PNAS
www.pnas.org

Supplementary Information for
Replication-Independent Instability of Friedreich's Ataxia GAA Repeats during Chronological Aging

Alexander J. Neil[1#], Julia A. Hisey[1#], Ishtiaque Quasem[1], Ryan J. Mcginty[1], Marcin Hitczenko[2], Alexandra N. Khristich[1], Sergei M. Mirkin[1*]

Sergei M. Mirkin
Email: sergei.mirkin@tufts.edu

This PDF file includes:

Supplementary text
Figures S1 to S5
Tables S1 to S4
Legend for Video S1
SI References

Other supplementary materials for this manuscript include the following:

Video S1
Video S2

Supplementary text

Bayesian analysis

Before discussing the details of our statistical model, we present a general overview of Bayesian analysis and the benefits it offers for our data. We recommend Gelman et al. (1) and McElreath (2) for a more thorough discussion of the philosophy behind and mechanics of Bayesian analysis. Broadly speaking, a Bayesian approach quantifies uncertainty about parameters of interest by defining probability distributions for those parameters. More specifically, for parameters ϕ, the goal is to estimate a posterior distribution

$$\text{Prob}(\phi \mid \texttt{data}) \quad \propto \quad \text{Prob}(\texttt{data} \mid \phi)\text{Prob}(\phi), \tag{1}$$

that, as a probability distribution, allows for intuitive statements about ϕ, of the form "Given the observed data, how likely is ϕ to be in [*some range*]?" The posterior determines the plausibility of ϕ by balancing how well ϕ fits the data, as given by the likelihood in the first term on the right-hand side of (Equation 1), with any other information about ϕ one may have, captured by the prior distribution in the second term on the right-hand side of (Equation 1). The prior can be chosen to be more or less agnostic about likely values of ϕ, and its influence on the posterior decreases as more data are collected. Especially in biological settings, where accumulation of knowledge and understanding about the mechanisms at play yields a sense of reasonable parameter values, prior information can be very useful for estimation. In particular, if any experiments similar to those in this paper were run in the future, the Bayesian construct would naturally allow the incorporation of knowledge gained from our experiments into further inference, via the prior.

The alternative, frequentist approach makes statements about functions of the data that have certain desirable characteristics, such as unbiasedness for point estimates or coverage for confidence intervals, upon repeated sampling of data. Thus, these statistics are evaluated based on their average behavior over infinitely many hypothetical datasets. Any associated probability statements relate to the performance of the statistical methodology itself, rather than the plausibility of the parameters. Besides having a notoriously awkward interpretation, such an approach may not provide useful or even sensible conclusions for the particular dataset one has.

As a simple example, consider an experiment with 5 independent binary events, each occurring with some unknown probability of interest p, so that the data one observes is $X = 0, 1, \ldots, 5$ successes. A common frequentist estimate for p is given by $\hat{p} = \frac{X}{5}$. Such an approach is unbiased, in the sense that if one repeated the experiment infinitely many times, to generate $X_1, X_2, X_3, \ldots$, the average of the corresponding estimates, $\hat{p}_1, \hat{p}_2, \hat{p}_3, \ldots$ would be p for any value of p. In practice, however, we often have data from only one experiment. Now, consider the case where $X = 0$, yielding $\hat{p} = 0$. Although this estimate is appropriate for long-run unbiasedness, it seems often the case that the experimenter knows that $p \neq 0$, making $\hat{p} = 0$ a somewhat nonsensical estimate. By contrast, a Bayesian approach infers a distribution for p, from which a more meaningful posterior mean can be calculated. In this case, the prior guides how likely p is to be very close to 0 or not (if $p = 0.25$, with 5 events, X will be 0 almost 25 percent of the time). This aspect of the Bayesian approach is a key motivation for its adoption in this work.

In addition, because fitting a Bayesian model effectively reduces to evaluating and weighing strength of evidence for parameters in the prior and data likelihood, one can readily quantify relative probability of

more layered models and complicated parameter structures. In our context, this opens the door to fitting hierarchical models that allow for similarities across genotypes and mutation types. While the analysis in this paper looked at each genotype and mutation type separately, imposing monotonicity of mutation frequencies across time was a simple extension of the basic model that greatly improved inference quality in cases with low mutation counts by making use of data from multiple time points.

Description of statistical model

Below, we provide details of the statistical model used to estimate and compare mutation frequencies. We fit the described model separately for each combination of genotype and mutation type, so for simplicity of notation, we assume a generic such combination and drop any relevant indexing. The goal of the analysis is to infer distributions for f_t, the frequency of mutation at time point t, for all observed time points. In this exposition, we focus on the most common experimental case in which data are observed at two time points. The approach in the case of more observed time points is a trivial extension of the model presented.

At each time point, $t = 0$ and $t = 6$, we observe data from I_t distinct plates, which we index by $i = 1, \ldots, I_t$. Specifically, the data observed for plate i at time t are:

$$\begin{aligned} N_{ti} &= \#\text{ of plated cells,} \\ M_{ti} &= \#\text{ resistant cells,} \\ K_{ti} &= \#\text{ resistant cells sampled,} \\ J_{ti} &= \#\text{ of sampled cells with given mutation type.} \end{aligned}$$

For our analysis, we decompose f_t into the product of the resistance frequency, r_t, and the frequency of resistant cells with the given mutation type, p_t, so that $f_t = r_t \times p_t$. The data likelihood for plate i at time point t conditions on the number of cells plated and the number of resistant cells sampled, and for a given (r_t, p_t) is defined by the binomial distributions

$$\begin{aligned} M_{ti} \mid N_{ti} &\sim \text{Binomial}\left(N_{ti}, r_t\right), \\ J_{ti} \mid K_{ti} &\sim \text{Binomial}\left(K_{ti}, p_t\right), \end{aligned} \tag{2}$$

with independence assumed across plates and time points.

