Molecular Cell, Volume *48*

Supplemental Information

Genome-wide Screen Identifies Pathways that Govern GAA/TTC Repeat Fragility and Expansions in Dividing and Nondividing Yeast Cells

Yu Zhang, Alexander A. Shishkin, Yuri Nishida, Dana Marcinkowski-Desmond, Natalie Saini, Kirill V. Volkov, Sergei M. Mirkin, and Kirill S. Lobachev

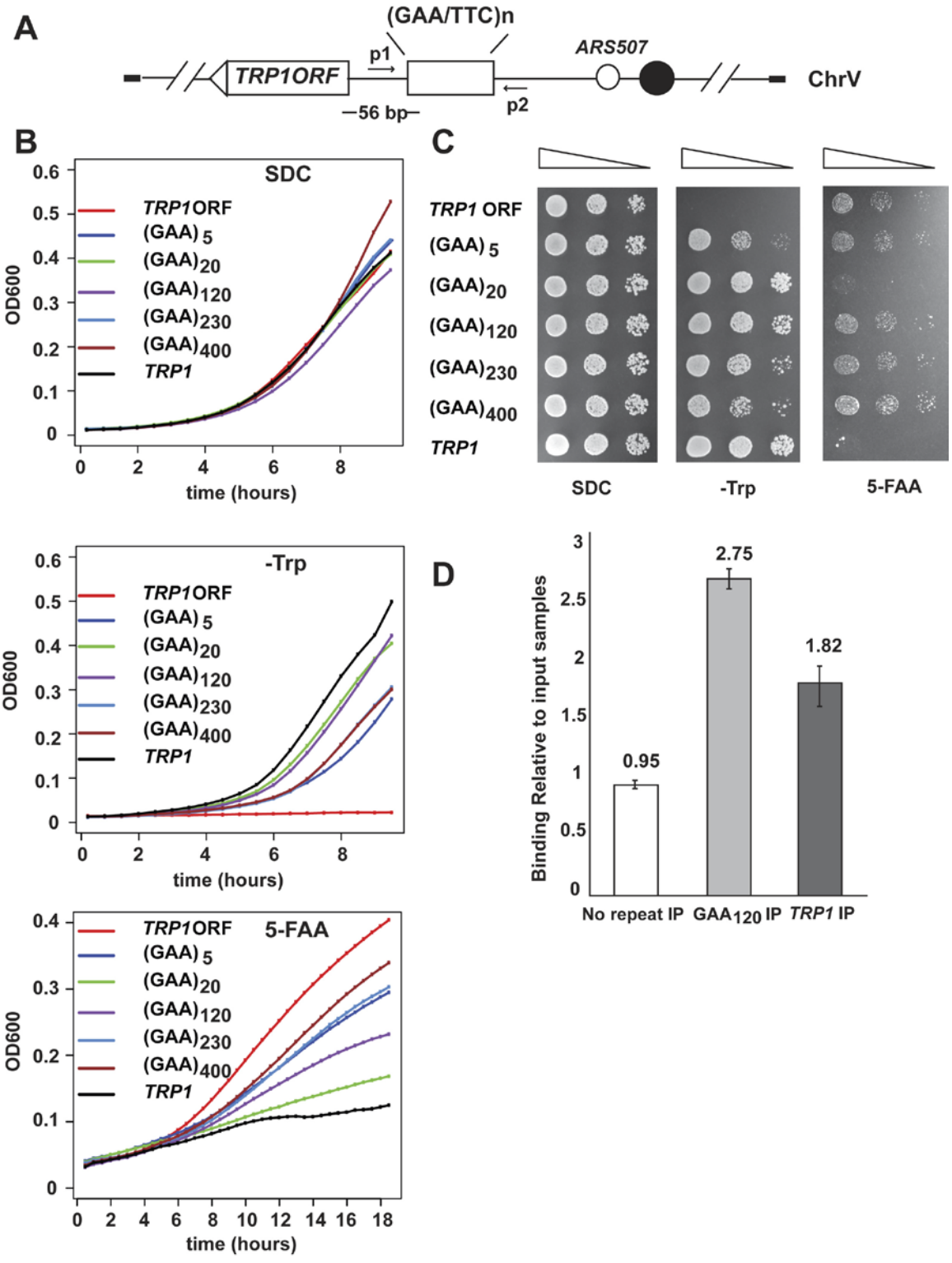
A
(GAA/TTC)n
p1
ARS507
TRP1ORF
ChrV
−56 bp−
p2
B
SDC
TRP1ORF
(GAA)5
(GAA)20
(GAA)120
(GAA)230
(GAA)400
TRP1
OD600
time (hours)
-Trp
5-FAA
C
TRP1 ORF
SDC
-Trp
5-FAA
D
Binding Relative to input samples
2.75
1.82
0.95
No repeat IP
GAA120 IP
TRP1 IP

Figure S1. GAA/TTC Repeats Can Recruit Transcription Initiation Factors and Drive Gene Expression, Related to Figure 3

(A) Diagram depicting the position of the repeats relative to *TRP1* ORF. Primers used for real-time qPCR are shown as black arrows.

