Mutagenic mechanisms of cancer-associated DNA polymerase ε alleles

Mareike Herzog[1,2], Elisa Alonso-Perez[3], Israel Salguero[1], Jonas Warringer[3], David J. Adams[2], Stephen P. Jackson ⬤[1,*] **and Fabio Puddu** ⬤[1,*]

[1]The Wellcome/Cancer Research UK Gurdon Institute and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK, [2]The Wellcome Sanger Institute, Hinxton CB10 1HH, UK and [3]Department of Chemistry and Molecular Biology, University of Gothenburg, Medicinaregatan 9 C, 413 90, Göteborg, Sweden

ABSTRACT

A single amino acid residue change in the exonuclease domain of human DNA polymerase ε, P286R, is associated with the development of colorectal cancers, and has been shown to impart a mutator phenotype. The corresponding Pol ε allele in the yeast *Saccharomyces cerevisiae* (*pol2-P301R),* was found to drive greater mutagenesis than an entirely exonuclease-deficient Pol ε (*pol2–4*), an unexpected phenotype of *ultra*-mutagenesis. By studying the impact on mutation frequency, type, replication-strand bias, and sequence context, we show that *ultra*-mutagenesis is commonly observed in yeast cells carrying a range of cancer-associated Pol ε exonuclease domain alleles. Similarities between mutations generated by these alleles and those generated in *pol2–4* cells indicate a shared mechanism of mutagenesis that yields a mutation pattern similar to cancer Signature 14. Comparison of *POL2 ultra*-mutator with *pol2-M644G*, a mutant in the polymerase domain decreasing Pol ε fidelity, revealed unexpected analogies in the sequence context and strand bias of mutations. Analysis of mutational patterns unique to exonuclease domain mutant cells suggests that backtracking of the polymerase, when the mismatched primer end cannot be accommodated in the proofreading domain, results in the observed insertions and T>A mutations in specific sequence contexts.

INTRODUCTION

Aneuploidies or large genomic rearrangements are a common feature of many types of cancer, but widespread hypermutation—the extensive accumulation of single nucleotide variants (SNVs) or small insertion/deletions (INDELs)—is comparatively less frequent (17% of adult cancers analysed in (1)). Hypermutation usually arises from exposure to mutagens, such as ultra-violet light or tobacco smoke, and/or from hereditary or acquired DNA repair defects leaving a trace of specific mutation signatures (2). DNA mismatch repair (MMR) inactivation, for example, has long been known to drive somatic hypermutation that leads to hereditary colorectal adenocarcinomas (3). More recently, specific mutator alleles altering the exonuclease (proofreading) domain of replicative DNA polymerases δ and ε (Pol δ and Pol ε), such as *POLE-P286R*, were found to foster SNV hyper-mutation in the presence of functional MMR, and drive the development of hereditary colorectal or endometrial cancers (4–6).

The proofreading activity of B-family DNA polymerases (such as Pol δ and Pol ε) is triggered by the presence of a base-pair mismatch between the template and the nascent DNA strand at the primer-template junction, originating from base mis-insertion or from a damaged template base (7,8), and, possibly, by other impediments to elongation. In these situations, the 3′ end of the primer is melted and moved to the spatially separate proofreading domain, where one or more nucleotides are exonucleolytically degraded (9–11). The primer end is then returned to the polymerase domain, where DNA synthesis can continue (12,13).

The origin of hypermutation in cancer cells with Pol ε proofreading domain mutants was originally ascribed to inactivation of Pol ε exonuclease activity (4). Subsequent work in the yeast *S. cerevisiae*, however, revealed that *pol2-P301R*—the ortholog of the human Pol ε mutation *POLE-P286R*—drives substantially greater mutagenesis than Pol ε exo⁻ (14), an exonuclease-deficient variant of Pol ε encoded by the yeast *pol2–4* allele (15). This *ultra*-mutator phenotype indicates that inactivation of the exonuclease activity is not primarily at the origin of the massive mutation