A key component of the analysis is our knowledge of the relative ordering among f_t, r_t across time points, specifically

(a.) $f_0 < f_6 < r_6$, and

(b.) $f_0 < r_0 < r_6$.

Because conditions (a.) and (b.) relate directly to f_t and r_t, we specify the model in terms of those

parameters, with $p_t = \frac{f_t}{r_t}$. Therefore, our four-parameter model is given by

$$\begin{aligned}
\text{logit}\,(f_0) &= \theta_0 && (3)\\
\text{logit}\,(f_6) &= \theta_0 + \theta_6 && (4)\\
\text{logit}\,(r_0) &= \theta_0 + \beta_0 && (5)\\
\text{logit}\,(r_6) &= \theta_0 + \theta_6 + \beta_6, && (6)
\end{aligned}$$

where $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$ maps a probability in $(0,1)$ to the real line. Conditions (a.) and (b.) are respectively satisfied by imposing the restrictions

$$\theta_6, \beta_6 > 0 \quad \text{and} \quad 0 < \beta_0 < \theta_0 + \theta_6. \tag{7}$$

The logistic model framework is flexible as it it allows extensions that assume more complicated sources of variation in the data, including commonality among similar genotypes or added variability due to plate attributes, by restructuring the additive effects that define f_t and r_t.

Adopting a Bayesian approach, the posterior distribution of the four parameters can be written as

$$\begin{aligned}
\text{Prob}\,(\theta_0, \theta_6, \beta_0, \beta_6 \mid \{N_{ti}\}, \{M_{ti}\}, \{K_{ti}\}, \{J_{ti}\}) \quad \propto \quad & \text{Prob}\,(\theta_0, \theta_6, \beta_0, \beta_6) \times \\
& \prod_{t=0,6} \prod_{i=1}^{I_t} \text{Prob}\,(M_{ti} \mid N_{ti}, r_t)\, \text{Prob}\,(J_{ti} \mid K_{ti}, p_t)\,,
\end{aligned}$$

where the the first line of the right-hand side represents a prior distribution for the four parameters and the second line represents the complete data likelihood, defined by the binomial distributions in (2) and the mappings from θ_t, β_t to r_t, p_t implied by (3)-(6).

Subject to the restrictions in (7), we adopt the following priors:

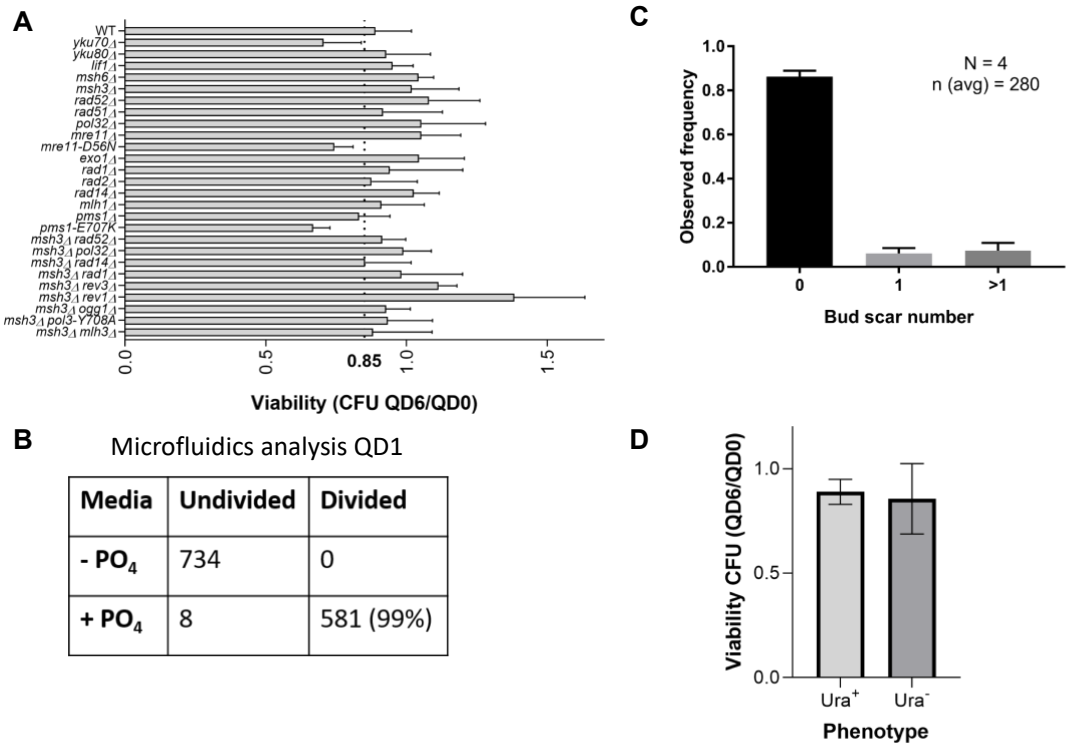
$$\begin{aligned}
\theta_0 &\sim \text{Normal}(-9, 3)\\
\theta_6 &\sim \text{Normal}(0, 3)\\
\beta_0 &\sim \text{Normal}(0, 10)\\
\beta_6 &\sim \text{Normal}(0, 10).
\end{aligned}$$

The prior for θ_0 represents a vaguely informative prior that reflects an expectation that mutation frequencies are rare, with a little over 95% of the mass falling between $1e-7$ and $5e-2$. The remaining three priors are uninformative by design, recognizing that differences between f_0 and f_6, and especially between f_t and r_t, might be substantial on the logit scale.

