercoils behind the polymerase and positive supercoils ahead of it (4–6). These alterations in the superhelical density could permit repetitive sequences to form structures that impede the progress of the polymerase and promote recombination. For example, excessive supercoiling in yeast leads to hyperrecombination at the highly repetitive rRNA-encoding DNA locus (7, 8). The DNA landscape may therefore influence the efficiency of transcription, and some of the elongation factors could be required to remodel the template to permit efficient transcription.

The *Saccharomyces cerevisiae* multisubunit THO complex, which has been identified as a possible elongation component, has been associated with many aspects of RNA and DNA metabolism (9–12). The complex consists of four tightly bound proteins (Hpr1, Tho2, Thp1, and Mft1) (13), two of which (Hpr1 and Tho2) are conserved from yeast to humans (14). Biochemical studies using natural templates have implicated the THO complex in recruiting the mRNA export proteins Sub2 (UAP56 in humans) and Yra1 (Aly1) to the mRNA in both yeast (15) and humans (14). In yeast, ChIP immunoprecipitation experiments indicate that the THO complex is recruited to actively transcribed genes (16–18).

The biochemical analysis of the function of the THO complex has not led to a consistent picture. Experiments using a *GAL1* promoted *Escherichia coli lacZ* reporter construct expressed in yeast suggested that transcription elongation of the *lacZ* gene is reduced in an *hpr1*Δ mutant (19). Further analysis using a P_{GAL}-*lacZ* system indicated that in a Tho⁻ mutant DNA:RNA hybrids are formed *in vivo* between the nascent transcript and the DNA template (20). Because the transcription of GC-rich *lacZ* constructs was THO-dependent, whereas that of many endogenous yeast genes was not, it was proposed that the THO complex is required for efficient transcription elongation of long and GC-rich genes (21). Moreover, the role of the THO complex in elongation has been questioned based on the insensitivity of Tho⁻ mutants to mycophenolic acid, a presumed inhibitor of transcription elongation (22).

Remarkably, the genetic analysis of Tho⁻ mutants has not resolved these puzzles and has provided little information on native genes that require THO complex function. Mutations in any of the four genes encoding the THO complex subunits do not result in inviability at normal growth conditions, suggesting that the THO proteins are not a core component of the elongation complex. However, one class of Tho⁻ mutants [hyperrecombination 1 (*hpr1*)] was first identified because a mutation in that gene increases the frequency of recombination between artificial tandem repeats constructed by transformation (23). Sequence similarity between Hpr1 and the topoisomerase Top1 as well as the lethality of *top1*Δ *hpr1*Δ double mutants (23, 24) are likely to reflect functional redundancy with respect to DNA metabolism. In *Drosophila*, loss of THO complex function results in only minor differences in transcription profiles as revealed by whole genome arrays (25). In both *Drosophila* and yeast, the apparent participation of the THO complex in some aspects of transcription and recombination contrasts with the absence of an effect of Tho⁻ mutations on resident genes.

In this report, we show that THO function is required for the transcription of several resident yeast genes containing multiple internal tandem repeats. The affected genes are not especially long, and neither the genes nor the repeats are GC-rich. The defect in transcription appears to be in transcription elongation, based on ChIP experiments designed to reveal RNA polymerase occupancy. Transcription is restored in Tho⁻ mutants when the repeats are removed from the gene. Because whole genome arrays comparing Tho⁺ and Tho⁻ strains do not reveal any general defects in transcription, these effects appear to be restricted to a subset of genes with internal repeats. The fact that the transcriptional defects in Tho⁻ mutants can be suppressed by overexpression of *TOP1* suggests a model in which the THO complex functions as an accessory complex that facilitates transcription past obstructive DNA configurations.

Results

***FLO11*-Dependent Adhesion Requires the THO Complex.** The gene knockout library of *S. cerevisiae* containing all viable single-gene deletions was screened to identify genes that are required for *FLO11* function. *FLO11*, a gene with many long internal tandem repeats, confers adhesion of cells to inert substrates, such as agar (26, 27). The screen used a P_{TEF}*FLO11* construct in which the *FLO11* gene was transcribed from the constitutive *TEF* promoter. This construct confers adherence to solid agar in S288c strains (Fig. 1*A*) and was used to avoid isolating mutations in

Author contributions: V.V., K.J.V., S.B., and G.R.F. designed research; V.V., A.J., and V.M.R. performed research; V.V., K.J.V., A.J., V.M.R., and S.B. contributed new reagents/analytic tools; V.V., V.M.R., S.B., and G.R.F. analyzed data; and V.V. and G.R.F. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: rtPCR, real-time PCR; YPD, yeast extract/peptone/dextrose; SC, synthetic complete media.

‖To whom correspondence should be addressed. E-mail: gfink@wi.mit.edu.