accumulation observed in *pol2-P301R* cells; and accordingly, biochemical work revealed that exonuclease activity is still detectable in the P301R mutant polymerase (∼20% to ∼60% of wild-type, depending on the assay) (16). Further structural analyses revealed that this amino acid residue change creates a barrier at the entrance of the exonuclease domain, possibly preventing the newly synthesized strand from accessing it (17). Inability to position the mismatched primer in the proofreading domain, and the observed increased mismatch extension ability, could explain the *ultra*-mutagenic phenotype, but *in vitro* polymerase assays failed to recapitulate *ultra*-mutagenesis (16). These observations open up the possibility that other cellular processes may be involved in the generation of mutations and motivated the studies we describe herein.

MATERIALS AND METHODS

Yeast strains and plasmids

Sequence alignment with Clustal Omega version 1.2.1 was carried out to determine the *Saccharomyces cerevisiae* residues orthologous to human Pol ε residues mutated in the literature analysed. Uniprot sequences used for alignment were *Homo sapiens* POLE (Q07864) and *S. cerevisiae* Pol2 (P21951). All *S. cerevisiae* strains used were derived from the laboratory strain W303 (*leu2–3,112 trp1–1 can1–100 ura3–1 ade2–1 his3–11,15 RAD5*). Polymerase mutant alleles were created by cloning an N-terminal *POL2* PCR fragment into pRS306 and generating the mutations of interest by site directed mutagenesis using the QuickChange Lightning Kit following manufacturer's instructions (Agilent Technologies). Polymerase mutant alleles were introduced into *MAT***a** haploid *S. cerevisiae* W303 strains, resulting in a full-length copy carrying the mutation and a non-mutated, truncated N-terminal fragment. Haploid *pol2* mutants were then mated to a wild-type isogenic *MATalpha* strain to generate heterozygous diploid mutant strains. Deletion of *MSH2* was introduced in wild-type W303 by one-step gene disruption. Disruptions were confirmed by PCR and whole-genome sequencing. Haploid double mutants *pol2 msh2Δ* were recovered by mating haploid *pol2* with haploid *msh2Δ* strains, sporulation, tetrad dissection and analysis. The genotypes of strains are described in Supplementary Table S1.

Growth rate and mutation assays

The growth rates of heterozygous diploid polymerase mutant strains were assessed by growing cultures to stationary phase, diluting them into rich medium and growing for 450 min. Growth was assayed by measuring absorbance at 595 nm wavelength. To determine forward mutation rates at the *LYP1* locus, single colonies were excised from agar plates with scalpels, inoculated in rich medium, and grown to saturation. Cell cultures were subsequently diluted 1:100,000 and plated on YPAD (1% yeast extract, 2% peptone, 2% glucose, 40 mg/l adenine) plates, or plated (50–250 μl) without dilution on SD (Synthetic Defined media) –LEU +Thialysine (50 mg/l) plates. Thialysine kills cells expressing an active Lyp1 transporter. Different amounts were used for

different strains to obtain countable plates. Mutation rates were calculated as described in (18).

Single-cell bottleneck propagation

Heterozygous diploid strains were grown on solid media at 30°C and propagated for 26 passages, every 2–3 days, through single-cell bottlenecks by means of repetitive isolation and single-colony picking. Cells were estimated to have undergone ∼20 generations per passage to generate a colony of 10^6 cells (∼500 generations during the entire propagation). Colonies were picked randomly to avoid bias towards adaptive or deleterious mutations (with the exception of lethal mutations). In each experiment, each strain was propagated in 18 parallel lines. Deviations from this number due to failures to sequence or to confirm the presence of the mutation at the end of the propagation are denoted in the relevant figures. Haploid strains were propagated for 13 passages. Standard YPAD non-selective rich medium was used. In each experiment, whole-genome sequencing of two random colonies for each strain was attempted at passage 0 and at the end of the propagation. Only mutations observed in both colonies (where a second colony was available), and absent from passage 0, were retained for further analysis.