Sensitivity to the choice of priors was studied by fitting models and comparing posterior distributions of f_t based on several assumed prior distributions. In cases where mutations were observed among sampled resistant cells ($J_{ti} > 0$ for some i), we found the results to be robust to the choice of priors, including other priors based on an understanding of the general range of mutation frequencies as well as more agnostic priors for θ_0. This fact should be of no surprise as the large numbers of plated cells provide a great deal of information about the true frequency of the mutations. The resulting impact on the posterior distribution is that the data likelihood dominates the prior distribution. The role of the prior has a more significant

impact on the estimated lower ranges of mutation frequencies when no mutations were observed ($J_{ti} = 0$ for all i), since the observed data are consistent with any choice of f_t sufficiently near 0. In this case, our chosen prior for θ_0 guides the lower end of plausible values based on our data to include, but be very likely larger than, $5e-8$. Most importantly, the upper end of the mutation frequency distribution remains data-driven for biologically sensible choices of prior, since sufficiently high values of f_t prove incompatible with data in which no mutations were observed.

Models were fit using the `rstan` package in `R`. For each model, four Markov Chain Monte Carlo chains were run for 2,000 iterations each, with the last 1,000 draws from each chain stored to generate a total of 4,000 posterior draws for each parameter. Paired values of θ_0 and θ_6 are converted to 4,000 paired draws of f_0 and f_6 per (3) and (4). These, in turn, can be used to generate a posterior sample of 4,000 estimates of the difference, $f_6 - f_0$, from which probabilistic assessments about the changes in mutation frequency over time can be made.



Media	Undivided	Divided
- PO_4	734	0
+ PO_4	8	581 (99%)

Fig. S1. Viability and assessments of quiescence. (A) Viability determined by CFUs on YPD from individual cultures at QD6 divided by QD0s. Bars represent mean viability and error bars represent 95% confidence intervals calculated via the Wilson method. (B) Data from 24hr observation of QD1 cells in a microfluidics chamber in media with (+PO_4) or without (-PO_4) phosphate. Cells were considered divided if they underwent at least one division in the 24hr period. (C) Analysis of bud scars per cell as visualized using calcofluor white staining of QD0 cells. Bars represent the frequency of cells with the designated number of bud scars. Error bars represent 95% confidence intervals calculated via the Wilson method. (D) Viability of a Ura^+ and a Ura^- strain was determined by CFUs on YPD from individual cultures at QD6 divided by QD0. Bars represent mean viability and error bars represent 95% confidence intervals calculated via the Wilson method.

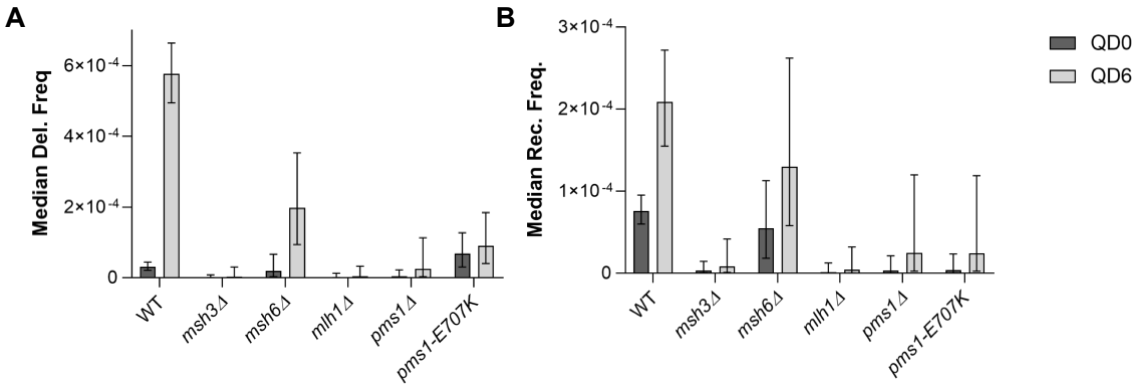


Fig. S2. Deletion (A) and recombination (B) frequencies at the beginning and end of chronological aging in mismatch repair deficient strains. Bars represent median deletion frequency or recombination frequency at QD0 (dark grey) and QD6 (light grey) timepoints. Error bars represent 95% credible intervals of the estimated distributions.

