Analysis of APOBEC-induced mutations in yeast strains with low levels of replicative DNA polymerases

Yang Sui[a,b], Lei Qi[a,b], Ke Zhang[a], Natalie Saini[c], Leszek J. Klimczak[d], Cynthia J. Sakofsky[c], Dmitry A. Gordenin[c], Thomas D. Petes[b,1], and Dao-Qiong Zheng[a,b,1]

[a]Ocean College, Zhejiang University, 316021 Zhoushan, China; [b]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC 27710; [c]Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709; and [d]Integrative Bioinformatics Support Group, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709

Contributed by Thomas D. Petes, February 28, 2020 (sent for review December 24, 2019; reviewed by Munira A. Basrai and Hannah L. Klein)

Yeast strains with low levels of the replicative DNA polymerases (alpha, delta, and epsilon) have high levels of chromosome deletions, duplications, and translocations. By examining the patterns of mutations induced in strains with low levels of DNA polymerase by the human protein APOBEC3B (a protein that deaminates cytosine in single-stranded DNA), we show dramatically elevated amounts of single-stranded DNA relative to a wild-type strain. During DNA replication, one strand (defined as the leading strand) is replicated processively by DNA polymerase epsilon and the other (the lagging strand) is replicated as short fragments initiated by DNA polymerase alpha and extended by DNA polymerase delta. In the low DNA polymerase alpha and delta strains, the APOBEC-induced mutations are concentrated on the lagging-strand template, whereas in the low DNA polymerase epsilon strain, mutations occur on the leading- and lagging-strand templates with similar frequencies. In addition, for most genes, the transcribed strand is mutagenized more frequently than the nontranscribed strand. Lastly, some of the APOBEC-induced clusters in strains with low levels of DNA polymerase alpha or delta are greater than 10 kb in length.

DNA replication stress | single-stranded DNA | APOBEC | mutation | DNA polymerase

Cells derived from solid metastatic tumors often have very high levels of chromosome rearrangements and aneuploidy, and it has been argued that such genomic changes may be a consequence of DNA replication stress (1, 2). In addition, when mammalian cells are exposed to drugs that inhibit DNA synthesis such as aphidicolin, there is an elevated rate of chromosome breakage at regions termed “fragile sites” (3).

Saccharomyces cerevisiae, like other eukaryotes, has three replicative DNA polymerases, alpha, delta, and epsilon; the catalytic subunits of these polymerases are encoded by the genes *POL1*, *POL3*, and *POL2*, respectively (4). DNA polymerase alpha synthesizes the RNA–DNA primers that are extended by the other replicative DNA polymerases (5). DNA polymerases delta and epsilon are responsible for most of the synthesis on the lagging and leading strands, respectively (6). However, the catalytic domain of DNA polymerase epsilon is not required for DNA replication or cell viability (7) and, in certain mutants of DNA polymerase epsilon, DNA polymerase delta is capable of DNA synthesis on the leading strand (8).

To examine the genomic changes that result from DNA replication stress, we constructed strains in which the levels of the replicative DNA polymerases alpha, delta, and epsilon, were regulated using a galactose-inducible promoter (9). We previously showed that strains with low levels of DNA polymerase alpha (10, 11) or delta (12) have very elevated rates of mitotic recombination, large deletions/duplications, and aneuploidy.

One possible explanation for the hyper-recombination phenotype associated with the low-polymerase strains is that low levels of DNA polymerase lead to stalled breakage-prone replication forks or forks in which synthesis of the leading and lagging strands has been uncoupled. Loss of coupling could result in large single-stranded regions at the fork. In addition, the high level of mitotic recombination in the low-polymerase strains likely reflects a high level of double-stranded DNA breaks (DSBs). Processing of these broken ends would represent another source of single-stranded DNA (13). We decided to look for single-stranded regions in low-polymerase strains using APOBEC3B (A3B), a member of the mammalian single-strand-specific cytosine deaminases.

In human cells, members of the APOBEC family have an antiviral role as well as reducing the rate of retrotransposition (14). APOBEC proteins can also mutate single-stranded chromosomal DNA resulting in high levels of mutations in multiple tumor types (15, 16). APOBEC and related proteins have been expressed in *S. cerevisiae* to detect single-stranded DNA in multiple studies (17–25), leading to a number of generalizations. First, expression of APOBEC substantially elevates mutation rates. Second, in yeast strains lacking uracil DNA glycosylase, most of the observed mutations are C-to-T (G to A in the reverse complement) mutations, as expected from the cytosine deaminase activity of APOBEC that converts C to uracil. Third, APOBEC-induced mutations are often clustered (18–20, 23–26). Fourth, the mutations are more abundant on the lagging-strand template than the leading-strand template (21).

In our analysis, we show that low levels of DNA polymerase alpha, epsilon, or delta greatly increase the numbers of APOBEC3B-associated mutations. Taken together with our

Significance

Perturbations in DNA replication cause high levels of chromosome rearrangements and it has been suggested that DNA replication stress promotes oncogenesis. In this study, we show that low levels of the DNA polymerases involved in replication in the yeast *Saccharomyces cerevisiae* lead to greatly elevated levels of single-stranded DNA. We suggest that these single-stranded regions are fragile, generating the DNA breaks that initiate translocations and other types of chromosome rearrangements.

Author contributions: T.D.P. and D.-Q.Z. designed research; Y.S. and L.Q. performed research; Y.S., K.Z., N.S., L.J.K., C.J.S., D.A.G., T.D.P., and D.-Q.Z. analyzed data; and Y.S., T.D.P., and D.-Q.Z. wrote the paper.

Reviewers: M.A.B., National Cancer Institute; and H.L.K., New York University School of Medicine.

Competing interest statement: C.J.S., D.A.G., T.D.P., and H.L.K. are coauthors on a 2019 methods review article.

Data deposition: The DNA sequencing data are on the National Center for Biotechnology Information (NCBI) Website: Sequence Read Archive (accession no. PRJNA314677).

See online for related content such as Commentaries.

[1]To whom correspondence may be addressed. Email: tom.petes@duke.edu or zhengdaoqiong@zju.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1922472117/-/DCSupplemental.

First published April 10, 2020.

previous results, our analysis suggests that the high level of genomic instability observed in strains with low levels of replicative DNA polymerases reflects the fragility of replication forks that have extended regions of single-stranded DNA.

Results

Experimental System and Rationale. Strains in which the synthesis of the catalytic subunits of DNA polymerases alpha (SY43-LA), epsilon (SY44-LE), and delta (SY45-LD) was regulated by the concentration of galactose in the medium were constructed (*SI Appendix*, Table S1). In these strains, the genes encoding the catalytic subunits of DNA polymerases alpha, epsilon, and delta (*POL1*, *POL2*, and *POL3*, respectively) were fused to the *GAL1,10* promoter; the abbreviations LA, LE, and LD indicate low alpha, low epsilon, and low delta DNA polymerases, respectively. In the strain SY46-WT, all polymerases are regulated by their native promoters. Relative to the wild-type (WT) strain, by Western analysis, strains with the *GAL-POL1* or the *GAL-POL3* fusions have 10% of the levels of alpha and delta DNA polymerases, when grown in low-galactose (0.005% galactose, 3% raffinose) medium, and a fivefold elevation of these polymerases when grown in high-galactose medium (10, 27). Low-galactose medium results in slow growth and elevated rates of mitotic recombination and large chromosome deletions and duplications (11, 12). Growth of cells in high-galactose medium restores normal growth rates (10–12) and reduces rates of genomic instability (10, 27). Since polymerase epsilon was not detectable by Western analysis, we examined expression of *POL2* by reverse-transcriptase PCR (details in *SI Appendix*). When grown in low-galactose medium, strains with the *GAL-POL2* fusion have about one-third the level of *POL2* expression as that of an isogenic wild-type strain grown in rich nutrient media (YPD); in high-galactose medium, strains with that fusion had about 1.3-fold more polymerase epsilon than the isogenic wild-type strain.

All four strains were transformed with a plasmid-borne copy of APOBEC3B which causes frequent C→U modifications in single-stranded DNA at the replication fork (21). The strains also had an *UNG1* mutation, preventing the conversion of uracil to an abasic site. The strains were grown from a single cell to a colony on solid medium containing low levels of galactose. The strains SY43-LA, SY44-LE, and SY45-LD were grown for a single passage (single cell to colony) on medium containing low galactose, whereas the wild-type strain SY46-WT was grown for 5 or 10 passages on rich growth medium containing glucose. To ensure that the APOBEC-induced mutation rate for the wild-type strain was the same on low-galactose and rich media, we measured the rate of *ura3* mutations (5-FOA-resistant derivatives) in the wild-type haploid SY10 (details in *SI Appendix*). The rates of *ura3* mutations were not significantly different in the two types of media, $7.1 \times 10^{-5} \pm 1.2 \times 10^{-5}$ (95% confidence limits) for cells grown in low-galactose medium and $6.3 \times 10^{-5} \pm 2.6 \times 10^{-5}$ for cells grown in rich nutrient media. Since we expected that the rate of mutations would be lower for SY46-WT than for the other strains, we grew this strain for additional passages to allow accumulation of similar numbers of mutations as in the low-polymerase strains.