Small-population bottleneck propagation

Haploid and heterozygous diploid polymerase mutant strains were propagated in a 1536 plate format in a non-selective complete synthetic medium (0.14% YNB, 0.5% ammonium sulphate, 0.077% complete supplement mixture [ForMedium], 2% (w/v) glucose and pH buffered to pH 5.8 with 1% (w/v) succinic acid). Plates were replicated using a ROTOR Robot (Singer Ltd, UK) and 1536 short pin pads every 2–3 days for 40 passages, through bottlenecks estimated to contain ∼10^4 cells. Effective population size is, however, likely to be smaller due to the population structure of cells in a colony. The number of cells in the bottleneck was calculated by estimates of pixel intensities using light transmission and conversion of pixel intensities into cell counts by calibration to a flow cytometry-based reference (19). In these conditions, wild-type BY strains undergo ∼6 doublings per growth cycle (passage) suggesting that cells underwent ∼240 generations over the duration of the experiment. Final populations were streaked for single colonies and the whole genome of 18–26 isolates per strain was sequenced. After reassigning strains to the correct genotypes and ploidy 12–38 isolates per strain were analysed.

DNA extraction, library preparation and whole genome sequencing

Genomic DNA extractions and library preparations were carried out as previously described (20). Libraries were sequenced using either HiSeq 2000 or HiSeq X (Illumina) to generate 125 or 150 bp paired-end reads, respectively.

Reference genome alignment

Sequencing reads were aligned to the *S. cerevisiae* S288c (R64–1-1) reference genome using BWA mem (-t 16 -p -T

0) and duplicates were marked with bamstreamingmarkduplicates (biobambam2 2.0.50) and stored in CRAM format (primary data). From these, reads were extracted with *samtools fastq* and subsequently re-aligned to a modified reference genome in which repetitive DNA regions were hard-masked and moved, as single-copy sequences, to *ad hoc* artificial chromosomes. Duplicates were marked with *bamsormadup SO = coordinate fixmate = 1.*

Confirmation of strain genotypes

Samples were automatically checked for their expected polymerase genotype using the script *deletion_check.pl.* Briefly, for point mutations the DNA sequence from the triplet coding for the residue in question was extracted from the sequencing data, translated and compared with the expected. Deletions and genetic mating type were determined as previously described ([20]). Ploidy was determined *a posteriori*, based on the distribution of the observed allelic frequencies (AF). Strains displaying a median of alleles with AF ~0.5 were classified as diploid, while strains with a median AF of ~1 were classified as haploid. In the rare cases in which the number of de-novo alleles was too low, generating intermediate AF, the genetic mating type, identified bioinformatically, was used.

Variant calling, consequence annotation and filtering

SNVs and small insertions/deletions (INDELs) were identified chromosome by chromosome using *samtools mpileup* (v.1.9), with the following options: *-g -t DP,DV -C0 -p -m3 -F0.2 -d10000*, followed by *bcftools call -vm -f GQ* (v.1.9). All mutations from each chromosome were merged with *bcftools concat*. All variants were annotated with the Ensembl Variant Effect Predictor (VEP; v95.3). INDELs were subsequently normalized with *bcftools norm -m-both –check-ref e* and sorted with *bcftools sort*. Low quality variants were flagged with *bcftools filter* with the following options *-m + -e 'INFO/DP<10' -e 'FORMAT/DV<3' -e 'TYPE = \'snp\' & QUAL<100' -e 'TYPE = \'indel\' & QUAL<30' -e 'FORMAT/GQ<40' -g 7*. Variants present in control samples were subsequently removed with *bcftools isec -w1 -C {sample_file} {control_file}*.

Further mutation filtering

SNV mutations were further filtered on the QUAL value and their prevalence across different sequencing samples. Given the relatively low number of mutated positions compared to the genome size, the vast majority of mutations are expected to be unique in different MA lines in single-cell bottleneck experiments, and shared mutations are likely to originate from systematic sequencing errors. Taking this into account, we removed mutations whose quality was below an arbitrary threshold that grows linearly with the prevalence of the mutation in different samples (Supplementary Figure S1), thus excluding mutations that are frequently observed and of lower quality. In small-population bottleneck experiments, many mutations are shared between different colonies from the final population, because of their shared ancestry. For this reason, a similar, less stringent threshold was used. Filters were designed to remove

~1–10% of all mutations. A similar rationale was used to filter INDELS. Small changes in the filtering parameters do not substantially alter the results of the subsequent analyses.