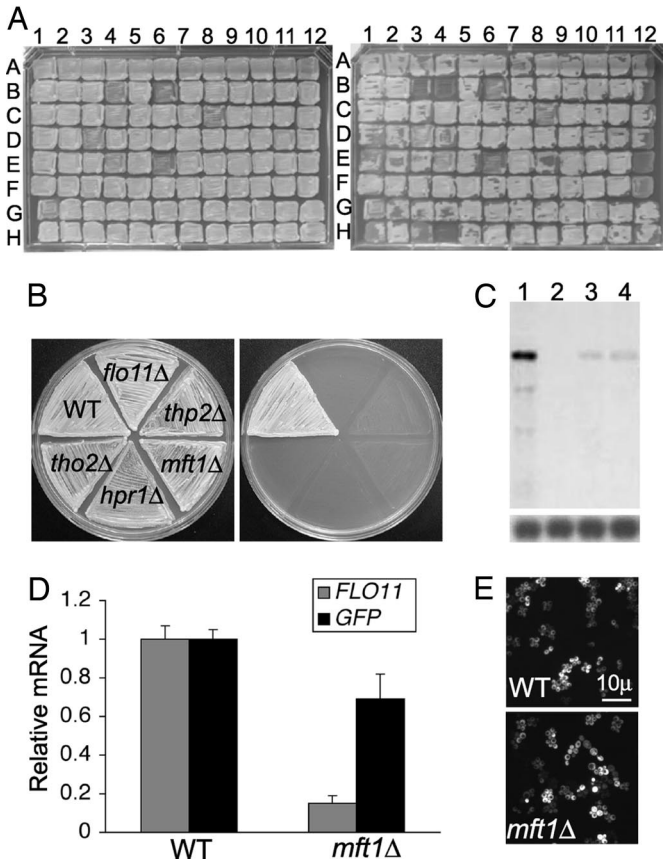

Fig. 1. The THO complex is required for *FLO11* expression. (*A*) A screen for promoter-independent factors required for *FLO11* function. (*Left*) Each mutant of the S288c deletion library was transformed with a P_{TEF}*FLO11* construct (B4126), patched on YPD plates, and grown at 30°C for 1 day. (*Right*) Adherence to agar was assayed after a wash of the plate. This wash removes cell patches that lack *FLO11* function because these cells fail to adhere to the agar. Plate 15 of the Invitrogen MATa collection is shown. The parental strain BY4741 (coordinate B3) is used as a negative control. BY4741 with P_{TEF}*FLO11* (coordinate H2) is a positive control. Whereas most mutant strains with P_{TEF}*FLO11* adhere to agar, the *thp2*Δ mutant shows a strong nonadherent phenotype (coordinate H4). (*B*) THO complex mutants (Tho$^-$) also are defective for adherence in the Σ1278b background. The strains shown are wild type (10560-23C), *flo11*Δ (L7558), *thp2*Δ (XY16), *mft1*Δ (XY118), *hpr1*Δ (XY189), and *tho2*Δ (XY191). (*Left*) A YPD plate after 3 days of incubation at 30°C. (*Right*) The same plate after wash. (*C*) Northern analysis shows a reduction of *FLO11* transcription in Tho$^-$ mutants. Lanes: 1, Σ1278b wild type (10560-23C); 2, *flo11*Δ (L7558); 3, *thp2*Δ (XY16); 4, *mft1*Δ (XY118). The blot was first hybridized with a *FLO11* probe (*Upper*) and then with an *SCR1* probe (*Lower*). (*D*) *GFP* transcription from the *FLO11* promoter in a P_{FLO11}*GFP* fusion is unaffected by Tho$^-$ mutants. The histogram compares *FLO11* (gray bars indicate strains as in *C*) and *GFP* mRNA levels [black bars indicate Tho$^+$ (L8225), *mft1*Δ (XY136)] by rtPCR. (*E*) GFP fluorescence indicates *FLO11* promoter functionality in Tho$^-$ mutants. Images show GFP fluorescence of exponentially growing Tho$^+$ (L8225) or *mft1*Δ (XY136) cells.

genes required for transcription regulation and initiation. Each mutant of the S288c deletion library was transformed with P_{TEF}*FLO11* and tested for agar adhesion. The screen identified <50 mutants with various extents of reduced adherence to solid agar. Among the mutants with the most nonadherent phenotypes and almost normal growth rates were the four single-gene deletions of the THO complex.

The nonadherent phenotype of the Tho$^-$ mutants is independent of the strain background and the promoter. Each member of the THO complex, *THP2*, *MFT1*, *HPR1*, and *THO2*, was separately deleted in a Σ1278b strain in which *FLO11* is under its native promoter at its resident site in the chromosome. Each of the four Tho$^-$ mutants also is strongly nonadherent in this background (Fig. 1*B*). Thus, our screen identified the THO complex as a novel promoter-independent regulator of *FLO11*.

Reduction of *FLO11* mRNA Levels in Tho$^-$ Mutants Requires the *FLO11* Coding Sequence. *FLO11* mRNA analysis by Northern blots as well as by real-time PCR (rtPCR) shows that *thp2*Δ and *mft1*Δ mutants have reduced levels of *FLO11* mRNA as compared with Tho$^+$ strains (Fig. 1 *C* and *D*). The reduction appears to be independent of the promoter sequence, because the *FLO11* levels are reduced both when *FLO11* is expressed from the *TEF1* promoter and from its native promoter. To determine whether the *FLO11* coding sequence was responsible, we analyzed the transcript levels in P_{FLO11}*GFP* strains in which *GFP* replaces the *FLO11* ORF. *GFP* mRNA and *FLO11* mRNA were compared by rtPCR in the corresponding Tho$^+$ and *mft1*Δ strains. The level of *FLO11* mRNA is reduced ≈85% in the *mft1*Δ mutant, whereas *GFP* expression from the *FLO11* promoter is nearly at wild-type levels in the Tho$^-$ mutant background (Fig. 1*D*). The lack of an effect of the Tho$^-$ mutants on P_{FLO11}*GFP* can also be visualized by the roughly equivalent GFP fluorescence in Tho$^+$ and *mft1*Δ strains (Fig. 1*E*). This result suggests that *FLO11* mRNA down-regulation in a Tho$^-$ mutant depends on the presence of the *FLO11* coding sequence.

***FLO11* Requires the THO Complex for Transcription Elongation Through the Repeats.** RNA polymerase (RNAP) II occupancy along the *FLO11* ORF was monitored in Tho$^+$ and Tho$^-$ strains by ChIP using an antibody to the Rpb3 subunit of the polymerase. The amount of *FLO11* DNA in the precipitate was assessed by PCR amplification. *FLO11* is an ORF of 4,104 nt, the middle third of which features 15 nearly perfect tandem repeats of 1,725 nt total length (28). We designed six primer pairs along *FLO11*: one in the promoter region, two in the 5′-end proximal region, two in the 3′end region, and one in the 3′ UTR (Fig. 2*A*). This ChIP analysis shows a gradual reduction in the level of RNAP II along *FLO11* in the *mft1*Δ *thp2*Δ mutant as compared with wild type (Fig. 2 *B* and *C*). The fact that the Tho$^-$ strain has comparable or slightly higher occupancy of RNAP II at the 5′ end of the *FLO11* ORF indicates that Tho$^-$ mutants do not reduce transcription initiation of *FLO11*. At the same time, reduced signal for the 3′ end probes of *FLO11* in the *mft1*Δ *thp2*Δ mutant suggests lower RNAP II occupancy along the *FLO11* ORF sequence. This result indicates that the THO complex is not involved in transcription initiation but rather in transcription elongation of *FLO11*.

To examine the role of the *FLO11* repeats on transcription, we constructed a *FLO11* allele that lacks the repeat-containing region (*flo11*::Δ*rep*) and compared the levels of *FLO11* transcription in Tho$^+$ and *mft1*Δ *thp2*Δ strains. *FLO11* expression is at least 65% reduced in *mft1*Δ *thp2*Δ compared with the wild-type strain, whereas *flo11*::Δ*rep* expression in the mutant strain is nearly the same as that in Tho$^+$ (Fig. 2*D*), suggesting that the repeat region in *FLO11* is the major obstacle to transcription elongation in the Tho$^-$ mutant background.

The obstacle to transcription caused by the repeats in a Tho$^-$ mutant has a profound consequence on the genetic stability of the repeats. The stability of the repetitive region was measured in a *FLO11*::*URA3* genomic construct that contains the *URA3* gene inserted among the *FLO11* repeats (Fig. 2*E*). Loss of the *URA3* gene is a direct measure of the gain or loss of integral numbers of repeats (28). In a Tho$^+$ strain, the repeats are relatively stable, being lost at $\approx 1.8 \times 10^{-5}$, whereas in the *thp2*Δ mutant the repeats are lost at 7.2×10^{-4} (Fig. 2*E*). The 40-fold higher frequency of segregants negative for uracil (Ura$^-$) in the Tho$^-$ strain compared with Tho$^+$ suggests a greater instability of the repeats region.