(B) Growth dynamics of GAA-containing and control strains in synthetic complete media (SDC), tryptophan drop-out media (-Trp) and on 5-FAA-containing media (5-FAA).

(C) GAA/TTC repeats promote *TRP1* ORF expression. Strains contain 0, 5, 20, 120, 230, 400 GAA/TTC repeats in front of *TRP1* ORF. Strains with *TRP1* ORF expressed from its natural promoter are used as a positive control. 10-fold serial dilutions of yeast cells were plated on SDC, -Trp or 5-FAA medium.

(D) GAA/TTC repeats recruit the transcription initiation factor Sua7. ChIP was carried out using TAP-tagged Sua7 in strains harboring no repeats, 120 GAA/TTC repeats, or the natural *TRP1* promoter. Precipitated DNA was quantified by real-time PCR using primers for the 5'end of *TRP1*. PCR amplifying the 3' region of *POL5* was used for data normalization. The graph shows relative enrichment of Sua7 in antibody treated samples compared with untreated samples. Experiments were done in triplicates; error bars represent the standard error of the mean.

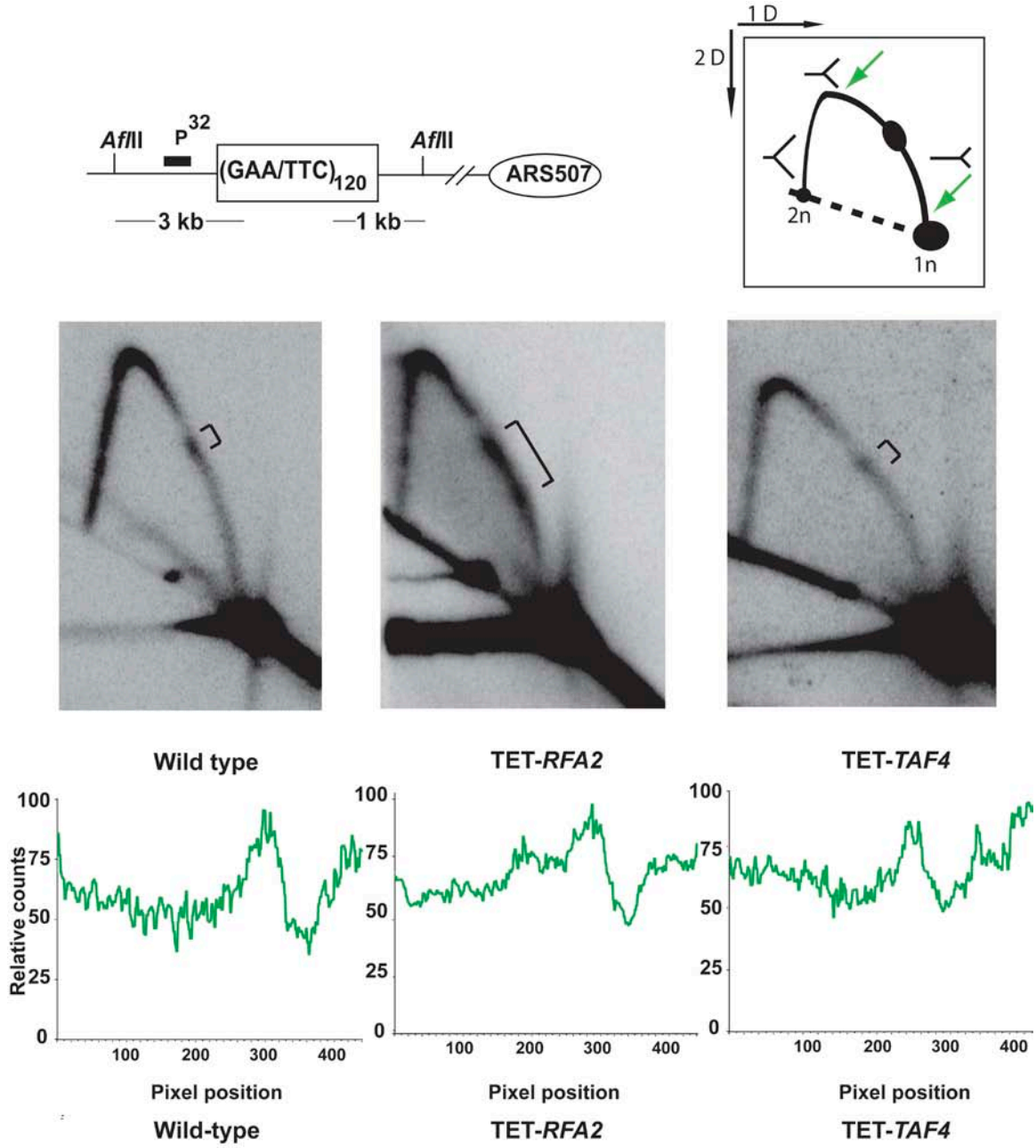

Figure S2. 2D Analysis of Replication Intermediates in Wild-Type, TET-*RFA2*, and TET-*TAF4* Strains, Related to Figure 5

The upper panel shows relative position of repeats, restriction sites and *ARS507* (not to scale). Replication intermediates are highlighted by *LYS2* specific probe (black solid

rectangle). Replication arrest at GAA/TTC tract is shown by bracket. Densitometry analysis of the Y-arc's long arm (marked by green arrows) is shown. Relative intensities of the arc are plotted against the distance from the monomers. The peaks depict the zones of replication arrests across the repeats. The TET-*RFA2* strain shows extended replication arrest in comparison with the wild-type and TET-*TAF4* strains.

Table S1. Mutants identified in the genome-wide screen, related to Figure 1, Table 1.

Genetic background	Essential genes		Non-essential genes	Mutants recreated