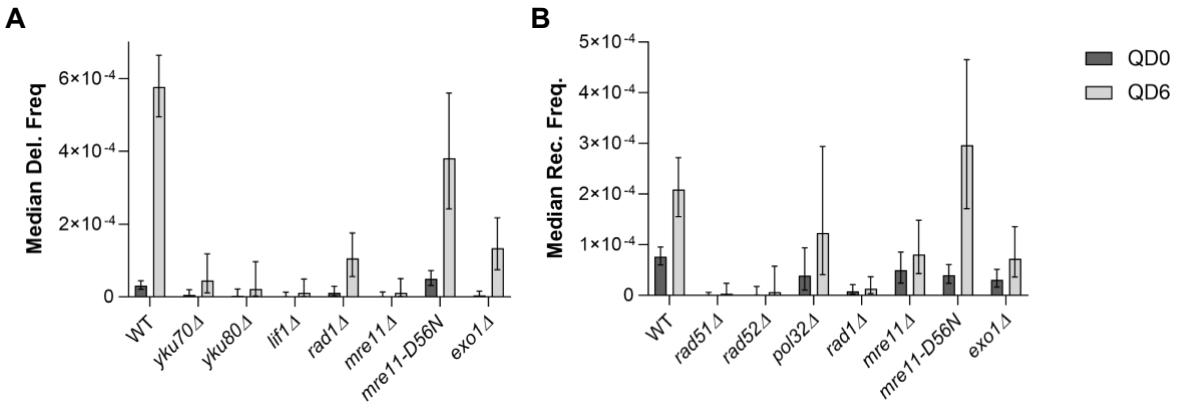


Fig. S3. Deletion and recombination frequencies at the beginning and end of quiescence in NHEJ and HR mutants. (A) Deletion frequencies at the beginning and end of chronological aging in NHEJ, Rad1, and Exo1 deficient strains. (B) Recombination frequencies at the beginning and end of chronological aging in HR, Rad1, and Exo1 deficient strains. Bars represent median deletion frequency or recombination frequency at QD0 (dark grey) and QD6 (light grey) timepoints. Error bars represent 95% credible intervals of the estimated distributions.

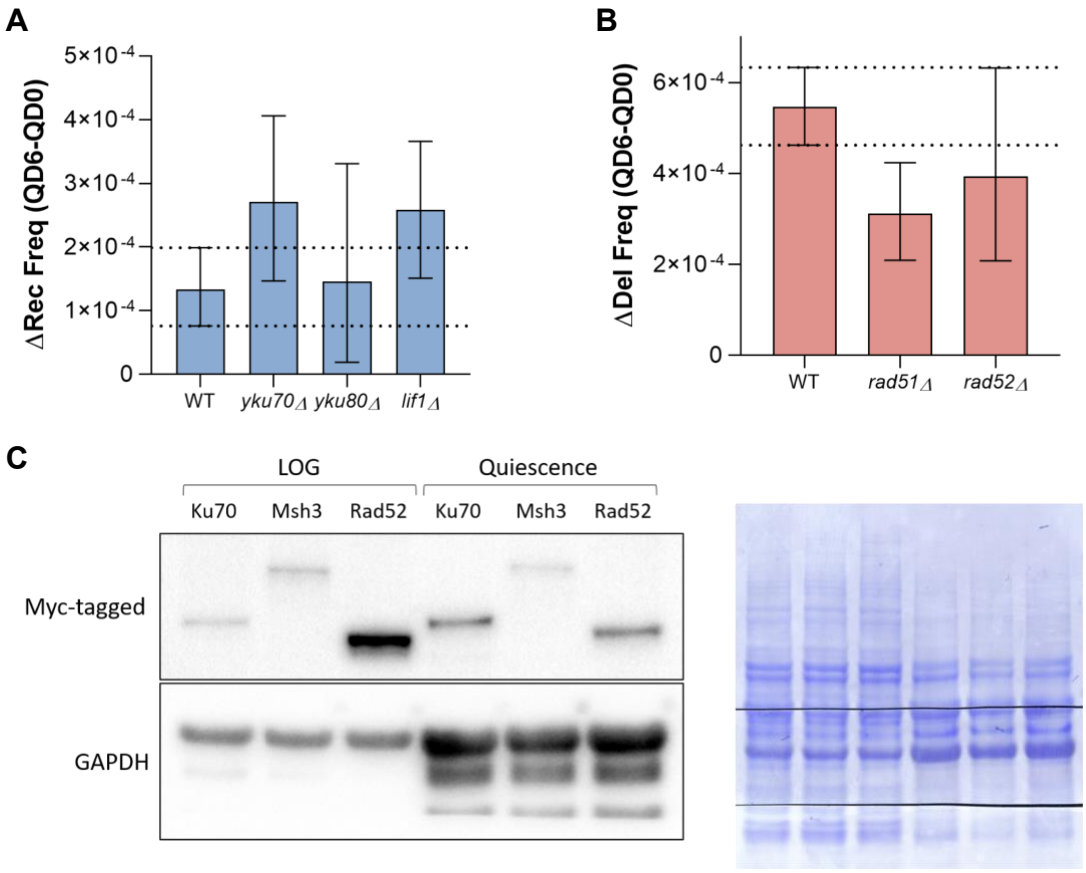


Fig. S4. Change in recombination and deletion frequencies in mutants required for the alternate pathway and repair protein levels in quiescence. (A) Eliminating NHEJ does not affect gene conversion frequency and (B) eliminating HR does not affect deletion frequency during chronological aging. Bars represent the median change in recombination frequency (Δrec freq) (blue) (A) or deletion frequency (Δdel freq) (red) (B) between QD0 and QD6. Error bars represent 95% credible intervals of the estimated distributions. Biological significance was defined as no overlap between the WT and mutant 95% credible intervals. Dashed lines indicate WT 95% credible interval. (C) Western blot showing abundance of myc-tagged Ku70, Msh3, and Rad52 proteins extracted from LOG or Q cells. Western blot probed using myc or GAPDH antibodies. Coomassie stained membrane (right) shows each lane's total protein in the same order as the western blot.