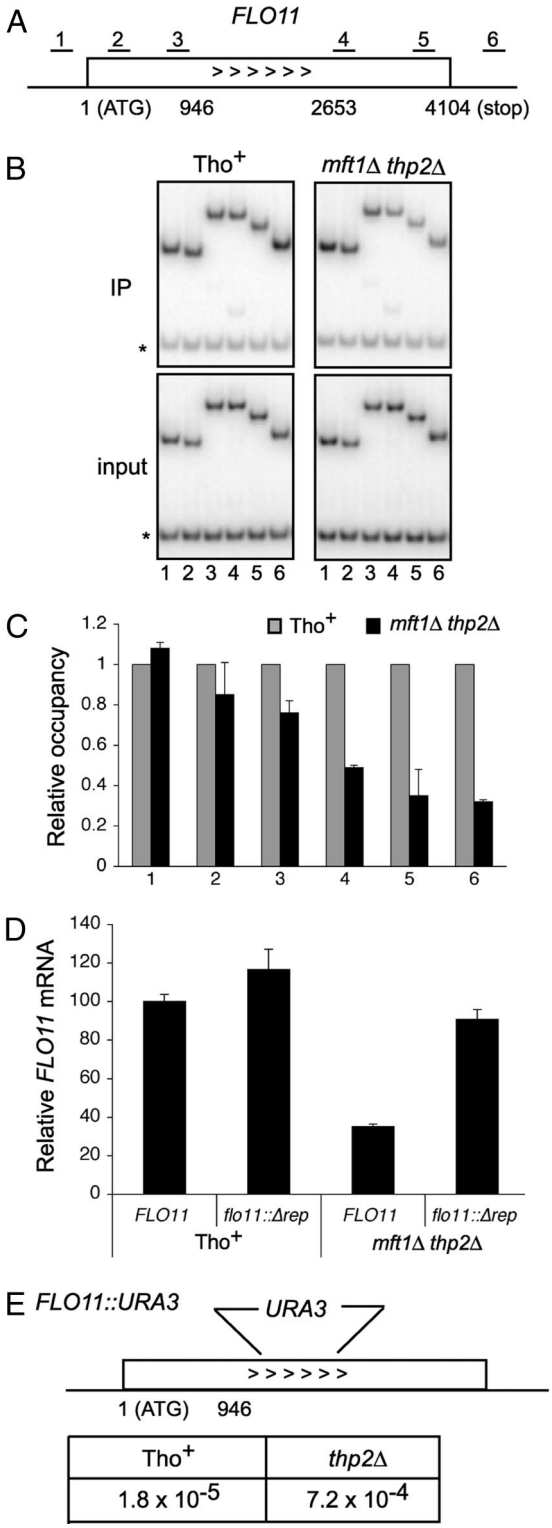

Fig. 2. Tho⁻ mutants show a transcription elongation defect through the repeats of *FLO11*. (*A*) *FLO11* probes. The relative position of six DNA fragments amplified for ChIP analysis is shown with respect to the start codon, stop codon, and the intragenic repeats of *FLO11*. Exact positions of the primers are given in Table 3. (*B*) Transcription elongation of *FLO11* is defective in Tho⁻ mutants. (*Upper*) RNAP II abundance along *FLO11* was monitored by anti-Rpb3 ChIP using the six primer pairs in *A*. (*Lower*) A 180-bp nontranscribed region on chromosome V. The strains shown are Tho$^+$ (L8046) and *mft1*Δ *thp2*Δ (XY269). (*C*) PhosphorImager quantitation of the elongation assay in *B* from two independent experiments. Each of the values for probes 1–6 in the Tho$^+$ strain was normalized to 1. The values for the double mutant were normalized to the corresponding wild-type probe. The raw occupancy values for the probes in the wild-type strain were typically between 3 and 8. (*D*) Removal of the repeats restores transcription of *FLO11* in a Tho⁻ mutant. Shown are the rtPCR results of the following strains: wild type (L8046), *flo11*::Δ*rep* (XY369), *mft1*Δ *thp2*Δ (XY269), and *mft1*Δ *thp2*Δ *flo11*::Δ*rep* (XY356). (*E*) *FLO11* repeats are less stable in a Tho⁻ mutant. The frequency of recombination was determined by measuring the frequency of Ura⁻ segregants obtained from a *FLO11*::*URA3* Tho$^+$ (XY266) or *thp2*Δ (XY454) strain in which *URA3* is flanked by *FLO11* repeats. Average measurements from four independent experiments are shown. Standard deviations were 0.6×10^{-5} for the Tho$^+$ strain and 8.7×10^{-5} for the *thp2*Δ mutant strain.

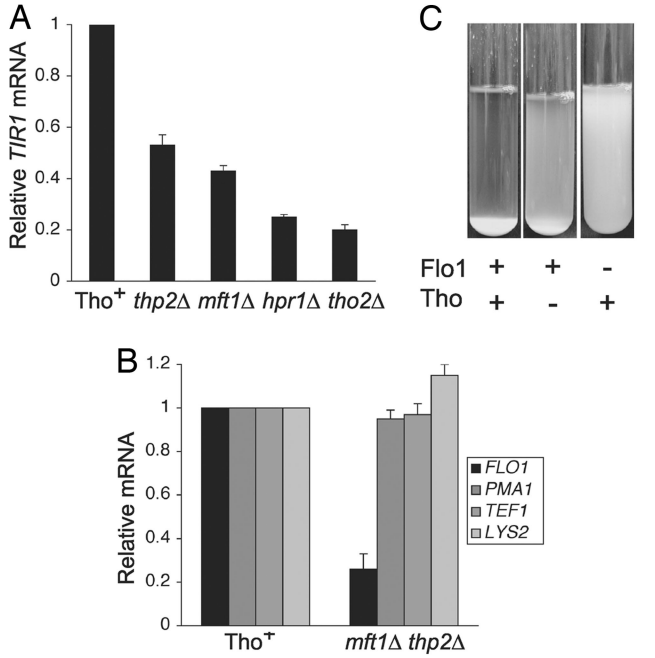

Fig. 3. Expression of other genes with repeats is reduced in Tho⁻ mutants. (*A*) *TIR1* requires the THO complex for transcription. Expression of *TIR1* was induced by a 90-min cold shock and measured by rtPCR. The strains shown are Tho$^+$ (10560–23C), *thp2*Δ (XY16), *mft1*Δ (XY118), *hpr1*Δ (XY189), and *tho2*Δ (XY191). (*B*) *FLO1* transcription is reduced in Tho⁻ mutants. The strains shown are Tho$^+$ (L8046) and *thp2*Δ *mft1*Δ (XY269). Expression of three other genes, *PMA1*, *TEF1*, and *LYS2*, was measured for comparison. (*C*) Reduced *FLO1* expression in Tho⁻ strains results in reduction of flocculence, a trait specified by *FLO1*. The strains shown are Tho$^+$ (L8046), *thp2*Δ *mft1*Δ (XY269), and, as the negative control, a strain that does not express *FLO1* (BY4741). Cells from an overnight culture were diluted to 0.1 OD and grown at 30°C for 24 h. The test tubes were vortex-mixed and immediately photographed.