	yTHC	DAmP	YKO	
DNA repair genes				
mre11			+[a]	+[b]
rad50			+	
xrs2			+	
sae2			+	+
tsa1			+	
Replication genes				
rfa2	+	+		+
pol12	+			+
pri2	+			+
pol1	+	+		
pol2				+
pol3	+			+
pol30	+	+		+
rfc2	+			+
rfc3	+			
rfc4	+			
rfc5	+			
dna2	+			+
mcm4	+			+
mcm5	+			
mcm7	+			
orc2	+			
orc4	+			+
rad27			+	+
rtt101			+	+
mms1			+	+
tof1			+	+
mrc1			+	+
csm3			+	
Telomere protection genes				
ten1		+		+
cdc13				+
Transcription initiation genes				
taf4	+			+
taf11	+			
taf12	+			
taf9				+
toa1				+
sua7				+
tfg1				+
spn1	+			+

[a] + indicates that effect of the mutant allele on GCRs was identified using this library
[b] + indicates mutant alleles that were used to calculate fragility rates and frequencies

Table S2. Complementation of hyper-GCR phenotype of TET-*TAF4* and TET-*POL3* by centromeric vectors carrying *TAF4* and *POL3* expressed from native promoters, related to Table 1.

Genetic background	Rate of GCRs ($X10^{-7}$)		
	YCp50[a]	*TAF4* plasmid	*POL3* plasmid
$(GAA)_{230}$	27 (18-41)[b]	33 (19-50)	30 (26-80)
TET-*TAF4* with $(GAA)_{230}$	3500 (2800-4800)	22 (18-33)	ND
TET-*POL3* with $(GAA)_{230}$	150 (110-250)	ND[c]	27 (23-55)

[a] YCp50, centromeric vector carrying *URA3* gene, was used as a control.

[b] Numbers in parentheses correspond to the 95% confidence interval

[c] Not determined

Table S3. Rates of GAA/TTC-induced GCRs with and without down-regulation of TET-*RFA2* and TET-*TAF4* using doxycycline, related to Table 1.

Genetic background	Rate of GCRs ($X10^{-7}$)	
	Without down-regulation	After down-regulation[a]
TET-*RFA2* $(GAA)_{120}$	270 (93-380)[b]	510 (210-1300)
TET-*TAF4* $(GAA)_{120}$	130 (110-170)	66 (47-110)

[a] 0.1 μg/ml doxycycline was used for down-regulation

[b] Numbers in parentheses correspond to the 95% confidence interval

Table S4. Frequency of fragility at $(GAA)_{120}$ in wild-type and TET-*TAF4* strains in dividing and non-dividing cells, related to Figure 3.