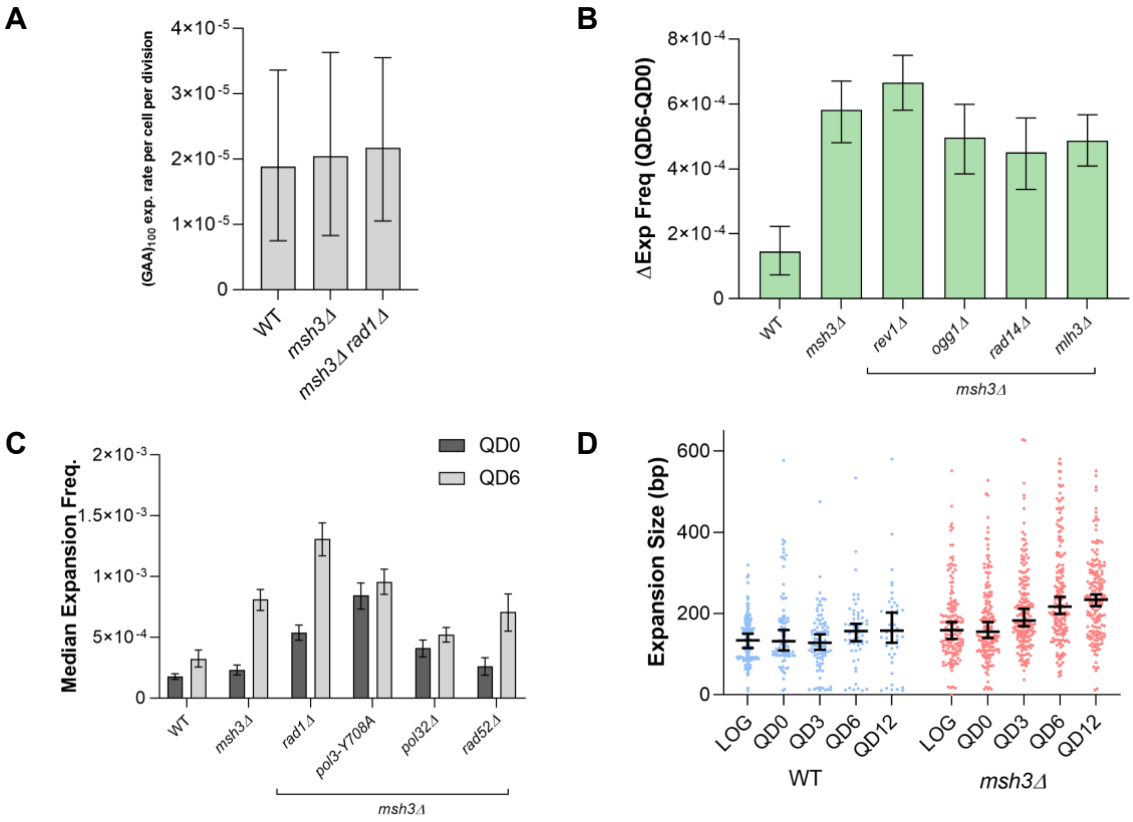


Fig. S5. Expansion rates, frequencies, and sizes in dividing and quiescent cells. (A) Rad1 and Msh3 deletion do not alter expansion rate in dividing cells. Bars represent $(GAA)_{100}$ expansion rates calculated using our FluCalc web tool. Error bars represent 95% confidence intervals. (B) Rev1, Ogg1 and Rad14 are not involved in expansions in chronological aging. Bars represent the median change in expansion frequency (Δexp freq) between QD0 and QD6. Error bars represent 95% credible intervals of the estimated distributions. Biological significance was defined as no overlap between the double mutant and *msh3Δ* 95% credible intervals. (C) Expansion frequency at the beginning and end of chronological aging in strains shown in Figure 6B. Bars represent median expansion frequency at QD0 (dark grey) and QD6 (light grey) timepoints. Error bars represent 95% credible intervals of the estimated distributions. (D) Expansion length as determined by agarose gel electrophoresis at QD0, QD3, QD6 and QD12 in WT and *msh3Δ* strains. Black bars and whiskers represent medians and 95% confidence intervals.

Table S1. Strains used in this study

Strain	Genotype	Comments
CH1585	*MAT***a**, *leu2-Δ1*, *trp1-Δ63*, *ura3–52*, *his3–200*	Shishkin et al. 2009
KS001	CH1585, *bar1::HIS3*	Shah et al. 2012
KS005	KS001, *ChrIII(75594-75641):: UR-(GAA)$_{100}$-A3-TRP1*	Shah et al. 2012
YEG101-1/2	KS005, *msh3::HphMX4*	
YEG102-1/2	KS005, *msh6::HphMX4*	
YAN134	KS005, *yku70::HphMX4*	
YAN131/132	KS005, *yku80::HphMX4*	
YAN195/196	KS005, *lif1::HphMX4*	
YAN175/176	KS005, *rad51::HphMX4*	
YAN135/136	KS005, *rad52::HphMX4*	
YAN150/151	KS005, *pol32::HphMX4*	
YAN153/155	KS005, *mre11::HphMX4*	
YAN191/192	KS005, *rad14::HphMX4*	
YAN189/190	KS005, *rad2:: HphMX4*	
JAH25	KS005, *mlh1:: HphMX4*	
JAH41	KS005, *pms1::LEU2*	
JAH16	KS005, *rad1::HphMX4*	
JAH205	KS005, *mre11-D56N*	
JAH275	KS005, *pms1-E707K*	
JAH285	KS005, *exo1::NatMX4*	
YAN206	YEG101, *rad14::NatMX4*	
YAN207	YEG101, *pol32::NatMX4*	
YAN204/205	YEG101, *rad52::NatMX4*	
JAH39	YEG101, *rad1::NatMX4*	
JAH48	YEG101, *rev3::NatMX4*	
JAH106	YEG101, *rev1::KanMX6*	
JAH94	YEG101, *ogg1::NatMX4*	
JAH184	YEG101, *pol3-Y708A*	
JAH238	YEG101, *mlh3::NatMX4*	