Other Genes with Repeats Require the THO Complex for Efficient Transcription. *TIR1* is a cell wall gene encoded by 765 nt, 261 of which are internal tandem repeats. The gene is required for hypoxic growth and is induced by cold shock as well as by low oxygen levels (29). The level of *TIR1* mRNA was measured in a Tho$^+$ strain and in Tho⁻ mutants after a 90-min cold shock at 15°C. There is an ≈50% reduction of *TIR1* expression in *mft1*Δ and *thp2*Δ and >75% reduction in *hpr1*Δ and *tho2*Δ mutants (Fig. 3*A*). When grown hypoxically, *mft1*Δ and *thp2*Δ mutants show a modest growth defect, and *hpr1*Δ and *tho2*Δ show a strong growth defect (Fig. 6, which is published as supporting information on the PNAS web site), consistent with the reduced *TIR1* transcript levels.

The expression of *FLO1*, another gene with long tandem repeats, also is reduced in Tho⁻ mutants (Fig. 3*B*). Flo1 is required for flocculation between yeast cells, and the reduction of *FLO1* mRNA levels is reflected in the reduced flocculation

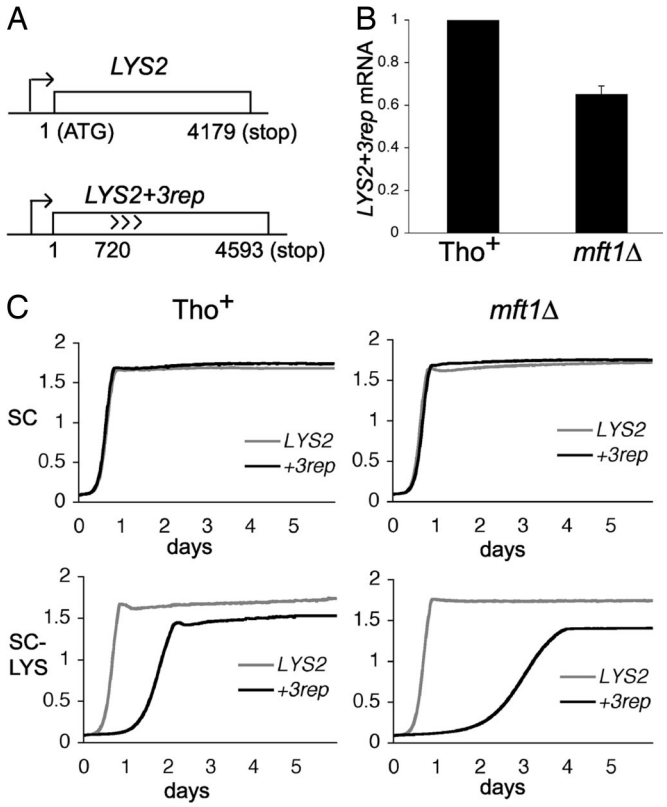

Fig. 4. *FLO1* repeats create THO dependence. (*A*) *LYS2*+*3rep* chimera. Three *FLO1* repeats (414 nt total) were inserted at position 720 of *LYS2*. (*B*) Insertion of *FLO1* repeats at *LYS2* leads to reduced expression in Tho$^-$ mutants. Tho$^+$ (XY299) and *mft1*Δ mutant strains (XY313) carrying the *LYS2*+*3rep* allele were grown in SC to an OD_{600} of 1 and then shifted to SC lacking Lys (SC-Lys) for 2 h. rtPCR data from two independent experiments is shown. (*C*) The growth defect of strains with a *LYS2*+*3rep* allele is greater in Tho$^-$ mutants. The strains shown are BY4741 (Tho$^+$ and *LYS2*), XY299 (Tho$^+$ and *LYS2*+*3rep*), *mft1*Δ (*LYS2*), and XY313 (*mft1*Δ and *LYS2*+*3rep*). The strains were grown overnight in SC 2% Glc, and diluted in a Bioscreen plate in SC 2% Glc or in SC-Lys/2% Glc in triplicate. The plate was incubated for 5.5 days with OD readings taken every 30 min.

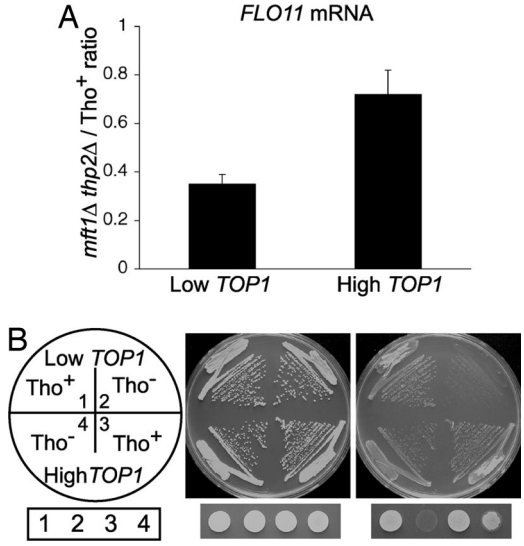

Fig. 5. *TOP1* overexpression partially restores *FLO11* mRNA levels and function. (*A*) Expression of *TOP1* from the strong *TEF* promoter restores *FLO11* expression in Tho$^-$ mutants. Low *TOP1* is the *FLO11* mRNA ratio in Tho$^-$/Tho$^+$ strains with *TOP1* under its own promoter (Tho$^+$, L8046; *mft1*Δ *thp2*Δ, XY269). High *TOP1* is the *FLO11* mRNA ratio in Tho$^-$/Tho$^+$ strains with *TOP1* under the *TEF* promoter (Tho$^+$, XY426; *mft1*Δ *thp2*Δ, XY427). (*B*) Overexpression of *TOP1* also partially restores adherence to agar of Tho$^-$ mutants. Strains: 1, Tho$^+$ (L8046); 2, *mft1*Δ *thp2*Δ (XY269); 3, Tho$^+$, high *TOP1* (XY426); 4, *mft1*Δ *thp2*Δ, high *TOP1* (XY427). The strains were streaked on a YPD plate, incubated for 2 days at 30°C, and photographed before and after the wash. Alternatively, 2×10^6 cells were spotted on a YPD plate, incubated for 1 day at 30°C, and washed.

of Tho$^-$ strains (Fig. 3*C*). There is no extensive sequence homology between the repeats in *FLO1* and those in *FLO11*. Several other genes with repeats (*FIT3* and *TIR4*; see Table 1, which is published as supporting information on the PNAS web site) show similar dependence on the THO complex.