Analysis of mutation numbers

The total number of SNPs/INDELs for each sequencing sample was calculated by counting the number of mutations passing all filters. In single-cell bottleneck propagations two colonies per MA line were sequenced and only mutations observed in both colonies (and absent from passage 0) were retained for further analysis; where a second colony was not available because of sequencing failure, all mutations in the only available colony were retained. For small population bottleneck propagation all mutations present in each colony from the final population were considered. Mutation rates are given in terms of SNV(INDEL)/haploid genome/passage and converted to SNV/generation/bp assuming a haploid genome size of 12 071 326 bp and 20 generations from single cell to colony.

Analysis of mutation types

In small-population propagation experiments, one single mutagenic event is likely observed in more than one colony picked from the final-population. Thus, mutations derived from small-population experiments were initially grouped by MA line and, in each line, when the same mutation was observed in more than one colony, only one instance was retained. Mutations from single-cell and small-propagation experiments were then pooled for the purpose of the subsequent analyses. Analysis of the frequency of different SNV classes was carried out by grouping the mutations by genotype (irrespectively of the type of propagation or ploidy), counting the number relevant mutations, and summing complementary pairs (e.g. A>C + T>G).

Analysis of replication strand bias

The relative position of each mutation with respect to the nearest replication origin was calculated in two steps. First, a replication model was built using the coordinates of replication origins obtained from OriDB (http://cerevisiae.oridb.org; only using 'confirmed' origins); location of each origin was calculated as the midpoint of the ARS region; termination points were arbitrarily defined as the midpoint of each inter-origin span; leading-strand regions were defined as regions comprised between an origin and the termination point to its immediate right; lagging-strand regions were defined as regions comprised between an origin and the termination point to its immediate left. Second, each mutation was localized to an inter-origin span (thus, mutations located before the first origin or after the last origin of each chromosome were discarded); the distance between each mutation and the origin to its immediate left was calculated, and normalized for the size of the inter-origin span in which the mutation was located, so that a distance of 50% coincides to the midpoint termination zone, and a distance of 100% coincides to the subsequent origin. To avoid over-weighting shared mutations originating from small-population bottleneck experiments, only a distinct set

of mutations was considered. The density of each mutation type was then plotted as a function of the relative distance of mutations from the origin to the immediate left.

Analysis of mutation patterns and comparison with mutation signatures

Mutational patterns were obtained by calculating the frequency of the 96 trinucleotide contexts (channels) in which mutations belonging to one of the six main classes (C>A, C>G, C>T, T>A, T>C, T>G) occurred. The remaining mutations (G>T, G>C, G>A, A>T, A>G, A>C) were reverse complemented along with their context and assigned to the appropriate channel. The frequency of each channel was then normalized by the relative abundance of each trinucleotide in the yeast genome.

Comparison with mutational signatures identified in cancers was calculated using cosine similarity (21) and the *cos_sim_matrix* function of the *MutationalPatterns* R package (22). Cancer signatures were obtained from https://cancer.sanger.ac.uk/cancergenome/assets/signatures_probabilities.txt.

Analysis of mutation sequence context

After extracting the context (five nucleotides) in which each mutation occur, mutations were classified as right or left mutations depending on where they occurred in relation to origins of replication and presumed termination points (see Analysis of replication strand bias). To exclude as much as possible mutations introduced by Pol ε synthesizing DNA beyond the midpoint of each replicon, only mutations occurring in the first and last third of each replicon were considered. Sequence logo images from the context of each mutation class were obtained with the *ggseqlogo* R package (23).