Table S2. Primers used in this study

Name	Sequence
For making knockouts	
MSH3F	gtacttttgagagccaaaagcagtgcaaatagatttattttgttgaatctattaacaataCAGCTGAAGCTTCGTACGC
MSH3R	gcataagaaattgctataccatcgtgcgtgccagtacctcttcccacttcgtctaataatCATAGGCCACTAGTGGATCTG
MSH6F	gataagatttttaattggagcaactagttaattttgacaaagccaatttgaactccaaaCAGCTGAAGCTTCGTACGC
MSH6R	taaagcatggatgtcttaatgatttaaatttcagaaaaccatttaattgagtattcgtttCATAGGCCACTAGTGGATCTG
YKU70F	gcgctcagtcactaatgcatttggcaatagtggagaacttaacgatcaagtggatCAGCTGAAGCTTCGTACGC
YKU70R	gcctttggatgattggatcttctgacttctcagattctaaaattttatttcgCATAGGCCACTAGTGGATCTG
YKU80F	ccctggaatcattgatcatcagttggtgaatccattccagtagcacctgttCAGCTGAAGCTTCGTACGC
YKU80R	ataatgtcaagtgagtcaacaactttcatcgtggatgtttcaccatcaatCATAGGCCACTAGTGGATCTG
LIF1F	cattggatgacttatttatgtagggtgctatgtcccagctgacggagttcattagCGGATCCCCGGGTTAATTAA
LIF1R	ataggctatgtttctatatccgtttccgattctgtctgcaagcaaggttccctaTCATAGGCCACTAGTGGATCTG
RAD51F	aaatgttggaaatgcaccactaccgttcttcaaccaatctagtttagctatTTAGAACGCGGCTACAATTA
RAD51R	aaagaggagaattgaaagtaaacctgtgtaaataaatagagacaagagaccaaatacctaCCCTGATTCTGTGGATAACC
RAD52F	cgaatggcgtttttaagctattttgccactgagaatcaacaaatgcaaacaaggaggttGCCAGATCTGTTTAGCTTGCCT
RAD52R	ggtttcacgcggtacttgattcccagccccttctagcatatgaggccccagttcttTATCATCGATGAATTCGAGCTCGTT
POL32F	ataatatttcacattaactaacaaccagaaataggctttagttaactcaatcggtaattaCAGCTGAAGCTTCGTACGC
POL32R	catttgtattatacattacatcacaattagtaatggaaagtgtttggaaaaaaaagaagaCATAGGCCACTAGTGGATCTG
MRE11F	ttaagagaatgcagacaattgacgcaagttgtacctgctcagatccgataaaactcgactCAGCTGAAGCTTCGTACGC
MRE11R	tcgcgaaggcaagcccttggttataaataggatataatataatatagggatcaagtacaaCATAGGCCACTAGTGGATCTG
RAD14F	agagtttggatcttcgtagtgaaggtatcgaacgtaacgctatgactcccCGGATCCCCGGGTTAATTAA
RAD14R	tatacataaccaacatttaaatgtcaatttcttcagtttctagcccgcagCATAGGCCACTAGTGGATCTG
RAD2F	ctagtataacccattcgaacctccgtggaggcattaaaagggagagtgaaaTTAGAACGCGGCTACAATTA
RAD2R	aaggaccgtatatatctactattcctggatcggttgactttgttaacatgcagaaacaCCCTGATTCTGTGGATAACC
MLH1F	atgtctctcagaataaaagcacttgatgcatcagtggttaacaaaattGCCAGCTGAAGCTTCGTACGC
MLH1R	ttaacacctctcaaaaactttgtatagatctggaaggttggctatttccaCATAGGCCACTAGTGGATCTG
PMS1F	gtacatagctagaactctagaaagcacagattaataccgattctaatacagattCGGCATCAGAGCAGATTGT
PMS1R	attatgaacgactctggtcatagtcttcttattcagcggtttacctatcattatACTGGAACAACACTCAACCCT
EXO1F	ctttttttcttacgcgtctttagcaaaggcgggaagtacaataactagttttgtgcacaaCCCCGGGTTAATTAAGGCG
EXO1R	gtcctcacatgcggcgtgcattgttcatagcggggcaaacatacttgtggcttaatttgacGAGCGCCCAATACGCAAAC
RAD1F	cttatctcctggagtaagctatagccacagtcaatatcgcgtctaatgaaaGGCATCAGAGCAGATTGTA
RAD1R	aaagattcaaagagcatgtctaacttataacatatacggtcgaagtcaccaaatgaataCCCTGATTCTGTGGATAACC
REV3F	atgtcgagggagtcgaacgacacaatacagagcgatacggttagatcatcCGGATCCCCGGGTTAATTAA
REV3R	gcgagacatatctgtgtctagattaccaatcatttagagatattaatGCTGCATAGGCCACTAGTGGATCTG
REV1F	acagattttctcaaaataaatcgatactgcatttctaggcatatccagcgCCAGATCTGTTTAGCTTGCCTCG
REV1R	gatattacaggtaatgttcgcaaactgcgtgtttactgtatgctgaaatgTATCATCGATGAATTCGAGCTCGTT
OGG1F	atgtcttataaattcggcaaacttgccattaataaaagtgagctatgtGCCAGCTGAAGCTTCGTACGC
OGG1R	ctaatctatttttgcttctttgatgtgaagatcagacaattcaactttCATAGGCCACTAGTGGATCTG
MLH3F	cataaaccagcgaggctttcaaggaagaatgaacgtgaactcgtcaactcCGGATCCCCGGGTTAATTAA
MLH3R	gcgcaatttaaaatgcaggcgacaaaccttgttccaggattaaggttctCGCATAGGCCACTAGTGGATCTG
For checking knockouts (internal)	
MSH3_int_F	GCCATTGAGTGCATTACATCC
MSH3_int_R	AAATGAGGCAACTGAGTGGT