For comparison, we also measured the mRNA levels of several ORFs without internal repeats of various lengths and expression levels: *PMA1* (2,757 nt) and *TEF1* (1,377 nt), which are highly expressed genes, and *LYS2* (4,179 nt), a gene expressed at lower levels. Expression of all three genes is unaffected in the *mft1*Δ *thp2*Δ mutant (Fig. 3*B*).

Intragenic Repeats Confer THO Dependence. An in-frame segment containing three *FLO1* repeats (a total of 414 nt) was inserted into the *LYS2* gene to test the effect of these repeats on transcription of that gene (Fig. 4*A*). *LYS2* is not affected by Tho$^-$ mutants when transcribed from its cognate promoter (Fig. 3*B*). *LYS2*+*3rep* expression is 35% less in the *mft1*Δ mutant than in the Tho$^+$ strain (Fig. 4*B*). This difference is reflected in the growth defect of the *LYS2*+*3rep mft1*Δ mutant compared with the *LYS2*+*3rep* Tho$^+$ strain in media that lacks lysine (Fig. 4*C*). These data suggest that *FLO1* repeats confer THO dependence.

Overexpression of *TOP1* Suppresses the Tho$^-$ Defect. The partial homology between Hpr1 and Top1 (23) and the lethality of *hpr1*Δ *top1*Δ or *mft1*Δ *top1*Δ double mutants suggested an overlap between topoisomerase and THO complex function. To test this possibility, we constructed a Tho$^-$ strain and a Tho$^+$ strain that contained the *TOP1* gene under the highly expressed *TEF* promoter and compared these strains with Tho$^-$ and Tho$^+$ strains without the overexpression construct. *TOP1* overexpression in Tho$^-$ mutants partially restores *FLO11* mRNA levels (Fig. 5*A*) as well as adherence to agar (Fig. 5*B*). Thus, Top1 partially complements THO complex function for the efficient transcription of genes with long internal repeats.

Discussion

The yeast genes affected by mutation of the THO complex have a number of similarities. The most salient attribute is that they are genes with many long tandem internal repeats. Genes with long internal tandem repeats are not a feature restricted to the yeast genome. It is estimated that 5% of human genes also have tandem repeats (30). As we showed previously (28), most of the yeast genes with internal repeats encode cell wall proteins, and the repeats are essential for cell surface interactions, such as adhesion. Here we show that alleles of genes with internal repeats require the THO complex for maximum expression and are genetically unstable in Tho$^-$ mutants. Our data showing that overexpression of topoisomerase I suppresses the Tho$^-$ defect in *FLO11* transcription further implicates the THO complex in DNA topology.

Although several previous studies using recombinant constructs have suggested that the THO complex was required either for genes of high GC content or for especially long genes, the yeast genes whose expression is dramatically affected do not have a high GC content [*FLO11* 46% (50% for the region of repeats)] (Table 1). The *FLO11* and *FLO1* genes are longer than the average yeast gene; however, transcription of yeast genes of equivalent size (*RPB1* and *LYS2*) is unaffected in Tho$^-$ mutants under standard growth conditions, and a third THO-dependent

gene, *TIR1*, is only 765 nt long. Moreover, in a Tho$^+$ strain there is little difference between the expression of the long (4.1 kb) or short (2.5 kb) form of the *FLO11* gene. However, efficient transcription of the wild-type *FLO11* gene containing the repeats depends on a functional THO complex, whereas a *FLO11* gene without the repeats (*flo11*::Δ*rep*) is expressed at the same level in both Tho$^+$ and Tho$^-$ strains.

The presumed importance of the THO complex for maintaining the topology of the DNA template contrasts with the failure of previous studies to identify phenotypic effects of Tho$^-$ mutants on native genes. In addition, we failed to detect any dramatic global change in the level of transcription for most genes as measured by whole genome microarrays in yeast. A similar analysis in *D. melanogaster* concluded that "the vast majority of genes are transcribed and exported independently of THO" (25). We posit that for most genes the activity of Top1 is sufficient to prevent the topological impediments to transcription elongation. However, for genes that have repeated obstructive sequences, such as the *FLO* genes, the stress on the system overwhelms the ability of Top1 to correct the defect. Under these conditions, the THO complex becomes essential.

This view raises the question of whether the THO complex is required only for efficient transcription of genes with long tandem repeats, which we think is unlikely. First, not all genes with tandem repeats show a phenotype in the Tho$^-$ strains (Table 1). Of course, many of these genes with repeats are expressed at extremely low levels and may, like *TIR1*, only require the THO complex upon induction or some environmental stress condition that requires enhanced transcription. Second, other genes whose transcription creates aberrant structures under stress conditions could also require the THO complex. For example, DNA:RNA hybrids, or R loops, have been detected during transcription in Tho$^-$ mutants (20), and increased levels of recombination have been associated with R-loop formation in Topo$^-$ (6, 31) as well as in splicing mutants (32). It is in this sense that we posit the THO complex as a protein complex whose function is to repattern the transcription complex, permitting efficient transcription elongation when transcription stalls.

Materials and Methods

Yeast Strains and Growth Conditions. Strains in two genetic backgrounds, S288c and Σ1278b, were used in these studies (Table 2, which is published as supporting information on the PNAS web site). The deletion library is in the S288c background, which has a mutation in the *flo8* gene (33). Because *FLO8* encodes a transcription factor required for *FLO11* expression, the screen of the library for mutations that caused the Flo$^-$ phenotype was performed with a P_{TEF}*FLO11* construct. This construct not only permits the screen of the S288c deletion library but also reports *FLO11* promoter-independent transacting mutations. Each of the Tho$^-$ mutants is a complete deletion of the respective THO gene. After the Tho$^-$ mutants were identified in the S288c screen, each was transformed into the Σ1278b 10560-23C strain and found to have a similar nonadherent phenotype.

For the yeast deletion library transformation, mutant strains in 96-well plates were preincubated with the *URA3*/*CEN* P_{TEF}*FLO11* plasmid B4126 and standard PEG/LiOAc/TE/ssDNA mixture (where TE is 10 mM Tris/1 mM EDTA, pH 7.5) for 3 h at 30°C, followed by a 45-min heat shock at 42°C. Transformants were grown on synthetic complete media (SC) lacking Ura, with the media first as a liquid (3 days) and then as a solid (2 days). A pool of transformants for each mutant was patched on a yeast extract/peptone/dextrose (YPD) rectangular plate and tested for adhesion after 1 day of growth at 30°C by a gentle wash under running water.

The S288c *FLO8*$^+$ strain L8046 was prepared by transforming a pRS305-based BglII-cut integrating plasmid that contains a Σ1278b copy of *FLO8* (B4241) into the S288c *flo8*$^-$ strain L4242. Strains with a *FLO11* allele that lacks the repeats region, *flo11*::Δ*rep*, were constructed in two steps. First, the *URA3* marker was amplified from a plasmid with primers V271 and V272 targeting the ends of the *FLO11* repeats region. Second, these *FLO11*::*URA3* strains were streaked on plates containing 5-fluoroorotic acid to loop out the *URA3* marker.