From 7 to 10 independent colonies of each strain were sequenced to determine the frequency and position of APOBEC-induced mutations (Fig. 1). Sequence-diverged repeated genes (primarily related to retrotransposons, Dataset S1. "List of repeated genes excluded from primary sequence analysis.") were excluded from our analysis because short-sequence reads make it difficult to map new mutations to a specific element; in addition, it is difficult to distinguish de novo mutations from polymorphisms or gene conversion events between diverged elements.

As in previous studies, these diploid strains were generated by mating isogenic derivatives of the haploid strains W303-1A and YJM789 that differ by about 55,000 single-nucleotide polymorphisms (SNPs) (25). Since mitotic crossovers result in loss of heterozygosity (LOH) of polymorphisms that are centromere-distal to the crossover, DNA sequencing or SNP-specific microarrays can be used to identify the breakpoints of recombination events in these strains (9, 28). Thus, in our experiments, we could correlate the location of recombination events with the presence of APOBEC-induced mutations.

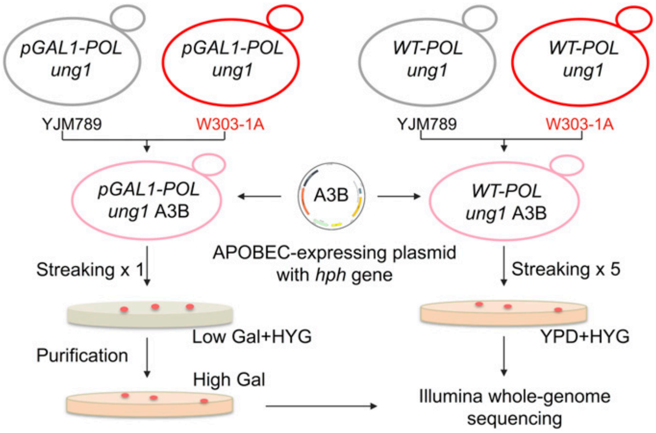

Fig. 1. Experimental design. The experimental diploid strains were homozygous for the *ung1* mutation and fusions of the *GAL1,10* promoter to genes encoding the catalytic subunits of DNA polymerases alpha (SY43-LA), epsilon (SY44-LE), and delta (SY45-LD). In SY46-WT, transcription of the DNA polymerase genes was regulated by their wild-type promoters.

Elevated Levels of APOBEC-Induced Mutations in Strains with Low Levels of Replicative DNA Polymerases.

Frequency of APOBEC-induced mutations. The types and locations of these mutations are in Dataset S2-1, and the numbers of mutations in each isolate of the four strains are in Dataset S2-2. The average number of mutations per isolate for strains subcultured once were: 2,400 (SY43-LA), 763 (SY44-LE), and 2,421 (SY45-LD). The number of mutations per isolate in the wild-type strain SY46-WT was 645; however, isolates of SY46-WT were passaged either 5 or 10 times instead of the single passage used for the other strains. We calculated the rate of mutations/base/cell division by dividing the number of mutations in each strain by the genome size times the number of isolates times the number of cell divisions; since an average colony has about 3×10^7 cells, the number of cell divisions necessary to create a colony is about 25. For SY46-WT, we multiplied the numbers of cell divisions by 5 or 10, depending on the number of passages. The rates of mutations in strains with low levels of DNA polymerases alpha, epsilon, and delta (normalized to the rate observed in the wild-type strain) were elevated 28-, 9-, and 28-fold, respectively (Fig. 2*A*). The simplest interpretation of these results is that the levels of single-stranded DNA in cells under replication stress are substantially increased. It should be pointed out that the genome-wide mutation rate in the wild-type isolates expressing A3B is about 1,000-fold greater than the spontaneous mutation rate in a wild-type strain in the absence of A3B (29); strains with low levels of DNA polymerase alpha and delta elevate single-base mutations 2-fold and 30-fold, respectively (10, 12), much smaller mutator phenotypes than those caused by expression of APOBEC.

The locations of mutations on the chromosomes of SY43-LA, SY44-LE, SY45-LD, and SY46-WT are in *SI Appendix*, Figs. S1–S4, respectively. As expected, in all strains, greater than 99% of the observed mutations are C-to-T alterations or G-to-A alterations (Dataset S2-2). Using a motif-finding program, we found

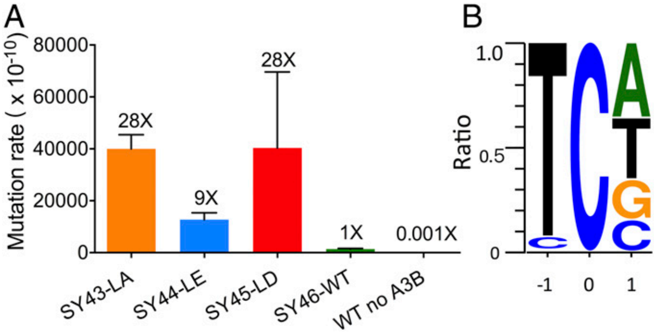

Fig. 2. Mutation rates and motifs of APOBEC-induced mutations. (*A*) Mutation rates in APOBEC-containing diploid strains. Rates were calculated based on the average number of mutations per isolate divided by the number of divisions in low-galactose medium. (*B*) Motif for APOBEC3B-induced mutations.

that 5′ tC characteristic of APOBEC3B mutagenesis accounted for more than 90% of the mutations (Fig. 2*B*).

Since our strains were diploid, most of the mutations were heterozygous. A small fraction of the new mutations were homozygous. These homozygous mutations were in regions of the chromosome with terminal or interstitial LOH events, and presumably were a consequence of A3B-induced mutations that occurred prior to an interhomolog recombination event. Based on the data in Dataset S6 ("Relationship between loss of heterozygosity and APOBEC-induced mutations."), the proportions of the genome within LOH events were: 0.01 (SY43-LA), 0.03 (SY44-LE), 0.02 (SY45-LD), and 0.002 (SY46-WT).

Strains with low levels of replicative DNA polymerases have elevated mutations on the lagging-strand template. The complementary strands of open reading frames (ORFs) are distinguishable in several ways: by replication (leading/lagging templates; LD/LG), by transcription (transcribed/nontranscribed; TR/NTR), and by transcript orientation relative to the closest replication origin (WL and WR, Watson orientation [*Saccharomyces* Genome Database] [SGD] located to the left or right of the closest origin, respectively; and CL and CR, Crick orientation located to the left or right of the closest origin, respectively). Based on the location and transcriptional orientation of the genes relative to efficient replication origins (5) and the specificity of APOBEC-induced modifications, we could examine the mutation frequency as affected by replication and by transcription separately.

Fig. 3*A* shows some details of this analysis for SY43-LA. In this figure, we show four genes, two located close to the origin on the left side (WL and CL) and two close to the origin on the right side (WR and CR); the dark blue arrows indicate their transcriptional orientation. The top and bottom strands of the duplex are depicted as in the SGD. As shown in Fig. 3*A*, the top strand of the genes located close to the left of the origin is the leading-strand template and the bottom strand is the lagging-strand template. Similarly, for sequences located close to the right side of the origin, the top and bottom strands are the lagging- and leading-strand templates, respectively. Lastly, we assume that A3B exclusively modifies C to U on single-stranded DNA. With these assumptions, all mutations can be assigned to one of eight classes of strands (WL/TR, WL/NTR, CL/TR, CL/NTR, WR/TR, WR/NTR, CR/TR, and CR/NTR) (Fig. 3*A*). Of the 2,289 mutations analyzed in SY43-LA, the observed numbers in each class are shown outside of parentheses. The expected numbers (shown in parentheses) are based on the numbers of the APOBEC tC motif (Ga on the opposite strand).

The eight classes of strands can be condensed into four groups: transcribed leading-strand templates (LD/TR), nontranscribed leading-strand templates (LD/NTR), transcribed lagging-strand templates (LG/TR), and nontranscribed lagging-strand templates (LG/NTR) (Dataset S3-1 "Distribution of APOBEC3B-induced mutations as regulated by lagging- and leading-strand DNA replication templates and by transcription."; Fig. 3*B*). We can also determine the frequencies of mutations of the lagging-strand relative to the leading-strand templates, and the transcribed versus the nontranscribed strand. From this analysis, we found that SY43-LA had about three times more mutations on the lagging-strand template than the leading-strand template, and about 50% more mutations on the transcribed strand than the nontranscribed strand (Dataset S3-1). Both of these differences were significant with P values <0.0001.