MSH6_int_F	CGATATTGCAGACGACAAAGAG
MSH6_int_R	CTCGTTCATCCACTAACCATTG
YKU70_int_F	GCATGAAGATATCAGACAAGAAGC
YKU70_int_R	TGCTCGATGAACGGAACC
LIF1_int_F	CTGTTGTGTTAACGATTCTTACCG
LIF1_int_R	GGAGAGTTTAACTCAGAAATTGGC
RAD51_int_F	AGATCGGAGCTGATTTGTTTGAC
RAD51_int_R	CTTCACCGCCACCAATATCC
RAD52_int_F	GCCAAGAAATCTGCCGTTAC
RAD52_int_R	TGAGCTTTCGCTGATTTCATCC
POL32_int_F	GACCACGCCAGAAGAAACAA
POL32_int_R	GCTGTCGTTTCCAACAAGTC
MRE11_int_F	GATGGTGAAGGTGACATGGTT
MRE11_int_R	GTTAGCAACACGTCCCACA
RAD14_int_F	CCGATGACCAAGAATTTGAATCTG
RAD14_int_R	CTGGATGCTCCTTAGAACACTG
RAD2_int_F	ACCGTCTCAGCAGGAGGATA
RAD2_int_R	GAGATTTAGTGGGAACGTCCTC
MLH1_int_F	GCGTTGATGGAAAGGTGTGT
MLH1_int_R	CAATGGCAGATAATTCGGCG
PMS1_int_F	TACCGCCAGAAGCGAAAGTAAT
PMS1_int_R	CACACTCCGAACCGATGACA
EXO1_int_F	CAGCGGGAGGGAAAACTGAT
EXO1_int_R	CTCTGTTGGCTAGAGGTTGGTG
RAD1_int_F	CGAACTGGCACCGAATTTCT
RAD1_int_R	CTAAAGTAGGCCCCTGAAGG
REV3_int_F	GCATGCACACCCCTCATAGTAAGT
REV3_int_R	TGGCATTTGACTCTGGCAAGTTCC
REV1_int_F	AATGTGTAGGGTCGGCATTG
REV1_int_R	TGTGCAACCATTCGTTCCTTGA
OGG1_int_F	ACTGTGGCAATCAGGACGCC
OGG1_int_R	GGCTCGAGAAGTTAGCTCCTC
MLH3_int_F	GTTTGGGGCAATAATACCACCGG
MLH3_int_R	CGAACATCAAGGATGAAGACGGG
For checking knockouts (external)	
MSH3_chk_F	GTGTTCAAATCACGGTATGTGG
MSH3_chk_R	AAGGGGCAGTCACTTAACTCAG
MSH6_chk_F	TGACATAATGAATGGCTTCTGG
MSH6_chk_R	CCCGTTAACAATCCTAATCTGG
YKU70_chk_F	TTAATTGACTCTCGGTAGCCAAGTT
YKU70_chk_R	CGTCTTTATATATTGAATTTCGGCTT
YKU80_chk_F	TGCCGAGCACTTCAACCAATTT
YKU80_chk_R	CCGTCAGGGCATTTGTTGTCAT
LIF1_chk_F	GACGTGTCAATGCATAGAACTG
LIF1_chk_R	GATGCGATACTATAATACTCTTTGCC
RAD51_chk_F	CAATTCGCAAGAAACGCACT
RAD51_chk_R	AAGTAGTCATCGGGAAGAAGAGTA
RAD52_chk_F	ACGTCGCTAAAGATGGTATGGTA
RAD52_chk_R	CTAGAGGATTTTGGAGTAATAAATAATGATG
POL32_chk_F	TTTCCACTACGGTGTAACTTTCC
POL32_chk_R	TGTCCTTCGGATGGTATATTAGG
MRE11_chk_F	CCAATCATTTCGACCGTCACTC
MRE11_chk_R	CACAAGGGGACGGTTAATGAGG