The *LYS2+3rep* strain that has three *FLO1* repeats inserted at position 720 nt of *LYS2* was prepared in the following way. A *FLO1rep–URA3–FLO1rep* cassette was amplified from the genomic DNA of strain KV133 (28) with primers K428 and K429 to create overhangs for in-frame integration at *LYS2* in the strain BY4741. Transformants that were Ura$^+$ and Lys$^-$ were then streaked on SC plus 5-fluoroorotic acid or SC-Lys plates to force *URA3* popouts, leaving behind *FLO1* repeats in *LYS2*. The *LYS2+3rep* chimera construct was confirmed by sequencing.

Strains were grown in YPD, unless selective media were required. Cold shock and anaerobic growth experiments were based on previously described protocols (29). For cold shock, cultures were grown at 30°C to OD_{600} 1.0 and then shifted to 15°C for 90 min; strains were grown hypoxically on YPD plates supplemented with 0.5% Tween 80 and 20 μg/ml ergosterol (Sigma, St. Louis, MO) and placed in a hypoxic chamber with an AnaeroPack sachet (Mitsubishi Gas Chemical America, New York, NY) for 3 days at 30°C. A Bioscreen apparatus (Labsystems, Chicago, IL) was used for the growth comparison of *LYS2+3rep* strains. Several reagents were used for selection or counterselection during the preparation of strains: 0.2 mg/ml geneticin (GIBCO, Carlsbad, CA), 0.3 mg/ml hygromycin (Sigma), 0.1 mg/ml nourseothricin (Werner BioAgents, Jena, Germany), and 1 mg/ml 5-fluoroorotic acid (USBiological, Swampscott, MA).

The frequency of Ura$^-$ segregants of *FLO11*::*URA3* Tho$^+$ (XY266) or *thp2*Δ (XY454) strains was determined after growth on YPD plates for 1 day at 30°C, followed by plating on SC plus 5-fluoroorotic acid to count colony-forming units.

Primers and Plasmid Construction. Primers are listed in Table 3, which is published as supporting information on the PNAS web site. Primer pairs for rtPCR analysis were designed with Primer Express software. The primer pairs along *FLO11* for ChIP analysis were designed to yield products of 250–300 bp. Primers for amplification of an untranscribed region on chromosome V were as previously described (34). The plasmid B4126 was constructed by transferring a StuI/AgeI fragment that contains *FLO11* from B4050 (35) into a p416TEF CEN plasmid linearized with EcoRI and XhoI.

mRNA Analysis. Total RNA was isolated from 10-ml cultures grown to an OD_{600} of 1.0 by using hot acid phenol. DNaseI treatment was carried out for 30 min (Epicentre Biotechnologies, Madison, WI). Reverse transcription of 0.3 μg of RNA was performed for 30 min at 48°C with 12.5 units of MultiScribe reverse transcriptase (Applied Biosystems, Framingham, MA) and 2.5 μM random hexamers. One-seventh of the cDNA product was used for rtPCR analysis with reagents from Applied Biosystems and the ABI 7500 rtPCR system. Probes at the 3′ end of ORFs were used when available. Normalization was to *ACT1*, except when analyzing Σ1278b Tho$^-$ mutants, where we noticed a slight up-regulation of *ACT1* in Tho$^-$ mutants compared with other controls. In those cases, normalization was to *SCR1*, a gene transcribed by RNAP III. The histograms present data from two to four independent experiments.

Northern hybridization was performed on 10-μg RNA samples after gel electrophoresis. The blots were first hybridized with a *FLO11* probe and then with an *SCR1* probe.

ChIP. ChIP were performed as previously described (34). Briefly, cells were grown to an OD_{600} of 0.8–1.0, fixed with formaldehyde, lysed, and sonicated. The lysates were immunoprecipitated with an

anti-Rpb3 antibody (NeoClone, Madison, WI) bound to Protein G Sepharose beads (Amersham Biosciences, Piscataway, NJ). Overnight incubation at 4°C was followed by four washes. The protein/DNA complexes were eluted, and the cross-links were reversed with pronase (Calbiochem, San Diego, CA). DNA was analyzed by concurrent PCR of a *FLO11* region and an untranscribed region on chromosome V. All samples were resolved on a 6% polyacrylamide gel, and the signals were quantitated by a PhosphorImager and ImageQuant software. Occupancy value for each of six regions along *FLO11* was calculated as a ratio (immunoprecipitation sample/input sample) of ratios (*FLO11* specific signal/untranscribed region signal).

The ChIP assays were performed both on strains in the S288c and Σ1278b backgrounds. Although there were quantitative differences in the relative enrichment of both backgrounds, the polymerase occupancy in the Tho$^-$ strains was reduced in the 3′ end of the *FLO11* strain. Better enrichment of the specific signal in the immunoprecipitation sample was observed for S288c than for Σ1278b strains.

Bioinformatics. The GC content of DNA sequences was determined with EMBOSS GEECEE software.

We thank Phil Sharp, Rob Wheeler, and Claire Tseng for critical reading of the manuscript; members of the G.R.F. laboratory for discussions and reagents; Milan de Vries for the large-scale transformation protocol; the laboratory of Susan Lindquist (Whitehead Institute for Biomedical Research/Massachusetts Institute of Technology) for use of the Bioscreen machine; George Bell and Kim Walker for help with bioinformatics; and Tom DiCesare for help with graphics. K.J.V. acknowledges financial support from the Harvard University Bauer Center, the Belgian American Educational Foundation, and from the Fund for Scientific Research–Flanders (Belgium). This work was supported by National Institutes of Health Grants GM56663 (to S.B.) and GM35010 (to G.R.F.). G.R.F. is an American Cancer Society Professor.