The distributions of mutations between leading- and lagging-strand templates and between transcribed and nontranscribed strands were examined in two different ways. In one analysis (shown in Fig. 3*A*), we examined only the two genes that were closest to the origins. In the second analysis, we considered the mutations in all genes, assigning each gene to the closest origin (Dataset S3-1). Although there were small quantitative differences

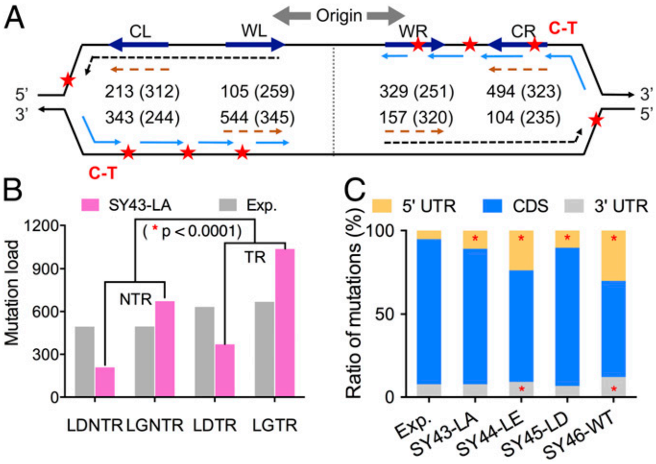

Fig. 3. Strand-specific APOBEC-induced mutations and frequencies of mutations within ORFs and in sequences flanking ORFs. (*A*) Numbers of strand-specific mutations in SY43-LA. A replicating DNA molecule is shown with the expected pattern of leading- and lagging-strand templates. The thin black dashed lines show the replicating leading strands with arrows indicating the 3′ ends. The short blue lines represent synthesis on the lagging strand. Red stars show the location of APOBEC-induced mutations on single-stranded DNA. As indicated, a C-to-T mutation on the upper DNA strand in the region to the right of the origin reflects a mutation on the lagging strand, whereas in sequences located to the left of the origin, the C-to-T mutation would result in a G-to-A mutation on the upper strand. The thick blue arrows show genes located in the Watson (W) or Crick (C) transcriptional orientations (based on the *Saccharomyces* Genome Database) to the left (L) or right (R) of the replication origin. The dotted orange lines indicate transcripts. The numbers outside of parentheses show APOBEC-induced mutations on the upper or lower strands (data in Dataset S3-1) in SY43-LA isolates; in parentheses, we show the expected numbers of mutations based on the numbers of tC motifs in each strand of each class of gene (Dataset S3-1). (*B*) Numbers of mutations as a function of location relative to flanking replication origins in SY43-LA. Based on the numbers of mutations shown for each class of strands in A and Dataset S3-1, we determined the observed (shown in pink) and expected (shown in gray) numbers of mutations in the leading nontranscribed strand (LD NTR), the lagging nontranscribed strand (LG NTR), the leading transcribed strand (LD TR), and the lagging transcribed strand. (*C*) Ratio of mutations in 5′ UTRs, ORFs, and 3′ UTRs. We show the observed and expected proportions of mutations in the region 5′ to the coding sequence, within the coding sequence (CDS), and in the region 3′ to the coding sequence. Red asterisks indicate a significant departure ($P < 0.0001$) from the expectation based on χ^2 analysis.

in the distributions, the patterns were similar for both types of comparisons.

The same analyses were also done for the other strains (Dataset S3-1). When we examined the genes located closest to the origins, we found the following biases for mutations on the lagging-strand templates (mutations on the lagging-strand template/mutations on the leading-strand template): 3.0 for SY43-LA ($P < 0.0001$), 1.3 for SY44-LE ($P = 0.002$), 4.3 for SY45-LD ($P < 0.0001$), and 2.0 for SY46-WT ($P < 0.0001$). The biases in favor of mutations on the transcribed strand were: 1.6 for SY43-LA ($P < 0.0001$), 1.2 for SY44-LE ($P = 0.02$), 2.2 for SY45-LD ($P < 0.0001$), and 1.2 for SY46-WT ($P = 0.2$). Previously, Di Noia and Neuberger (30) concluded that the transcribed and nontranscribed strands were equally susceptible to activation-induced deaminase (AID) during somatic hypermutation in mammalian cells. The amount of single-stranded DNA at the replication forks under these conditions is likely much lower than under the conditions of replication stress in our experiments.

When mutations were assigned to intervals located between origins, there is a strong bias for mutations in the lagging-strand template for SY43-LA (threefold) and SY45-LD (sixfold), and weaker biases for SY44-LE (1.4-fold) and SY46-WT (twofold) (Fig. 4). The twofold bias in SY46-WT is similar to that observed in a wild-type yeast strain expressing APOBEC3B previously (21).

In agreement with the results of Lada et al. (31), we find that the 5′ untranslated region (UTR) (average of 83 bp upstream of the initiation codon) had a significantly greater density of APOBEC-induced mutations in all four strains ($P < 0.0001$) (Fig. 3*C*); in the different strains, the ratios of observed:expected mutations varied between about 2 and 4 (Dataset S3-2 “Density of mutations in ORFs and flanking intergenic regions.”). In all four strains, the density of mutations within the coding sequence is significantly less than expected from a random distribution. Lastly, the density of mutations in the 3′ untranslated region of the gene (average of 145 bp) is more variable (Fig. 3*C*). The sequence coordinates used for this analysis are in Dataset S3-5 (5′ UTRs) and Dataset S3-6 (3′ UTRs).

In all strains, the density of mutations was between two- and threefold higher in the 500-bp regions located upstream of the top 300 highly expressed genes than for all other genes ($P \leq 0.0001$) (Dataset S3-3 “Effect of transcription on mutation rates.”). In contrast, the density of mutations of the 363 genes with no detectable expression was significantly reduced compared to all other genes ($P \leq 0.001$). The data on gene expression were derived from SI appendix, table S1 of Rong-Mullins et al. (32); the genes and expression levels used in our comparisons are in Dataset S3-4 (“Expression level of genes used in analysis.”).

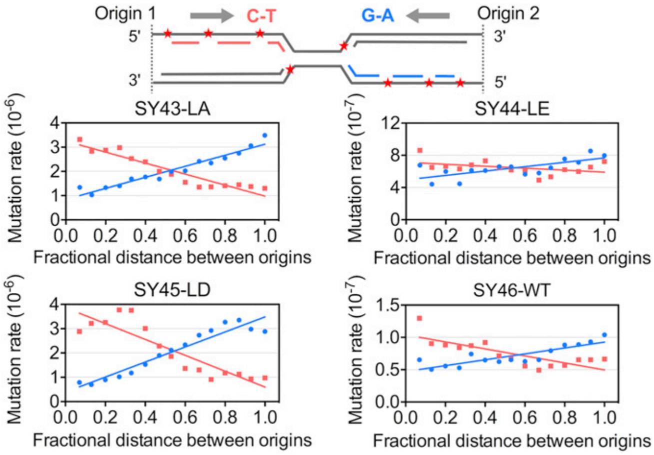

Fig. 4. APOBEC-induced mutation rates as a function of the distance from the replication origins. For this analysis, we divided each interorigin distance into 15 intervals and determined the numbers of C-to-T and G-to-A mutations on the top strand within these intervals. This approach was used previously by Hoopes et al. (21).

APOBEC-induced mutations in the tRNA and ribosomal RNA (rRNA) genes. In agreement with previous studies (22, 33), we found that tRNA genes are a preferred substrate for APOBEC. Based on the rates of mutations/bp/cell division in all other yeast genes (Dataset S2-2), the number of mutations in tRNA genes (Dataset S3-1), and the number of bases encoding tRNA genes in the haploid genome (21,939), we calculated the rates of A3B-induced mutations in tRNA. These rates are (normalized to the wild-type rate in parentheses): 6.7×10^{-5} (10), 4.9×10^{-5} (7), 5.5×10^{-5} (8), and 7.0×10^{-6} (1) for strains SY43–SY46, respectively. Relative to the mutation rates for non-tRNA genomic sequences in the same strain, the rates in the tRNA genes were elevated 17-fold (SY43-LA), 42-fold (SY44-LE), 14-fold (SY45-LD), and 52-fold (SY46-WT). In addition, as observed by others (22), the mutations occur preferentially on the nontranscribed strand. In all four strains, there were about 10-fold more mutations on the nontranscribed strand than the transcribed strand (Fig. 5 *A* and *B*).

From the sequencing coverage of ribosomal DNA compared to the coverage of single-copy sequences (Dataset S7-1), we calculated the number of rRNA genes per cluster per isolate as 63 (SY43-LA), 59 (SY44-LE), 75 (SY45-LD), and 160 (SY46-WT). It has been noted previously that yeast strains under replication stress have a smaller number of rRNA genes than wild-type strains (12, 34, 35). The A3B-induced mutations in the rDNA are listed in Dataset S7-3 and the locations of the mutations within the 9.1-kb repeats are shown in Fig. 5*C*. Fig. 5*D* shows the mutation rates (bp/cell division) for transcribed and nontranscribed strands for the 35S, 5S, and nontranscribed spacers (sum of NTS1-1 and NTS1-2). The rates of mutations (bp/cell division) averaged over the entire repeat and the rates normalized to the wild-type strain (shown in parentheses) were: 1.1×10^{-5} (64) for SY43-LA, 5.7×10^{-6} (35) for SY44-LE, 6.7×10^{-6} (41) for SY45-LD, and 1.6×10^{-7} for SY46-WT (Dataset S7-2 “Number of APOBEC3B-induced mutations within the rRNA genes.”). We also examined the rate of mutations in each strand of the various segments of the rDNA gene (Fig. 5*D* and Dataset S7-2).