RAD14_chk_F	CGTTTGCTAAGTTGTAGGGAGA
RAD14_chk_R	GTACGAGTGACAAATGGGATATCA
RAD2_chk_F	TTGATGTTTCCAGAGGATGTGA
RAD2_chk_R	CCCAATTCCAGAGGAGTGAT
MLH1_chk_F	GTAATCGCGCTAGCATGCTA
PMS1_chk_F	GGTTTGTGGGCGTACTGTC
PMS1_chk_R	AAATGAGCTCCAATCACGTAATTCC
EXO1_chk_F	GTATTACGTCCAAACTAAGTTCGCG
EXO1_chk_R	GACCGCTAGCGGCTTGATTAG
RAD1_chk_F	ATGTAATCAACCTGTCCCGTCC
RAD1_chk_R	AGTGGAAGATGAATTGCGGATGA
REV3_chk_F	CGAGTGCAGTGCGTCTAGAAATAGTGT
REV1_chk_F	TACGGCAACCTTTAAGCACC
REV1_chk_R	GAGTCGGCCATTCCAATACC
OGG1_chk_R	TATCTTCCAACGCCTTGGTG
MLH3_chk_F	TTTGCGTTTATTTGCGAGCG
MLH3_chk_R	GGTTTTGACAACGTGATGAGG
For making CRISPR guides	
PMS1E707K_gRNA_F	CTGCGATTTAAAAACTGTCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
PMS1E707KR_gRNA_R	TGACAGTTTTTAAATCGCAGCGATCATTTATCTTTCACTGCGGAG
MRE11D56NF_gRNA_F	ATTCACGTGAAAAAGATCACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
MRE11D56NR_gRNA_R	GTGATCTTTTTCACGTGAATCGATCATTTATCTTTCACTGCGGAG
POL3Y708AF_gRNA_F	TAACTCTGTCTATGGTTTTACAGGAGCGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
POL3Y708AR_gRNA_R	TCGCTCCTGTAAAACCATAGACAGAGTTACGATCATTTATCTTTCACTGCGGAG
For making CRISPR donor templates	
PMS1E707K_don_F	TTATCATAGTGACCAGAAAAGTTGATAACAAATATGATCTGTTTATTGTCGATCAGCATGCAAGTGATaagAAGTATAATTTCGAAACACTGCAGGCAGTG
PMS1E707K_don_R	CTACCGGCTGAGGTATTATCAATTTCTGCGATTTAAAAACTGTCACtGCCTGC
MRE11D56N_don_F	AAAACTTTCCATGAAGTCATGATGCTGGCCAAAAATAACAACGTAGACATGGTTGTACAGagtGGTaATCTTTTTCACGTGAATAAGCC
MRE11D56N_don_R	CTCGCAAGGCTTGTCACCCATGCAACATAATCTC
POL3Y708A_don_F	AGAAGGATCCATTCAAAAGAGATGTTTTAAATGGTAGACAATTGGCTTTG
POL3Y708A_don_R	CATAAGCAGTAACAGATGAAGAAATGGCTAAACATGGCAATTTACCCACTGTCGCTCCTG
For checking CRISPR mutations	
PMS1E707K_chk_F	TACCGCCAGAAGCGAAAGTAAT
PMS1E707K_chk_R	AAATGAGCTCCAATCACGTAATTCC
MRE11D56N_chk_F	CCAATCATTTCGACCGTCACTC
MRE11D56N_chk_R	ATGACCCCATATCACCATATCC
POL3Y708A_chk_F	GTGCAAAGGCGCTAAAGGTG
POL3Y708A_chk_R	CCAGAATAAACCTGCATAACGC
For detecting mutations	
expF	CTCGATGTGCAGAACCTGAAGCTTGATCT
expR	GCTCGAGTGCAGACCTCAAATTCGATGA
delF	CCTTTGCCATTTATTGTCGCAGTAAGGAAAAGCGCAGA
delR	CGACCGAATTTCTTGAAGACGAAAGGGCCTCGTGATAC
recF	ATGACCCACTCAGGTGTTAAA
recR	ATGTAAATACTAGTTAGTAGATGATAGTTGATTT
URA3_seq_1	GCCATTTATTGTCGCAGTAAGG

URA3_seq_4	CGAATTTCTTGAAGACGAAAGGG
Delseq_1A	GGAGCACAGACTTAGATTGGTAT
Delseq_3	GCCCCTATTTATTCCAATAATATCGTG

Table S3. Media for starving and maintaining quiescent yeast

Component	Final Concentration
Phosphate limited media	
Yeast Nitrogen Base – Amino Acids – Ammonium Sulfate – Sugar – Phosphate (Sunrise Science Products 1532)	0.71g/L
Potassium Chloride	1 g/L
Ammonium Sulfate	5g/L
Dextrose	20g/L
Dropout mix – uracil – YNB (US Biological D9535)	2g/L
Uracil	1g/L
Potassium Phosphate	0.5g/L
No Phosphate Media	
Same as above with the following modifications:	
Dextrose	2g/L
Potassium Phosphate	0g/L

Table S4. Microhomologies at deletion junctions from quiescent cells

Age	Sample	Microhomology
QD0	1-9	A
QD0	2-4	AAAG
QD0	5-15	AGAAGAAG
QD0	6-16	GG
QD6	1-3	A
QD6	1-8	GAA
QD6	1-11	T
QD6	2-2	G
QD6	2-6	TTTA
QD6	3-8	CTAA
QD6	3-14	AAGG
QD6	4-1	CTCA
QD6	4-8	TC
QD6	5-1	A
QD6	6-8	AAG

Supp Video 1 (separate file): Video of 24hr observation of QD1 cells in a microfluidics chamber in media with (+PO_4) phosphate. 1s of video is equivalent to 1.6hrs of observation.

Supp Video 2 (separate file): Video of 24hr observation of QD1 cells in a microfluidics chamber in media without (-PO_4) phosphate. 1s of video is equivalent to 1.6hrs of observation.

SI References

1. Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian Data Analysis (Chapman & Hall/CRC, Boca Raton, FL), 2nd edition.
2. McElreath R (2016) Statistical Rethinking (Chapman & Hall/CRC, Boca Raton, FL).