Time after inoculation (hours)	Fragility frequency ($x10^{-7}$)	
	Wild-type	TET-*TAF4*
37	5 (4-5)[a]	29 (12-38)
40	4 (2-9)	45 (16-110)
43	7 (2-10)	47 (43-160)
46	10 (3-14)	100 (53-150)
49	12 (6-18)	170 (100-300)
52	15 (10-23)	350 (140-440)
55	22 (14-24)	630 (440-770)
58	23 (17-37)	460 (290-520)
61	34 (31-46)	520 (360-840)
64	45 (30-54)	740 (610-890)
68	51 (34-66)	980 (720-1400)
82	100 (73-140)	1500 (900-2500)
93	150 (100-240)	2300 (1500-3500)
115	350 (160-480)	1900 (1600-3000)

[a] Numbers in parentheses correspond to the 95% confidence interval

Table S5. Expansion frequencies of $(GAA)_{100}$ in dividing and non-dividing cells, related to Figure 4.

Time after inoculation (hours)	Expansion frequency ($x10^{-5}$)	
	Wild-type	TET-*TAF4*
37	57 (39-76)[a]	210 (130-550)
43	35 (27-54)	700 (340-1400)
49	18 (11-23)	700 (600-1600)
55	12 (9-21)	860 (510-1000)
68	14 (6-21)	540 (380-830)
82	12 (8-23)	400 (320-520)
115	12 (10-19)	580 (350-800)

[a] Numbers in parentheses correspond to the 95% confidence interval

Supplemental Experimental Procedures

Yeast Strains

All strains used in this study are derivatives of BY4742 (Open Biosystems). The genotype of the query strain, HMK246, used in the screen is: MATα, *ura3*-Δ, *leu2*-Δ, *his3*-Δ, *lys2*-Δ, *rpl28-Q38K, mfa1Δ::MFA1pr-HIS3, V34205::lys2::*$(GAA)_{230}$, *V29617::hphMX*. The strain was constructed in several steps. First, in BY4742 *LYS2* was inserted in the left arm of chromosome V telomere distal to *CAN1*. Second, the mating type of this strain was switched to MATa using pJH132 plasmid (gift from Dr. James Haber). The *MFA1pr-HIS3* cassette was PCR amplified from y2454 (Smith et al., 2004) and inserted into the *MFA1* locus. The mating type was switched back to MATα. Third, the *hphMX* cassette was inserted telomere proximal from *CAN1* on chromosome V. Fourth, the *rpl28- Q38K* mutation was obtained by selecting resistant colonies on YPD media supplemented with 5 mg/L cycloheximide. Finally, the $(GAA)_{230}$ tract was moved to the *Bam*HI site of *LYS2* via *delitto perfetto* (Storici et al., 2001).

The three collections of tester strains: yTHC, DAmP, YKO strains, were purchased from Open Biosystems. The strains where fragility was assessed have the following genotype: MATa, *bar1*-Δ, *trp1*-Δ, *his3*-Δ, *ura3*-Δ, *leu2*-Δ, *ade2*-Δ, *lys2*-Δ, *met15*-Δ, *V34205*::*ADE2*, *lys2*::(GAA)n. *LYS2* contains either 120 or 230 of GAA/TTC repeats. As a control, the strain with no repeats was used. To verify the effect of mutants uncovered from the screen, TET-alleles were created by replacing the natural promoters of the essential genes with tetracycline repressible promoters (Belli et al., 1998). pCM225 (Euroscarf) was used as a template for PCR amplification to generate a replacement cassette containg *kanMX*-TETp flanked by 50 bp homology to the region of

integration. Non-essential genes were knocked out by the *kanMX* or *hphMX* cassettes using one-step integration.

For testing GAA repeats' ability to serve as promoter elements, *TRP1* ORF amplified from pFL35 (Bonneaud et al., 1991) was used to replace the promoter part and 3246 bps of *LYS2* 5' region using *delitto perfetto* technique. As a result *TRP1* ORF was positioned 56 bps away from 0, 5, 20, 120 or 230 repeats where GAA strand serves as the sense strand for transcription. The isolate with 400 GAA/TTC repeats in front of *TRP1* ORF was the result of spontaneous expansion of 230 repeats. The size of the expanded repeats is approximate and was estimated by PCR amplification of the region and comparison against a known DNA ladder. *TRP1* ORF was also brought adjacent to 20, and 120 copies of repeats with TTC strand serving as the sense strand for transcription. As a positive control, *TRP1* with the natural promoter was used to replace the promoter and 5' part of *LYS2* in the strain carrying no repeats. After integration, *TRP1* ORF and *TRP1* were sequenced to confirm that no mutations were introduced into *TRP1* during transformation.