APOBEC-induced mutations are associated with replication-termination regions, regions with high-GC content, and other chromosome motifs known to slow DNA replication forks. In our previous studies, we found that breakpoints for mitotic recombination events in strains with low-alpha or low-delta polymerase were enriched for motifs (such as quadruplex structures) associated with slow-moving DNA replication forks (11, 12). We examined the SY43-LA–SY46-WT strains for enrichment for these motifs (Dataset S2-3 “Association of various chromosome motifs with APOBEC-induced mutations.”). In general, SY43-LA and SY45-LD had similar patterns of APOBEC-induced enrichments. Significant enrichments were observed for replication–termination sequences, regions of high-GC content, meiotic recombination hotspots, noncoding RNA genes, tRNA genes, and Rrm3p binding sites. For all comparisons, we corrected the expected number of events to account for the numbers of the tC motif in each element of interest. The association of high levels of APOBEC-induced mutations at replication-termination sequences and tRNA genes is likely because these sequences stall replication forks (36). The association with high-GC sequences may be related to our previous observation that such regions elevate the rate of mitotic recombination and mutagenesis (37). The association with meiotic recombination hotspots, which are GC-rich regions in yeast (38), may represent the same effect. High-GC regions, meiotic recombination hotspots, and tRNA genes were also enriched for APOBEC-induced mutations in SY44-LE and SY46-WT. Lastly, using the information in SGD (https://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/), we examined whether the mutation

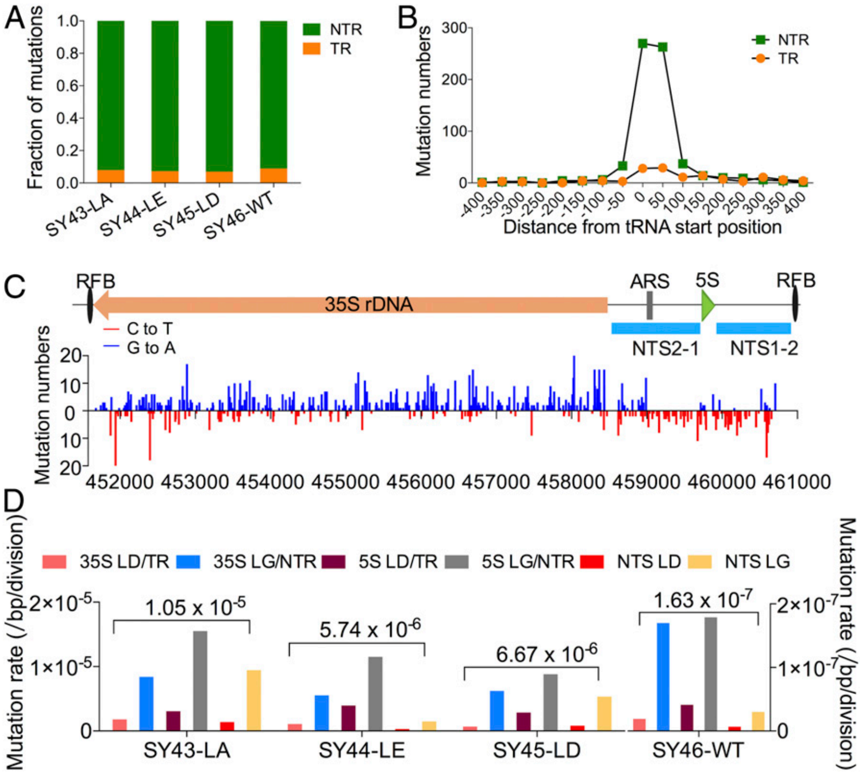

Fig. 5. APOBEC-induced mutations in tRNA and ribosomal RNA genes. (*A*) Elevated levels of mutations in nontranscribed strand of tRNA genes. (*B*) Location of mutations within tRNA genes in SY43-LA. The numbers of mutations in 50-bp increments upstream, within, and downstream of tRNA genes are shown for both the transcribed (TR) and nontranscribed (NTR) strands. (*C*) Distribution of mutations within 9-kb rDNA repeats. RFB shows the position of the replication fork block, and NTS regions are nontranscribed spacers. Red and blue represent C-to-T and G-to-A mutations, respectively. Most of the G-to-A changes occur to the left between the *ARS* element and the RFB, as expected if they represent mutations on the lagging strand; similarly, the C-to-T mutations occur between the *ARS* and the RFB on the right side, as expected for mutations on the lagging strand. (*D*) Rate of mutations in different regions of rDNA repeats. Based on the data in Dataset S7, we calculated the approximate mutation rates for strand-specific mutations in different segments of the rDNA repeat.

rates within introns were different from the whole-genome rate. The intron rates (normalized to the whole-genome rate of 1) were similar to the rates for the whole genome: 1.0 (SY43-LA), 1.4 (SY44-LE), 0.9 (SY45-LD), and 1.6 (SY46-WT). In summary, except for the enrichment of APOBEC mutations in the tRNA genes, most of the significant enrichments or depletion are threefold effects or smaller.

Mitotic Recombination Events and Chromosome Alterations Induced by Low Levels of Replicative DNA Polymerases. We previously showed that low levels of replicative DNA polymerases elevated the rates of chromosome rearrangements (9). These events were detected by looking at LOH for homolog-specific SNPs (12). The common classes of LOH events observed in the current experiments were interstitial LOH (gene conversions), terminal LOH events (crossovers or break-induced replication [BIR] events), and deletions/duplications (Dataset S6 "Expected and observed numbers of APOBEC-induced mutations at LOH breakpoints."; *SI Appendix*, Fig. S5). The location of the LOH events and the depiction of the various classes of LOH events are in Datasets S6-1 and S6-2, respectively. As expected from our previous studies (11, 12), we also observed an elevated number of whole-chromosome changes (Dataset S6-3).

We examined the breakpoints of interstitial and terminal LOH events (details in *Materials and Methods* and *SI Appendix*) to determine whether these regions were associated with elevated levels of A3B-induced mutations (Datasets S6-4 and S6-5). In the strains SY44-LE and SY45-LD, there were approximately 2-fold more mutations than expected, indicating a nonrandom association between APOBEC-induced mutations and the breakpoints of LOH events. However, it should be noted that many LOH breakpoints had no associated APOBEC-induced mutations, and the number of mutations associated with LOH breakpoints was a small fraction (0.01 or less) of the total mutations. Lastly, we found that one of the SY43-LA isolates (SY43-LA-6) had a level of interstitial LOH that was about 10-fold higher than the average of the other isolates. This isolate was excluded from the LOH analysis described above. It is likely that an A3B-induced mutation in SY43-LA-6 within a replication protein resulted in elevated levels of DSBs and, consequently, a higher rate of interstitial LOH events.

APOBEC-Induced Mutations Are Often Clustered. In yeast, as in mammalian cells, APOBEC-induced mutations are often clustered at a much higher density than expected from a random distribution (16). This clustering likely reflects two factors: the infrequent formation of long and/or persistent regions of single-stranded DNA in the predominately double-stranded yeast genome, and the processivity of APOBEC on single-stranded DNA (39). Long stretches of single-stranded DNA could be formed during replication (as discussed above), R-loop formation, or by exonucleolytic processing of DSBs (16). We used two different methods to assign mutations to clusters, one utilizing a Poisson distribution (P-method) and one using a negative binomial distribution (B-method) (18). The details of each method are briefly described below with additional details in *SI Appendix*.

Analysis of clusters using the P-method. This method was designed to look for significant clusters of A3B-induced mutations within one strand (Watson or Crick). For this method, we divided the genome into adjacent nonoverlapping 5-kb intervals. Based on the number of mutations of the same type (C to T or G to A) observed in each isolate and the size of the genome, we calculated the number of mutations expected for these intervals, assuming a Poisson distribution. The calculations to determine which 5-kb intervals had significantly more mutations than expected for a random distribution are summarized in Dataset S4 ("Basic parameters of analysis of mutation clusters within 5-kb windows."). Datasets S4-1–S4-3 outline the calculation used to assign mutational clusters. For this analysis, clusters of C-to-T and G-to-A mutations were considered separately. The 5-kb intervals with significantly more mutations of C to T (or G to A) than expected for a random distribution of these types of mutations were considered a significant cluster. If two or more adjacent 5-kb intervals had significant clusters of the same type of mutation, these intervals were considered a single cluster.

The 5-kb intervals with significantly high levels of mutations are shown in Datasets S4-4–S4-7. Based on Datasets S5-1 and S5-3, the median numbers of clusters per isolate were: 59 (SY43-LA), 9.5 (SY44-LE), 45 (SY45-LD), and 3.5 (SY46-WT). The total numbers of clusters for each strain were: 463 (SY43-LA), 115 (SY44-LE), 436 (SY45-LD), and 44 (SY46-WT). The percentages of A3B-induced mutations that were in these clusters were (Dataset S5-7): 19% (SY43-LA), 7% (SY44-LE), 41% (SY45-LD), and 2.5% (SY46-WT). It is noteworthy that, although SY43-LA and SY45-LD had similar numbers of clusters and a similar number of genomic mutations, SY45-LD had twice as many mutations within the clusters as SY43-LA.

As described above, we assigned clusters based on a significant excess of G-to-A mutations or a significant excess of C-to-T changes within the 5-kb window (Dataset S5-2 "List of clusters in all isolates identified by the P-method."). Although most (>70% in all strains) of these clusters were "pure" (having only one type of change), some had one or more different mutations. In these mixed clusters, generally, one type of mutation predominated. For example, in SY43-LA-2, there was a cluster located on chromosome I (cluster ID 92) that contained 11 G-to-A mutations and one C to T (Dataset S5-2). One likely interpretation of mixed clusters (discussed further below) is that they represent two independent mutagenic events.