1. Lee TI, Young RA (2000) *Annu Rev Genet* 34:77–137.
2. Maniatis T, Reed R (2002) *Nature* 416:499–506.
3. Arndt KM, Kane CM (2003) *Trends Genet* 19:543–550.
4. Liu LF, Wang JC (1987) *Proc Natl Acad Sci USA* 84:7024–7027.
5. Brill SJ, Sternglanz R (1988) *Cell* 54:403–411.
6. Drolet M (2006) *Mol Microbiol* 59:723–730.
7. Christman MF, Dietrich FS, Fink GR (1988) *Cell* 55:413–425.
8. Kim RA, Wang JC (1989) *Cell* 57:975–985.
9. Fan HY, Cheng KK, Klein HL (1996) *Genetics* 142:749–759.
10. Prado F, Piruat JI, Aguilera A (1997) *EMBO J* 16:2826–2835.
11. Fan HY, Merker RJ, Klein HL (2001) *Mol Cell Biol* 21:5459–5470.
12. Jimeno S, Rondon AG, Luna R, Aguilera A (2002) *EMBO J* 21:3526–3535.
13. Chavez S, Beilharz T, Rondon A, Erdjument-Bromage H, Tempst P, Svejstrup J, Lithgow T, Aguilera A (2000) *EMBO J* 19:5824–5834.
14. Reed R, Cheng H (2005) *Curr Opin Cell Biol* 17:269–273.
15. Zenklusen D, Vinciguerra P, Wyss JC, Stutz F (2002) *Mol Cell Biol* 22:8241–8253.
16. Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, Hurt E (2002) *Nature* 417:304–308.
17. Abruzzi K, Lacadie S, Rosbash M (2004) *EMBO J* 23:2620–2631.
18. Kim M, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S (2004) *EMBO J* 23:354–364.
19. Chavez S, Aguilera A (1997) *Genes Dev* 11:3459–3470.
20. Huertas P, Aguilera A (2003) *Mol Cell* 12:711–721.
21. Chavez S, Garcia-Rubio M, Prado F, Aguilera A (2001) *Mol Cell Biol* 21:7054–7064.
22. Jensen T, Boulay J, Olesen J, Colin J, Weyler M, Libri D (2004) *Mol Cell* 16:235–244.
23. Aguilera A, Klein H (1990) *Mol Cell Biol* 10:1439–1451.
24. Sadoff B, Heath-Pagliuso S, Castano I, Zhu Y, Kieff F, Christman M (1995) *Genetics* 141:465–479.
25. Rehwinkel J, Herold A, Gari K, Kocher T, Rode M, Ciccarelli F, Wilm M, Izaurralde E (2004) *Nat Struct Mol Biol* 11:558–566.
26. Roberts RL, Fink GR (1994) *Genes Dev* 8:2974–2985.
27. Lambrechts MG, Bauer FF, Marmur J, Pretorius IS (1996) *Proc Natl Acad Sci USA* 93:8419–8424.
28. Verstrepen K, Jansen A, Lewitter F, Fink G (2005) *Nat Genet* 37:986–990.
29. Abramova N, Sertil O, Mehta S, Lowry C (2001) *J Bacteriol* 183:2881–2887.
30. O'Dushlaine CT, Edwards RJ, Park SD, Shields DC (2005) *Genome Biol* 6:R69.
31. Masse E, Drolet M (1999) *J Biol Chem* 274:16659–16664.
32. Li X, Manley J (2005) *Cell* 122:365–378.
33. Liu H, Styles C, Fink G (1996) *Genetics* 144:967–978.
34. Keogh M, Buratowski S (2004) *Methods Mol Biol* 2557:1–16.
35. Guo B, Styles C, Feng Q, Fink G (2000) *Proc Natl Acad Sci USA* 97:12158–12163.

Voynov *et al.* 10.1073/pnas.0606546103.

Supporting Information

Files in this Data Supplement:

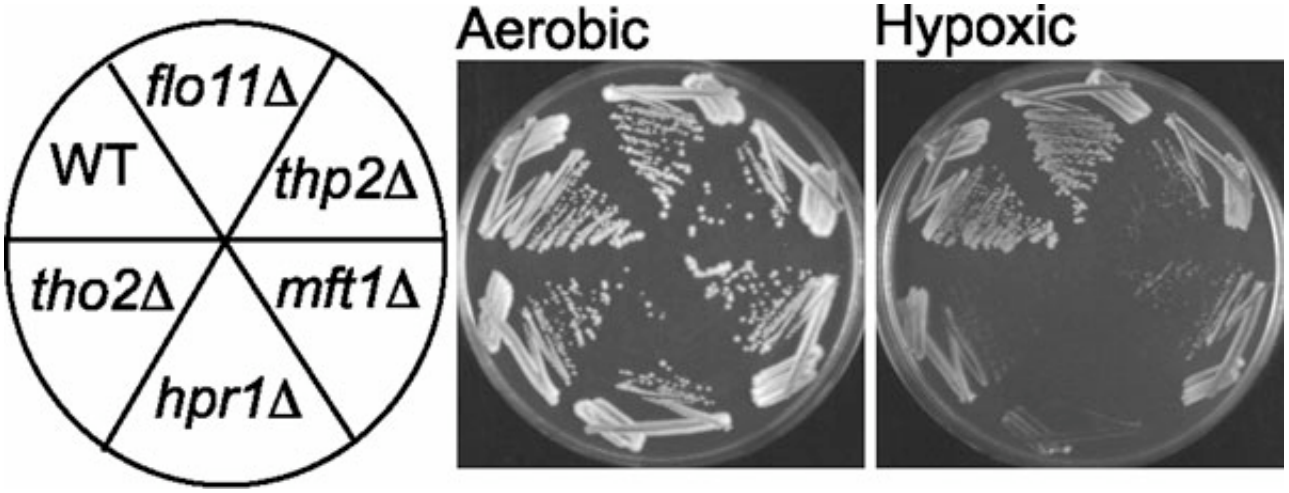

Fig. 6. Tho$^-$ mutants show a hypoxic growth defect. The strains shown are wild type (10560-23C), *flo11D* (L7558), *thp2D* (XY16), *mft1D* (XY118), *hpr1D* (XY189), and *tho2D* (XY191).

The strains were streaked on yeast extract/peptone/dextrose plates supplemented with 0.5% Tween 80 and 20 mg/ml ergosterol, and grown at 30°C aerobically or in a hypoxic chamber for 3 days.

Table 1. Gene expression in Tho$^-$ mutants

ORF	Tho$^-$/Tho$^+$	Length, nt	GC, %	Repeats length (position)	Repeats GC, %
FLO11	0.15-0.35	4,104	0.46	1725 (937-2,661)	0.50
FLO1	0.22-0.30	4,614	0.45	2391 (841-3,231)	0.47
TIR1	0.21-0.51	765	0.47	261 (343-603)	0.51
FIT3	0.31-0.47	615	0.51	306 (76-381)	0.52
TIR4	0.47-0.80	1,464	0.47	399 (373-771)	0.50
PIR1	0.96-1.02	1,026	0.46	375 (208-582)	0.48
HSP150	0.91-1.11	1,164	0.49	399 (286-684)	0.50
RPB1	1.01-1.28	5,202	0.41	486 (4,651-5,136)	0.50
PMA1	0.91-0.96	2,757	0.42	No repeats	N/A
TEF1	0.92-0.96	1,377	0.44	No repeats	N/A
LYS2	1.12-1.18	4,179	0.40	No repeats	N/A
lacZ	0.06-0.22*	3,075	0.56*	No repeats	N/A

Cumulative real-time PCR data is presented as a ratio of expression in Tho$^-$ mutants to expression in Tho$^+$ strains. Length of the full ORF and of the repeats region if applicable, as well as GC content of each, also are shown for comparison.