The strains for studying large-scale repeat expansions were derived from the strains described in Shishkin et al., 2009. Importantly, it was modified from the originally published cassette by increasing the length of the intron by 269 bps, making its overall length 974 bps. As a result, even relatively small expansions would drive the overall length of the intron above the splicing threshold of ~1.1 kb. The adjusted selectable cassette was created as follows. First, a DNA segment containing the multiple cloning site of pYES3 (Shishkin et al., 2009) was removed by the digestion with *Pdi*I and *Pme*I followed by the vector's re-ligation. Second, a plasmid pTRP1-ISR was generated

by cloning a downstream integration sequence, corresponding to the positions 75641-75895 of chromosome III (*Saccharomyces* genome database) into the *Cla*I site of the modified pYES3. Third, the plasmid pISL-TRP1-ISR was generated by cloning an upstream integration sequence, corresponding to the positions 75227-75594 of chromosome III into the *Zra*I site of the pTRP1-ISR. Fourth, an artificially split *URA3* cassette was amplified from the pYES-Int (Shishkin et al., 2009) and cloned between the *Bgl*II and *Sal*I sites of pISL-TRP1-ISR. (GAA)100 repeat was then cloned between the *Nco*I and *Cla*I fragment of the resultant pISL-URA-TRP1-ISR plasmid. To balance the overall lengths of the intron in the repeat-containing cassettes, a unique 269 bp-long sequence from the coding part of the *tet*R gene of the pACYC184 plasmid (NEB) was amplified and cloned into *Sph*I located in the intron of the split *URA3* gene of the pISL-URA-TRP1-ISR plasmid. The plasmid was digested by *Smi*I and followed by the integration into chromosome III using a selection for tryptophan prototrophy.

Measurement of GAA/TTC fragility and expansion rates

Strains were grown on YPD plates at 30^{O}C for 3 days; a minimum of 14 independent colonies for each strain were taken for fluctuation tests to calculate the rate of fragility or expansions. Each colony was diluted in 250 μl of water and 10-fold serial dilutions were applied to obtain 100-200 colonies on YPD and 50-100 colonies on either canavanine-containing plates for the fragility assay or 5-FOA-containing plates for the expansion assay. Canavanine plates contain 60 mg/L L-canavanine and 5 mg/L adenine; the low amount of adenine is to differentiate mutations at *CAN1* (white colonies) from arm loss events (red colonies). The formula $\mu = f/ln(N\mu)$ was used to calculate the rate of fragility

or expansion (Drake, 1991). 95% confidence intervals were calculated as described in Dixon, 1969. Since GAA repeats are prone to size variations, only the colonies with the correct GAA tract sizes (pre-screened by PCR) were used for tests.

For complementation tests, wild-type and TET-*TAF4* strains were transformed with YCp50 or pKL155 containing wild-type *TAF4*, while TET-*POL3* strain was transformed with YCp50 or pBL304 carrying wild-type *POL3* (Morrison et al., 1993). Strains were grown on uracil drop-out plates for 4 days and appropriate dilutions were plated on uracil drop-out media for measuring total number of cells and on uracil drop-out media containing canavanine for estimation of arm loss events.

To estimate rates of fragility when the expression of essential genes is down-regulated, we first determined the optimal concentration of doxycycline that does not cause cell death (data not shown). This concentration, depending on the allele, varied from 0.1 μg/ml to 2 μg/ml of doxycycline in YPD media. Strains were grown on doxycycline plates for 4 days before fluctuation tests were carried out.

For monitoring GAA fragility and expansion in actively dividing and non-dividing cells, ~ 5 yeast cells were inoculated in 5 ml YPD and grown at 30^{o}C. Samples were then taken at indicated time-points and plated on plates for fluctuation tests as mentioned above. A minimum of 8 cultures were used in this experiment for each strain. Wild-type and TET-*TAF4* strains reached saturation in YPD at around 43 hours and 49 hours after inoculation, respectively. Frequencies instead of rates were compared between the actively dividing and arrested cells.