From these data, we calculated average, minimum, and maximum numbers of mutations per cluster as: 7.1/4/69 (SY43-LA), 4.9/3/17 (SY44-LE), 16/3/105 (SY45-LD), and 3.7/3/6 (SY46-WT). The main differences are that low-polymerase delta results in clusters that include more mutations than observed in low-polymerase alpha or low-polymerase epsilon, and that the wild-type strain has substantially fewer mutations per cluster than any of the low-polymerase strains.

The average lengths of the clusters (calculated from Dataset S5-2) in the three low-polymerase strains were similar: 3,786 bp (SY43-LA), 3,320 bp (SY44-LE), and 4,648 bp (SY45-LD). These values likely represent a minimal estimate, since mutations that extend a cluster beyond one 5-kb interval may not have a significant concentration of mutations in an adjacent 5-kb interval to be counted as a statistically significant cluster. In the low-polymerase strains, the sizes of the clusters had a very wide range, varying from <60 bp to >20 kb. In the wild-type strain, the average cluster is relatively small (1,557 bp), and the range of cluster sizes (15 to 4,300 bp) is also narrower than for the low-polymerase strains. In general, the clusters are much larger than the size of an alpha DNA polymerase-generated primer (35 to 50 bp) or an Okazaki fragment (150 to 400 bp) (40).

The number of large (≥16 mutations) clusters in SY45-LD relative to the other strains is particularly striking. The 172 large clusters in the four strains (Dataset S5-5 "Mapping long clusters to leading- or lagging-strand templates.") were distributed as follows: 22 (SY43-LA), 3 (SY44-LE), 147 (SY45-LD), and 0 (SY46-WT). It should be noted, however, that the number of large clusters in SY45-LD is strongly affected by the large number of large clusters in two isolates, SY45-LD-4 and SY45-LD-5. Large clusters were concentrated on the lagging-strand template (Dataset S5-5).

We also calculated whether there was a significant excess of clusters that include at least two adjacent 5-kb "hot" samples. Based on the number of 5-kb hot intervals (either C-to-T or G-to-A mutations) in each isolate, we used a Monte Carlo simulation to predict the probabilities of one or more pairs of adjacent hot 5-kb segments in each isolate. Based on these calculations and correcting for multiple comparisons, we found that 5 of 7 isolates of SY43-LA, 5 of 10 isolates of SY44-LE, and 4 of 7 isolates of SY45-LD had significant overrepresentations of adjacent C-to-T or G-to-A (or both) pairs of hot 5-kb regions (Dataset S5-3). None of the isolates of SY46-WT had significantly elevated pairs of hot 5-kb intervals.

Large mutational clusters could result from events produced by APOBEC-induced modifications on a large single-stranded region in one cell cycle or represent multiple events produced in different cell cycles. Two types of analyses were done to address this issue (details in *SI Appendix*). First, we tested whether large clusters in one isolate had an overrepresentation of mutational clusters in the same genomic region in other isolates of the same strain. No significant tendency of this type was observed for SY43-LA, SY44-LE, or SY46-WT, although SY45-LD had significant overlaps of mutational clusters in multiple isolates. Second, we examined whether the mutational clusters were associated with a single homolog or were found on two different homologs. This analysis was done by determining whether the APOBEC-induced mutations were linked to homolog-specific SNPs. Of 33 large clusters examined, 21 had a significant excess of mutations from a single homolog. These numbers, however, are not significantly different from those expected if the large clusters were generated in two separate cell cycles, as discussed in *SI Appendix*.

Analysis of clusters using the B-method. In previous studies of clusters of APOBEC-induced mutations, a different method was used to determine the significance of clusters (18). In brief, based on the number of mutations in the genome of the isolate, the probability of observing a certain number of mutations within a certain number of base pairs was calculated using a negative binomial distribution. Probabilities were calculated for all groups of mutations with an intermutational distance below an arbitrary threshold regardless of whether the mutation was a C-to-T or a G-to-A change. For these analyses, we used intermutational distance thresholds of 2, 5, or 10 kb. In Datasets S5-1, S5-7, and S5-8, we show the clusters identified by the B-method and compare the clusters obtained by the two methods. The relative numbers of clusters identified in all isolates are highest for the 2-kb analysis, intermediate for the 5-kb analysis, and lowest for the 10-kb analysis (Dataset S5-7). This result is expected since more than one of the 2-kb clusters are often included within the 10-kb clusters. The fraction of mutations within the clusters is similar for the 2-kb, 5-kb, and 10-kb clusters within each genotype, although SY45-LD includes about twice as many mutations in clusters as the other strains (Dataset S5-7).

Despite the differences of the P- and B-methods, there is considerable overlap between the clusters identified by the two methods. In all strains, the number of mutations included within clusters by the P-method is smaller than that included within clusters by the B-method. In part, this difference is because the P-method includes closely linked C-to-T and G-to-A mutations as part of the same cluster. Despite these differences, the locations of the clusters obtained by the different methods are in good agreement (*SI Appendix*, Fig. S6). For example, 74% of the P-method clusters overlap with the 10-kb B-method clusters, and 67% of the 10-kb B-method clusters overlap with the P-method clusters (Dataset S5-8).

Strand-switch clusters (analyzed by the P-method). As described previously, many of the 5-kb intervals that were significantly enriched for A3B-induced mutations were immediately adjacent to other 5-kb intervals that were significantly hot for the same type of mutation. For example, in SY43-LA, of 564 hot 5-kb regions, 166 (29%) were adjacent to other hot 5-kb regions. For SY44-LE and SY45-LD, the comparable percentages were 29% (40 of 138) and 49% (303 of 613), respectively. In contrast to these strand-coordinated clusters, in all of the strains analyzed, 5-kb regions with one type of mutation (for example, G to A) were very rarely adjacent to 5-kb regions with significant numbers of the other type of mutation (C to T). Only 16 examples were observed for the whole dataset, 8 from SY43-LA and 8 from SY45-LD.

There are a number of possible sources of strand switches. One source is the result of 5′ to 3′ exonuclease-mediated processing of broken DNA ends resulting from a DSB (16). One common type of mitotic recombination is gene conversion, and many conversion events unassociated with crossovers are a consequence of synthesis-dependent strand annealing (SDSA) (13). One-ended SDSA events can produce a strand switch in which a tract of C-to-T changes is adjacent to a tract of G-to-A changes (Fig. 6*A*); note that, by this mechanism, the C-to-T tract is always to the left of the G-to-A tract on the top (Watson strand) (24, 27). By this criterion, 11 of the 16 strand switches were not consistent with the repair of a DSB.

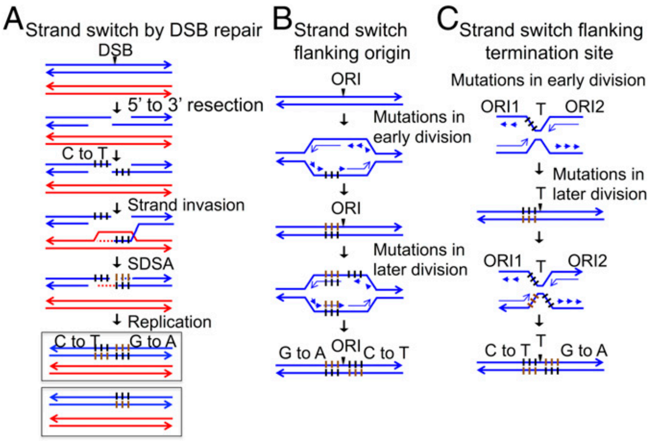
A
Strand switch by DSB repair
DSB
5′ to 3′ resection
C to T
Strand invasion
SDSA
Replication
C to T
G to A
B
Strand switch flanking origin
ORI
Mutations in early division
ORI
Mutations in later division
G to A
ORI
C to T
C
Strand switch flanking termination site
Mutations in early division
ORI1
T
ORI2
Mutations in later division
T
ORI1
T
ORI2
C to T
T
G to A

Fig. 6. Mechanisms that could give rise to a mutational cluster with a strand switch. Although most clusters had primarily G-to-A or C-to-T changes, some clusters had a tract of G-to-A mutations adjacent to a tract of C-to-T mutations (strand-switch clusters). Several mechanisms could give rise to such clusters (16). (*A*) Strand switch by repair of a DSB. Two chromatids are shown as paired blue and red lines. A DSB results in two ends that are processed by 5′ to 3′ resection (13). The single-stranded regions are subsequently modified by APOBEC to generate C-to-T alterations. Following strand invasion, DNA synthesis is primed from the invading end. The invaded end is extruded and reanneals to the other broken end (SDSA). Replication of the DNA molecule with the APOBEC-induced mutations would result in a strand-switch cluster with the C-to-T mutations shown as short black lines and the G-to-A mutations as short brown lines. (*B*) Strand switch by mutations flanking replication origin. Mutations are introduced in one cell cycle on the lagging-strand template to the left of the origin, resulting in G-to-A mutations on the upper strand. In a subsequent cell cycle, mutations are introduced on the lagging-strand template to the right of the origin, resulting in a cluster with a strand switch. (*C*) Strand switch flanking termination region. Mutations occur on the lagging-strand template to the left side of the termination region (T) in one cell cycle and on the lagging-strand template to the right of T in a subsequent cell cycle.