**lacZ* data is as in Chavez S, Beilharz T, Rondon A, Erdjument-Bromage H, Tempst P,

Svejstrup J, Lithgow T, Aguilera A (2000) *EMBO J* 19:5824-5834.

Table 2. Yeast strains

Strain	Genotype	Source
BY4741	S288c *MATa his3^1 leu2^0 met15^0 ura3^0*	
orf?	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 orf::KanMX4*	Invitrogen deletion library
mft1?	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 mft1::KanMX4*	Invitrogen *mft1*Δ
thp2?	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 thp2::KanMX4*	Invitrogen *thp2*Δ
10560-23C	☒1278b *MAT*ᵹ *ura3-52 his3::hisG leu2::hisG*	Fink laboratory collection
L7558	☒1278b *MAT*ᵹ *ura3-52 his3::hisG leu2::hisG flo11::kanMX4*	Fink laboratory collection
XY16	☒1278b *MAT*ᵹ *ura3-52 his3::hisG leu2::hisG thp2::kanMX4*	This study
XY118	☒1278b *MAT*ᵹ *ura3-52 his3::hisG leu2::hisG mft1::kanMX4*	This study
XY189	☒1278b *MAT*ᵹ *ura3-52 his3::hisG leu2::hisG hpr1::kanMX4*	This study
XY191	☒1278b *MAT*ᵹ *ura3-52 his3::hisG leu2::hisG tho2::kanMX4*	This study
L8255	☒1278b *MAT*ᵹ *ura3-52 trp1::hisG his3::hisG leu2::hisG flo11::yEGFP-URA3*	Fink laboratory collection
XY136	☒1278b *MAT*ᵹ *ura3-52 trp1::hisG his3::hisG*	This study

	leu2::hisG flo11::yEGFP-URA3 mft1::kanMX4	
L8046	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52*	Fink laboratory collection
XY266	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 FLO11::URA3*	This study
XY269	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4 mft1::HygR*	This study
XY356	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4 mft1::HygR flo11::Δrep*	This study
XY369	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 flo11::Δrep*	This study
XY299	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 LYS2+3rep*	This study
XY313	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 LYS2+3rep mft1::KanMX4*	This study
XY426	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 NatNT2-*P_{TEF}*TOP1*	This study
XY427	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4 mft1::HygR NatNT2-*P_{TEF}*TOP1*	This study
XY454	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4 FLO11::URA3*	This study

Table 3. Primers used in this study

ORF, position	Name	Sequence
ChIP		

FLO11 5' f -296	V241	TGTCTTATCTGAGGAATGTCCGTG
FLO11 5' r -44	V242	ATTAGAACCACAACATGACGAGGG
FLO11 f +93	V243	CTCCGAAGGAACTAGCTGTAATTCT
FLO11 r +345	V244	TTCGTTGTAACCGTATAGTTGGACG
FLO11 f +603	V247	GACAATAATTGTGGCGGTACGAAG
FLO11 r +918	V248	AGTGCATGTCTTAGATGTGGTAGT
FLO11 f +2752	V249	ACCGAAACTACCATTGTTCCAACT
FLO11 r +3058	V250	GTTTCGCTTGGACTGGTTGAACAT
FLO11 f +3737	V253	AGTCATCTGTTGGTACTAACTCCG
FLO11 r +4037	V254	CCTTGGTAAGTACTCGAGATAGAAGG
FLO11 3' f +258	V255	ACAAGTACCGGTAGTATTGGCAC
FLO11 3' r +523	V256	CCTCTTATTCATCAAAGCCTGGTC

Chr V no-ORF f	contr f	GGCTGTCAGAATATGGGGCCGTAGTA
Chr V no-ORF r	contr r	CACCCCGAAGCTGCTTTCACAATAC
Real-time PCR		
SCR1 f 384	A47	CGGCCGGGATAGCACATA
SCR1 r 437	A48	CGCCGAAGCGATCAACTT
TEF f 349	V216	GCTGGTGGTGTCGGTGAATT
TEF r 410	V217	GCGTGTTCTCTGGTTTGACCAT
FLO11 f 3038	V290	GTTCAACCAGTCCAAGCGAAA
FLO11 r 3104	V291	GTAGTTACAGGTGTGGTAGGTGAAGTG
GFP f 24	V292	CACTGGTGTTGTCCCAATTTTG
GFP r 94	V293	CACCGGAGACAGAAAATTTGTG
FLO1 f 4268	V298	TAGCTGCTGAGACGATTACCAA
FLO1 r 4348	V299	GCGTGATTAGATCTTGAAAGCGAA

LYS2 f 2081	V308	GTTGCGGTAGGGCTGATGA
LYS2 r 2150	V309	TGCGTATCTATTTCTCCTAATTCGATT
PMA1 f 2451	V363	GTCTTCTATCCCATCCTGGCAAT
PMA1 r 2532	V364	CCAACCGAATAAGGTAAACATGGT
RPB1 f 4268	V369	GTGTTTCGGAAAATGTCATTCTTG
RPB1 r 4346	V370	GACTCCTCATCGATCATCACATCA
HSP150 f 99	V379	CACTTACAGCGGTGGTGTTACC
HSP150 r 162	V380	GGAGATTGGTTGAACGGCAAT
FIT3 f 409	V387	AACACCTGGTCTCCAAGTAGTACTTCT
FIT3 r 481	V388	TGGTTGCAGTGGTTGAAGCA
TIR1 f 624	V401	CAAGACCTCTGCCATCTCTCAA
TIR1 r 690	V402	TTGCTCAGAAACAGCCTTGGT
PIR1 f 844	V403	CCACAAGCTGGTGCTATCTATGC

PIR1 r 910	V404	CACCGATGGCCAAGTTACCT
TIR4 f 1125	V451	TGACAACACTCTTGTCACCAAAGA
TIR4 r 1188	V452	TGGAACGGCAGATGTTGAAG
ACT1 f 1038	V453	GGCTTCTTTGACTACCTTCCAACA
ACT1 r 1103	V454	GATGGACCACTTTCGTCGTATTC
Strain constr.		
flo11::Δrep f	V271	TGCACTAAGAAGACTACTACTCCAGTACCAACCCCATCA
		AGCTCTACTACTGAAAGTTCTTGATTCGGTAATCTCCGA
flo11::Δrep r	V272	GAACAGAAGAGCTTTCAGTGCTAGAGCTGAATGGGGTT
		GAAGATGGAGCGGGTAATAACTGATATAA
LYS2+rep f	K428	GATAGTTTACCTGATCCAACTAAGAACTTGGGCTGGTG
		CGATTTCGTGGGGTGTATTCACCTAAGTCAATCTAACT
		GTACTGTCCCTGA
LYS2+rep r	K429	TGGAGTCTCCACAACACAGGTTCTCTCTGGGAAGGCTT

		CAGCATTGTCCTGGAAAATGTC-GATAGAGCTGGTGAT
		TTGTCCTGAA