Analysis of yeast growth dynamics

Yeast strains were grown overnight in 5ml of YPD medium. Cultures were washed with distilled water once and resuspended in 10 ml of distilled water. 2 µl cultures were inoculated into 150 µl of SDC, -Trp or 0.75 g/L 5-FAA medium in the 96-well Costar flat bottom plate. The optical density of the culture was measured every 30min using a 600nm filter in the Biotek Synergy H4 plate reader at 30^{O}C with constant shaking. For building the growth curve, the mean of the OD values for each strain was plotted against the time-points used. Four technical and two biological replicates were used for each strain.

ChIP and qPCR

TAP-tagged *SUA7* was PCR amplified from *YPR086W* (Open Biosystems) and brought into strains containing no repeats, 120 copies of GAA/TTC repeats, or the natural *TRP1* promoter through one-step integration. Exponentially grown (OD600 = 0.8) cells for each strain were taken for ChIP procedure as described by Aparicio *et al*., 2005. Briefly, cells were cross-linked using 1% formaldehyde followed by 5 min incubation with 2.5M glycine. Cells were then harvested and lysed in a freezer mill. DNA was sheared using a sonicator to give fragments between 500 bp to 1 kb. Chromatin was then immunoprecipitated using Protein A-Sepharose beads. After centrifugation, the supernatants were taken as input samples and the proteins bound to the beads were eluted. The cross-linking was reversed in the eluate and the samples were kept as IP.

The binding of Sua7p with the GAA/TTC repeats and the *TRP1* promoter was measured by qPCR. The samples for PCR were prepared using qPCR kit (Clonetech) in

accordance to the manufacturer's instructions. PCR reactions were carried out on Applied Biosystems StepOne™ Real-Time PCR System. Each reaction was done in triplicates. 3' region of *POL5* was amplified and used as an endogenous control for normalization. Comparative CT method using ΔΔCt values was applied for relative quantification. The ratio of PCR amount from IP samples to input genomic DNA was calculated using the formula ratio=$2^{(-\Delta\Delta Ct)}$. Primers information used for strain construction and qPCR are available upon request.

2D analysis of replication intermediates

A colony prescreened for the full size GAA 120 repeats was inoculated into 800 ml YPD and grown overnight. Cells were arrested with alpha factor (50 ng/10^7 cells) at OD600 = 0.8. 2 μg/ml doxycycline was added into YPD during alpha factor arrest for TET-*RFA2* and *TET-TAF4* strains. ~90% cells were arrested in the G1 stage in the case of wild-type and TET-*RFA2* strains. In TET-*TAF4* strains only 70% cells could be arrested. Cells were washed and released into fresh YPD supplemented with pronase. Cells were harvested at 50 min, 50 min or 70 min after release for wild-type, TET-*RFA2* and TET-*TAF4* strains, respectively. Genomic DNA was then extracted as previously described (Friedman and Brewer, 1995). Genomic DNA was digested with *Afl*II such that GAA/TTC tract was positioned on the long arm of the replication arc. DNA was run in 0.4% agarose gel for 22 hours at room temperature in the first dimension and in a 1.2% agarose gel containing 0.3 μg/ml ethidium bromide for 10.5 hours at 4°C in the second dimension. Southern hybridization was carried out using a *LYS2*-specific probe at 64°C.

Quantification of 2D gels was carried out as previously described (Krasilnikova et al., 2004).

Supplemental References

Aparicio, O., Geisberg, J.V., Sekinger, E., Yang, A., Moqtaderi, Z., and Struhl, K. (2005). Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. Curr Protoc Mol Biol *Chapter 21*, Unit 21 23.

Bonneaud, N., Ozier-Kalogeropoulos, O., Li, G.Y., Labouesse, M., Minvielle-Sebastia, L., and Lacroute, F. (1991). A family of low and high copy replicative, integrative and single-stranded S. cerevisiae/E. coli shuttle vectors. Yeast *7*, 609-615.

Dixon, W.J., and Massey, F.J. Jr. (1969). Introduction to statistical analysis. . ed. (New York: McGraw-Hill, Inc.), p. 349.

Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci U S A *88*, 7160-7164.

Friedman, K.L., and Brewer, B.J. (1995). Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. Methods in enzymology *262*, 613-627.

Morrison, A., Johnson, A.L., Johnston, L.H., and Sugino, A. (1993). Pathway correcting DNA replication errors in Saccharomyces cerevisiae. The EMBO journal *12*, 1467-1473.

Smith, S., Hwang, J.Y., Banerjee, S., Majeed, A., Gupta, A., and Myung, K. (2004). Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *101*, 9039-9044.

Storici, F., Lewis, L.K., and Resnick, M.A. (2001). In vivo site-directed mutagenesis using oligonucleotides. Nat Biotechnol *19*, 773-776.