Alternatively, strand switches could be produced by events that occur in nearby regions in different cell cycles by several different mechanisms including: 1) mutation of the lagging-strand templates from an origin located between the adjacent tracts (Fig. 6*B*), 2) mutation of the lagging-strand templates in replication forks converging on a termination site (Fig. 6*C*), and 3) mutation of the lagging-strand template in one cell cycle next to a region mutated on the leading-strand template in a different cell cycle.

Strand switches that occur as a consequence of DSB repair should occur during one cell division and involve only one homolog, whereas the other mechanisms occurring in separate cell divisions could occur on one or two homologs. Using the same type of analysis that we employed to determine whether large mutational clusters were on one or two homologs, we found that one strand switch was on one homolog and six involved two homologs (Dataset S5-6 "Strand switches between clusters."). For nine strand-switch events, we could not determine whether one or two homologs were involved because too few of the APOBEC-induced polymorphisms were linked to heterozygous markers or because the mutations were in regions of LOH. Incorporating this information, we conclude that only 4 of 16 strand-switch events could be a consequence of DSB repair following symmetric bidirectional resection of a DSB.

Discussion

The results described above lead to the following conclusions: 1) Low levels of replicative DNA polymerases result in substantially elevated levels of single-stranded DNA; in strains with low levels of DNA polymerase alpha and delta, most of the mutations are introduced into the lagging-strand template, whereas in strains with low levels of DNA polymerase epsilon, mutations are elevated to similar extents on the leading- and lagging-strand templates. 2) Most of the APOBEC-induced mutations are associated with single-stranded regions generated during DNA replication rather than single-stranded regions associated with symmetric bidirectional processing of a DSB. 3) Some APOBEC-induced mutations are in strand-coordinated clusters exceeding 10 kb in size. 4) In tRNA genes, the nontranscribed strand is preferentially mutated (10-fold preference), but in most yeast genes, there is a significant small preference (about 1.5-fold) for mutating the transcribed strand.

Elevated Levels of Single-Stranded DNA in Strains with Low Levels of Replicative DNA Polymerases. In the wild-type strain expressing APOBEC (SY46-WT), the rate of mutations on the lagging-strand template is about 0.9 × 10^{-7}/bp/cell division and about 0.5 × 10^{-7}/bp/cell division on the leading-strand template. In SY43-LA, we found rates of mutation of 2.9 × 10^{-6} (32-fold increase relative to SY46-WT) and 1 × 10^{-6} (20-fold increase) on the lagging- and leading-strand templates, respectively. By a similar calculation, in SY45-LD, we found rates of mutation of 3.3 × 10^{-6} (37-fold increase relative to SY46-WT) and 0.6 × 10^{-6} (12-fold increase) on the lagging- and leading-strand templates, respectively.

In preliminary experiments (described in *SI Appendix*), we found that strains grown under high-galactose conditions also had elevated levels of APOBEC-induced mutations. Although asynchronous cells grown in high-galactose medium have about fivefold more DNA polymerase alpha and delta than observed in wild-type strains (10, 27), the level of DNA polymerase alpha during the S period is less than observed in wild-type strains (41). Consequently, it is not clear whether the mutations induced in strains grown in high levels of galactose reflect too much or too little of the replicative polymerases.

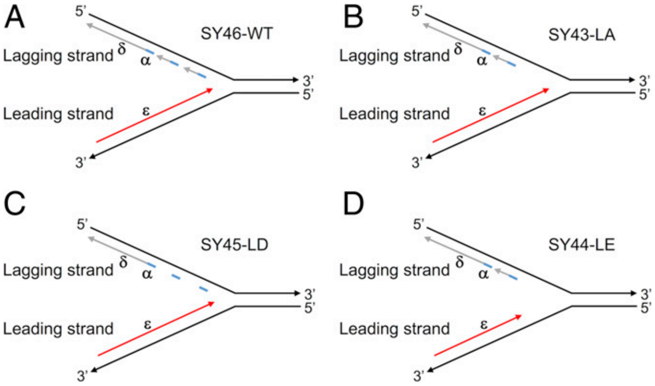

Fig. 7. Expected patterns of single-stranded DNA at replication forks in strains with low levels of replicative DNA polymerases. The template strands are shown in black, the Okazaki fragments introduced by DNA polymerase alpha are shown in blue, the sequences produced by DNA polymerase delta are shown in gray, and the sequences synthesized by DNA polymerase epsilon are in red. (*A*) Wild-type strain SY46-WT. In this strain, because synthesis of the leading- and lagging-strand templates is coordinated, the amount of single-stranded DNA is limited. Most of the single-stranded DNA is on the lagging-strand template. (*B*) Strain with low levels of DNA polymerase alpha (SY43-LA). Because of the low levels of DNA polymerase alpha, the density of Okazaki fragments is reduced, leading to an elevated level of single-stranded DNA on the lagging-strand template. (*C*) Strain with low levels of DNA polymerase delta (SY45-LD). A reduction of DNA polymerase delta results in less efficient extension of Okazaki fragments and an elevated level of single-stranded DNA on the lagging-strand template. (*D*) Strain with low levels of DNA polymerase epsilon (SY44-LE). The data indicate elevated levels of single-stranded DNA on both template strands. One possibility is that low levels of DNA polymerase epsilon result in increased recruitment of DNA polymerase delta to the leading-strand template (7, 8). This "unnatural" replication fork may have less coordination of the synthesis of leading and lagging strands and more single-stranded DNA on both templates.

As expected from current models of DNA replication (Fig. 7*A*), the lagging-strand template would be expected to have more single-stranded DNA than the leading strand; this tendency has been observed previously in wild-type strains (21), as well as in our studies of a wild-type strain (SY46-WT). The increase in single-stranded DNA, resulting in elevation of APOBEC-induced mutations on the lagging strand in SY43-LA, is expected, since a reduction in alpha DNA polymerase would be expected to reduce the number of primers on the lagging strand (Fig. 7*B*). In SY45-LD, the low levels of DNA polymerase delta in SY45-LD would be expected to result in an increased number of single-stranded regions on the lagging-strand template because of a delay in the extension of Okazaki fragments by DNA polymerase delta (Fig. 7*C*).

The elevated APOBEC-induced mutagenesis on the leading-strand template in SY43-LA and SY45-LD is more difficult to explain. In *Escherichia coli*, phage, and eukaryotes, synthesis of the leading- and lagging-strand templates is coupled (42, 43). Thus, delayed synthesis on the lagging-strand template might increase the level of single strandedness on the leading-strand template. In addition, as described previously DNA polymerase delta can replicate the leading strand in conditions in which synthesis by DNA polymerase epsilon is perturbed (8). This unnatural replication fork could result in an increase of single-stranded DNA on both the leading- and lagging-strand templates (Fig. 7*D*). Lastly, our analysis is based on the assumption that the probability of initiating replication from an origin is not substantially affected by low levels of the replicative DNA polymerases. Since Porcella et al. (41) found that low levels of alpha DNA polymerase led to significantly reduced origin efficiency, however, we cannot exclude the possibility that we are overestimating the frequency of mutations on the leading strand in SY43-LA.

Our conclusions about single-stranded regions on the leading and lagging strands based on APOBEC-induced mutations is supported by a striking agreement between the frequency of mutations along the chromosome and the mapping of Okazaki fragments. In *SI Appendix*, Fig. S7*A*, we show the mapping of Okazaki fragments observed by McGuffee et al. on yeast chromosome X (5); *SI Appendix*, Fig. S7*B* shows our mapping of APOBEC-induced mutations in SY43-LA.

Most of the APOBEC-Induced Mutations Are Associated with Replication-Associated Single-Stranded Regions Rather than Single-Stranded Regions Associated with Symmetric Bidirectional Resection of DSB-Associated Ends. The bias in favor of mutations on the lagging-strand template observed in strains with low levels of DNA polymerases alpha and delta is strongest in genomic regions located close to the replication origin (Fig. 4). This pattern argues that the target of APOBEC-induced mutagenesis primarily is single-stranded DNA generated during DNA replication. However, since strains with low levels of DNA polymerases alpha and delta also have high rates of mitotic recombination (9), another potential source of single-stranded DNA is a consequence of processing and repair of DSBs. Symmetric processing of two broken ends followed by APOBEC-induced modifications and repair by the SDSA pathway would be expected to result in clusters with a strand switch between C to T on the left side of the break and G-to-A clusters on the right side (Fig. 6*A*). Very few such clusters were observed in our analysis. In addition, most of the LOH regions observed in our study were not associated with elevated levels of APOBEC-induced mutations (Dataset S6). It should also be noted, however, repair of DSBs by BIR is often associated with clusters of APOBEC3A-induced mutations (25). Thus, the mutational clusters detected in our experiments may reflect a variety of mechanisms.

In examining the patterns of APOBEC-induced mutations in strains with high levels of DSBs produced by gamma rays, Sakofsky et al. (24) showed that mutational clusters with a single switch (the pattern expected from symmetric bidirectional processing and repair of a DSB) represented a third or less of the clusters. To explain the high frequency of strand-coordinated clusters, they suggested that such clusters reflected either long unidirectional resection or BIR. Although, in our previous studies of mitotic recombination in wild-type diploid cells we observed that reciprocal crossovers between homologs were three to four times more common than BIR events (44), the ratio of crossovers and BIR events involving sister chromatids was not determined. Thus, in our experiments, we cannot determine whether the clustered mutations are a consequence of unidirectional resection of a DSB followed by homologous recombination or BIR.

APOBEC-Induced Mutations Are Clustered with Some Clusters Exceeding 10 kb in Size. Clusters of APOBEC-induced mutations greater than 10 kb in size generated in a single cell cycle were observed in gamma-irradiated strains (24). In SY43-LA and SY45-LD, we observed totals of 144 clusters ≥7.5 kb in size and 56 ≥10 kb in size (Dataset S5-2, calculated by the P-method); no clusters of this length were observed in SY46-WT, and only 13 clusters ≥7.5 kb were found in SY44-LE. As discussed above, such clusters could be produced by APOBEC modification of very long single-stranded regions in one cell cycle or by two or more APOBEC modifications of the same strand in the same region in different cell cycles. We used two tests to determine whether large clusters were a consequence of multiple events or single events. One test (details in *SI Appendix*, *Identifying APOBEC-Induced Mutational Clusters*) was to determine whether regions with large clusters in one isolate also had hotspots at the same position in other isolates, as expected if the region was prone to APOBEC-induced mutations. In SY43-LA, no such tendency was observed, although in SY45-LD, we found a significant overrepresentation ($P < 0.001$) of regions that were hot in multiple isolates (discussed in *SI Appendix*); of the 90 pairs of 5-kb regions with clusters of APOBEC-induced mutations, 40 were found in more than one isolate. These results suggest that low levels of DNA polymerase delta result in more extensive single-stranded DNA in some chromosome regions than in others.

We also determined whether the strains with large numbers of mutations were derived from one or both homologs by analyzing the linkage of the APOBEC-induced mutations to homolog-specific SNPs. In 21 of 33 clusters examined, most of the mutations were derived from one homolog. However, this deviation from 50:50 is not significant ($P = 0.16$). In summary, we cannot rule out the possibility that the long clusters of APOBEC-induced mutations are a consequence of the modification of closely linked regions of the genome in different cell cycles rather than APOBEC-induced mutations that occurred in one cell cycle.

In tRNA Genes, the Nontranscribed Strand Is Preferentially Mutated (10-Fold Preference), but in Most Yeast Genes, There Is a Significant Small Preference (about 1.5-Fold) for Mutating the Transcribed Strand. As observed previously (22, 33), in our study, the tRNA genes are a preferred target for APOBEC-induced mutagenesis, and the nontranscribed strand is mutated about 10-fold more frequently than the transcribed strand. Several factors may influence the high rate of tRNA mutations. First, tRNA genes have high rates of R-loop formation (45) that would result in single-stranded DNA on the nontranscribed strand. In support of this model, yeast strains with mutations in two genes involved in removal of R loops (*rnh1 rn201*) have twofold elevated rates of APOBEC-induced mutations relative to a wild-type strain expressing APOBEC (22). A second factor is that secondary structures formed by single-stranded tRNA genes may be particularly prone to mutations, primarily within the loops of hairpin structures. Third, tRNA genes in yeast are associated with an open chromatin structure (46) which may allow increased access to APOBEC. Lastly, it is possible that the proteins associated with RNA polymerase III interact with APOBEC directly to facilitate mutagenesis.

If the high rate of mutations in tRNA genes is related to its transcription by RNA polymerase III, the 5S rRNA genes should have a similarly high rate of mutations. The mutation rates/bp in the 9.1-kb rRNA gene repeat in each strain are shown in Dataset S7-2 and Fig. 5*D*. As expected, the rates of mutations in the rRNA genes are elevated substantially in strains with low DNA polymerases relative to the wild-type strain SY46-WT (fold elevation shown in parentheses): SY43-LA (64), SY44-LE (35), and SY45-LD (41).

The rate of mutations in the rRNA genes is only slightly higher (two- to fourfold) than the rates in the single-copy genomic sequences in the same strain. Similarly, the rates of mutation of the 35S region of the rRNA gene (transcribed by RNA polymerase I) are only slightly elevated relative to single-copy genes in the same strain (comparison of Datasets S2-2 and S7-2). The rates of mutations in the 5S rRNA genes (transcribed by RNA polymerase III) are less than twofold different from those observed for single-copy genes in the same strain. Thus, the high rate of mutations observed for tRNA genes is not likely a consequence of RNA polymerase III transcription, but some other factor. In summary, the amount of single-stranded DNA in the rRNA genes in strains with low levels of DNA polymerase is roughly similar to that observed in most yeast genes.

Variants Arise during the Growth of the Low-Polymerase Strains that Have Elevated Rates of Recombination or Mutagenesis. Because of the very large numbers of mutations introduced by APOBEC, we cannot exclude the possibility that these mutations interact with the low levels of DNA polymerases to affect the frequency or position of APOBEC-induced mutations. As mentioned previously, most of the clusters with 16 or more mutations were derived from two (SY45-LD-4 and SY45-LD-5) of the seven SY45-LD isolates. We examined the SY45-LD isolates for mutations in about 70 genes involved in DNA replication or DNA repair. The SY45-LD-4 and SY45-LD-5 had mutations in both *PIF1* and *TOP1*, encoding a helicase involved in DNA synthesis/DNA repair, and topoisomerase I, respectively. Although none of the other isolates shared both mutations, it is not clear whether these mutations were responsible for the high frequency of large clusters of APOBEC-induced mutations; it is also unclear whether the mutations were in both copies of these genes. Nonetheless, the existence of novel phenotypes in the diploids expressing APOBEC suggests the utility of APOBEC for uncovering mutations that affect DNA replication or DNA repair.

Relationship between Elevated Levels of Mitotic Recombination and Elevated Levels of Single-Stranded DNA. We showed that reduced levels of replicative DNA polymerases result in increased amounts of single-stranded DNA, and we previously demonstrated that low levels of either DNA polymerase alpha or delta greatly elevated the frequency of mitotic recombination (11, 12). Using different methods, Feng et al. (47) provided evidence that treatment of *mec1* yeast strains with hydroxyurea caused accumulation of single-stranded DNA and DSBs.

One unresolved issue is the relationship between single-stranded DNA at the replication forks and the formation of DSBs. One possibility is that single-stranded DNA is prone to form secondary DNA structures that are substrates for structure-specific nucleases. A second alternative is the single-stranded DNA is more susceptible to DNA damage (for example, base modification) that is repaired in a manner that generates DNA breakage. Although the simplest interpretation of our results is that the elevated levels of single-stranded DNA in strains with low levels of DNA polymerases directly result in genetic instability, we cannot exclude the possibility that the genetic instability is a consequence of an extended S period or some other indirect consequence of low levels of replicative polymerases. This second possibility may be difficult to test since most treatments that extend the S period (for example, growth of cells in the presence of hydroxyurea) are likely to be associated with elevated levels of single-stranded DNA.

Summary. Our results show that lowering the levels of replicative DNA polymerases substantially elevates the amount of single-stranded DNA in yeast. This elevation most strongly affects the lagging-strand template in strains with low levels of DNA polymerase alpha and delta and affects both the lagging- and leading-strand templates in strains with low levels of DNA polymerase epsilon. It is likely that the increased level of single-stranded DNA is causally linked to the increased rates of LOH and chromosome rearrangements observed in strains with low levels of replicative DNA polymerases.

Materials and Methods

Strain Constructions. The diploids used in this study were hybrids generated by crosses of haploids isogenic with W303-1A and YJM789 and were closely related to the hybrid diploids used in many of our other studies (28, 48). The genotypes and sequences of primers used in the study are in *SI Appendix*, Table S1. Details of the constructions are in *SI Appendix*.

Analysis of APOBEC-Induced Mutations by DNA Sequencing. Whole-genome sequencing of *S. cerevisiae* strains was performed on an Illumina NextSEq. 500 sequencer using the 2 × 150 bp paired-end indexing protocol. DNA samples were prepared as described previously (28). The sequence coverage for each sample was ∼129-fold. The protocols used to define APOBEC-induced mutations, and to assign mutations to specific DNA strands (transcribed/nontranscribed or leading-/lagging-strand templates) are described in *SI Appendix*.

Data Availability. The sequencing data are freely available and are stored on the NCBI Website: Sequence Read Archive, PRJNA314677 (49).

Calculations concerning APOBEC-Induced Mutational Clusters. Two different methods were used to determine whether APOBEC-induced mutations were clustered, one based on a Poisson analysis (the P-method) and using a negative binomial distribution (the B-method) (18). The details of these methods are in the text above and in *SI Appendix*.

Analysis of LOH Events and Other Chromosome Alterations by DNA Sequencing. There are roughly 55,000 heterozygous SNPs in the diploid strains used in our study, and the coupling relationships of these SNPs is known by analysis of the haploid parental genomes (28). The average read depth in our sequence analysis was about 129. We calculated the number of reads for each SNP and divided that number by the average read depth for that isolate (read ratio). For example, for heterozygous SNPs, each SNP would be found about 65 times and, thus, the read ratios for each SNP would be about 0.5. LOH events were defined by one SNP having a read ratio of ≤0.1, and the alternative SNP having a read ratio of ≥0.7. A list of LOH events with their coordinates is given in Dataset S6-1, the different classes of events are depicted in Dataset S6-2, and the numbers of LOH events per isolate are summarized in Dataset S6-3. Other details about LOH analysis are in *SI Appendix*.

ACKNOWLEDGMENTS. We thank all members of the T.D.P. and Jinks-Robertson laboratories, Dave MacAlpine and Yin Yi for useful suggestions, and Duncan Smith for communicating unpublished data. The research was supported by grants from NIH and the US Army (R35GM118020 and W911NF1920082) (to T.D.P.), by the National Natural Science Foundation of China grants (31800055 and 31401058), the Fundamental Research Funds for the Central Universities (2019QNA4054), the Zhoushan City–Zhejiang University Joint Specific Project (2019C81055) (to D.-Q.Z.), and by the NIH Intramural Research Program Project Z1AES103266 (to D.A.G.).

1. H. Gaillard, T. García-Muse, A. Aguilera, Replication stress and cancer. *Nat. Rev. Cancer* **15**, 276–289 (2015).
2. M. Macheret, T. D. Halazonetis, DNA replication stress as a hallmark of cancer. *Annu. Rev. Pathol.* **10**, 425–448 (2015).
3. S. G. Durkin, T. W. Glover, Chromosome fragile sites. *Annu. Rev. Genet.* **41**, 169–192 (2007).
4. T. A. Kunkel, P. M. J. Burgers, Arranging eukaryotic nuclear DNA polymerases for replication: Specific interactions with accessory proteins arrange Pols α, δ, and ∈ in the replisome for leading-strand and lagging-strand DNA replication. *BioEssays* **39**, 1700070 (2017).
5. S. R. McGuffee, D. J. Smith, I. Whitehouse, Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol. Cell* **50**, 123–135 (2013).
6. S. A. Nick McElhinny, D. A. Gordenin, C. M. Stith, P. M. Burgers, T. A. Kunkel, Division of labor at the eukaryotic replication fork. *Mol. Cell* **30**, 137–144 (2008).
7. T. Kesti, K. Flick, S. Keränen, J. E. Syväoja, C. Wittenberg, DNA polymerase ε catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol. Cell* **3**, 679–685 (1999).
8. M. A. Garbacz *et al*., The absence of the catalytic domains of *Saccharomyces cerevisiae* DNA polymerase ∈ strongly reduces DNA replication fidelity. *Nucleic Acids Res.* **47**, 3986–3995 (2019).
9. D.-Q. Zheng, T. D. Petes, Genome instability induced by low levels of replicative DNA polymerases in yeast. *Genes (Basel)* **9**, E539 (2018).
10. F. J. Lemoine, N. P. Degtyareva, K. Lobachev, T. D. Petes, Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* **120**, 587–598 (2005).
11. W. Song, M. Dominska, P. W. Greenwell, T. D. Petes, Genome-wide high-resolution mapping of chromosome fragile sites in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2210–E2218 (2014).
12. D.-Q. Zheng, K. Zhang, X.-C. Wu, P. A. Mieczkowski, T. D. Petes, Global analysis of genomic instability caused by DNA replication stress in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E8114–E8121 (2016).
13. L. S. Symington, R. Rothstein, M. Lisby, Mechanisms and regulation of mitotic recombination in *Saccharomyces cerevisiae*. *Genetics* **198**, 795–835 (2014).
14. S. U. Siriwardena, K. Chen, A. S. Bhagwat, Functions and malfunctions of mammalian DNA-cytosine deaminases. *Chem. Rev.* **116**, 12688–12710 (2016).
15. S. A. Roberts *et al*., An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
16. K. Chan, D. A. Gordenin, Clusters of multiple mutations: Incidence and molecular mechanisms. *Annu. Rev. Genet.* **49**, 243–267 (2015).
17. B. Gómez-González, A. Aguilera, Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8409–8414 (2007).
18. S. A. Roberts *et al*., Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
19. B. J. Taylor *et al*., DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* **2**, e00534 (2013).
20. A. G. Lada *et al*., Genome-wide mutation avalanches induced in diploid yeast cells by a base analog or an APOBEC deaminase. *PLoS Genet.* **9**, e1003736 (2013).
21. J. I. Hoopes *et al*., APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during DNA replication. *Cell Rep.* **14**, 1273–1282 (2016).
22. N. Saini *et al*., APOBEC3B cytidine deaminase targets the non-transcribed strand of tRNA genes in yeast. *DNA Repair (Amst.)* **53**, 4–14 (2017).
23. K. Chan *et al*., An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat. Genet.* **47**, 1067–1072 (2015).
24. C. J. Sakofsky *et al*., Repair of multiple simultaneous double-strand breaks causes bursts of genome-wide clustered hypermutation. *PLoS Biol.* **17**, e3000464 (2019).
25. R. Elango *et al*., Repair of base damage within break-induced replication intermediates promotes kataegis associated with chromosome rearrangements. *Nucleic Acids Res.* **47**, 9666–9684 (2019).
26. A. G. Lada *et al*., AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biol. Direct* **7**, 47, discussion 47 (2012).
27. R. J. Kokoska, L. Stefanovic, J. DeMai, T. D. Petes, Increased rates of genomic deletions generated by mutations in the yeast gene encoding DNA polymerase δ or by decreases in the cellular levels of DNA polymerase δ. *Mol. Cell. Biol.* **20**, 7490–7504 (2000).
28. J. St Charles *et al*., High-resolution genome-wide analysis of irradiated (UV and γ-rays) diploid yeast cells reveals a high frequency of genomic loss of heterozygosity (LOH) events. *Genetics* **190**, 1267–1284 (2012).
29. Y. O. Zhu, M. L. Siegal, D. W. Hall, D. A. Petrov, Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2310–E2318 (2014).
30. J. M. Di Noia, M. S. Neuberger, Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).
31. A. G. Lada *et al*., Disruption of transcriptional coactivator Sub1 leads to genome-wide re-distribution of clustered mutations induced by APOBEC in active yeast genes. *PLoS Genet.* **11**, e1005217 (2015).
32. X. Rong-Mullins, M. C. Ayers, M. Summers, J. E. G. Gallagher, Transcriptional profiling of *Saccharomyces cerevisiae* reveals the impact of variation of a single transcription factor on differential gene expression in 4NQO, fermentable, and nonfermentable carbon sources. *G3 (Bethesda)* **8**, 607–619 (2018).
33. B. J. Taylor, Y. L. Wu, C. Rada, Active RNAP pre-initiation sites are highly mutated by cytidine deaminases in yeast, with AID targeting small RNA genes. *eLife* **3**, e03553 (2014).
34. E. X. Kwan *et al*., A natural polymorphism in rDNA replication origins links origin activation with calorie restriction and lifespan. *PLoS Genet.* **9**, e1003329 (2013).
35. D. Salim *et al*., DNA replication stress restricts ribosomal DNA copy number. *PLoS Genet.* **13**, e1007006 (2017).
36. A. M. Deshpande, C. S. Newlon, DNA replication fork pause sites dependent on transcription. *Science* **272**, 1030–1033 (1996).
37. D. A. Kiktev, Z. Sheng, K. S. Lobachev, T. D. Petes, GC content elevates mutation and recombination rates in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7109–E7118 (2018).
38. T. D. Petes, Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* **2**, 360–369 (2001).
39. L. Chelico, P. Pham, P. Calabrese, M. F. Goodman, APOBEC3G DNA deaminase acts processively 3′ –> 5′ on single-stranded DNA. *Nat. Struct. Mol. Biol.* **13**, 392–399 (2006).
40. D. J. Smith, I. Whitehouse, Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**, 434–438 (2012).
41. S. Y. Porcella *et al*, Separable, Ctf4-mediated recruitment of DNA Polymerase α for initiation of DNA synthesis at replication origins and lagging-strand priming during replication elongation. https://doi.org/10.1101/352567 (23 January 2020).

42. S. Kim, H. G. Dallmann, C. S. McHenry, K. J. Marians, τ couples the leading- and lagging-strand polymerases at the *Escherichia coli* DNA replication fork. *J. Biol. Chem.* **271**, 21406–21412 (1996).
43. Z. Yuan *et al*., Ctf4 organizes sister replisomes and Pol α into a replication factory. *eLife* **8**, e47405 (2019).
44. K. O'Connell, S. Jinks-Robertson, T. D. Petes, Elevated genome-wide instability in yeast mutants lacking RNase H activity. *Genetics* **201**, 963–975 (2015).
45. A. El Hage, S. Webb, A. Kerr, D. Tollervey, Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet.* **10**, e1004716 (2014).
46. Y. Kumar, P. Bhargava, A unique nucleosome arrangement, maintained actively by chromatin remodelers facilitates transcription of yeast tRNA genes. *BMC Genomics* **14**, 402 (2013).
47. W. Feng, S. C. Di Rienzi, M. K. Raghuraman, B. J. Brewer, Replication stress-induced chromosome breakage is correlated with replication fork progression and is preceded by single-stranded DNA formation. *G3 (Bethesda)* **1**, 327–335 (2011).
48. J. St Charles, T. D. Petes, High-resolution mapping of spontaneous mitotic recombination hotspots on the 1.1 Mb arm of yeast chromosome IV. *PLoS Genet.* **9**, e1003434 (2013).
49. Y. Sui *et al*, Bioproject: Saccharomyces cerevisiae Raw Sequence Reads. National Center for Biotechnology Information Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra/PRJNA314677. Deposited 6 September 2019.