RESULTS

A spectrum of DNA Pol ε *ultra*-mutator alleles

As an approach to investigate how *ultra*-mutator Pol ε mutants exert their genotoxic activities *in vivo*, we focused on *POLE* alleles (Figure 1A) originally described as drivers of colorectal and endometrial cancers (4–6). To avoid confounding factors that would arise from conducting such studies directly in cancer cell lines, such as a higher background of genomic instability, we introduced, where possible, the corresponding mutations in heterozygous state into the diploid yeast strain W303. Thus, we generated eight yeast strains, each carrying one of eight *pol2* alleles encoding the different aminoacid changes in evolutionary conserved residues (Figure 1A). Several colonies (18–54) for each POL2/*pol2*- heterozygous diploid strain were then independently cultured through single-cell bottlenecks (26 passages; ∼500 cell generations), allowing mutations to accumulate in the genome. Mutational events that occurred during the experiment were identified by whole-genome sequencing of each mutation accumulation (MA) line at the start and the end of the experiment. For comparison, we also carried out such analyses of MA lines containing wild-type *POL2* or a proofreading defective *pol2–4* allele (15,24).

Since at each passage colonies were randomly selected, we expect mutation rates and spectra to be unbiased with regard to their consequences on gene function, except for overlooking lethal mutations.

In line with previous estimates (1.67–3.8×10^{-10} SNV/generation/bp) (25–27), wild-type MA lines acquired a median of 0.1 SNV/haploid genome/passage (or ∼4.14×10^{-10} SNV/generation/bp), while strains carrying a proofreading-defective allele (*pol2–4*/*POL2*) displayed a modest increase over this rate (∼30% or 7.04×10^{-10} SNV/generation/bp; Figure 1B). Four *pol2* alleles led to either no detectable mutator phenotype (Q468R, D290V) or a hypermutator phenotype similar in magnitude to that of *pol2–4*/*POL2* cells (V426L, A480V; Figure 1B). In contrast, four other Pol ε alleles (P301R, M459K, S312F, L439V) accumulated considerably more mutations than would be expected by simple lack of exonuclease activity (3–21 times or 2.1–14.7×10^{-9} SNV/generation/bp), thus reflecting an *ultra*-mutator phenotype. Since the growth rates of these strains did not substantially differ from those of the other *POL2* mutants or from the wild-type strain (Supplementary Figure S2A), MA could be taken as an accurate reflection of mutation rates. Accordingly, when we characterised haploid strains carrying wild-type or various mutant *POL2* alleles for their rates of loss-of-function mutations yielding thialysine resistance (mainly occurring via inactivation of *LYP1* (28)), there was good concordance ($R = 0.81$) between these data and results from MA experiments (Figure 1C).

Notably, the presence of a wild-type polymerase reduced the mutagenic effects of the hypermutator allele as MA in heterozygotic diploid cells was substantially reduced compared to haploid cells (62–92%; Supplementary Figure S2B). This reduction exceeded the expected 50%, pointing to other factors impacting on Pol ε-driven mutagenesis such as the reduced ratio of mutagenic Pol ε to Pol δ, which could contribute to increase the frequency of *in trans* proofreading by Pol δ (29,30) in heterozygous diploid cells. Alternatively, in diploid cells, the wild-type polymerase may be preferentially expressed or used.

The haploid state also unmasked an *ultra*-mutator phenotype for *pol2-A480V*, which in heterozygosis did not accumulate significantly more mutations than the corresponding exonuclease deficient strain (compare blue and red *P*-values, Supplementary Figure S2B). In contrast to SNVs, the strongest *pol2-ultra* alleles only led to a small but statistically significant increase in the accumulation of INDELs in haploid cells (Supplementary Figure S2C). These findings are in line with previous studies of *pol2-P301R*, *pol2-L439V*, *pol2-V426L* and *pol2-D290V* cells using classical mutation rate assays (31), and together with our results show that *ultra*-mutagenesis—the accumulation of considerably more mutations than would be expected by loss of exonuclease activity—is a common outcome for mutations in the proofreading domain of Pol ε that are found in cancers.

Diploidization and accumulation of secondary Pol ε mutations in a competitive growth assay

Assessing MA in cells propagated through population bottlenecks of ∼3×10^4 cells broadly confirmed our initial ob-
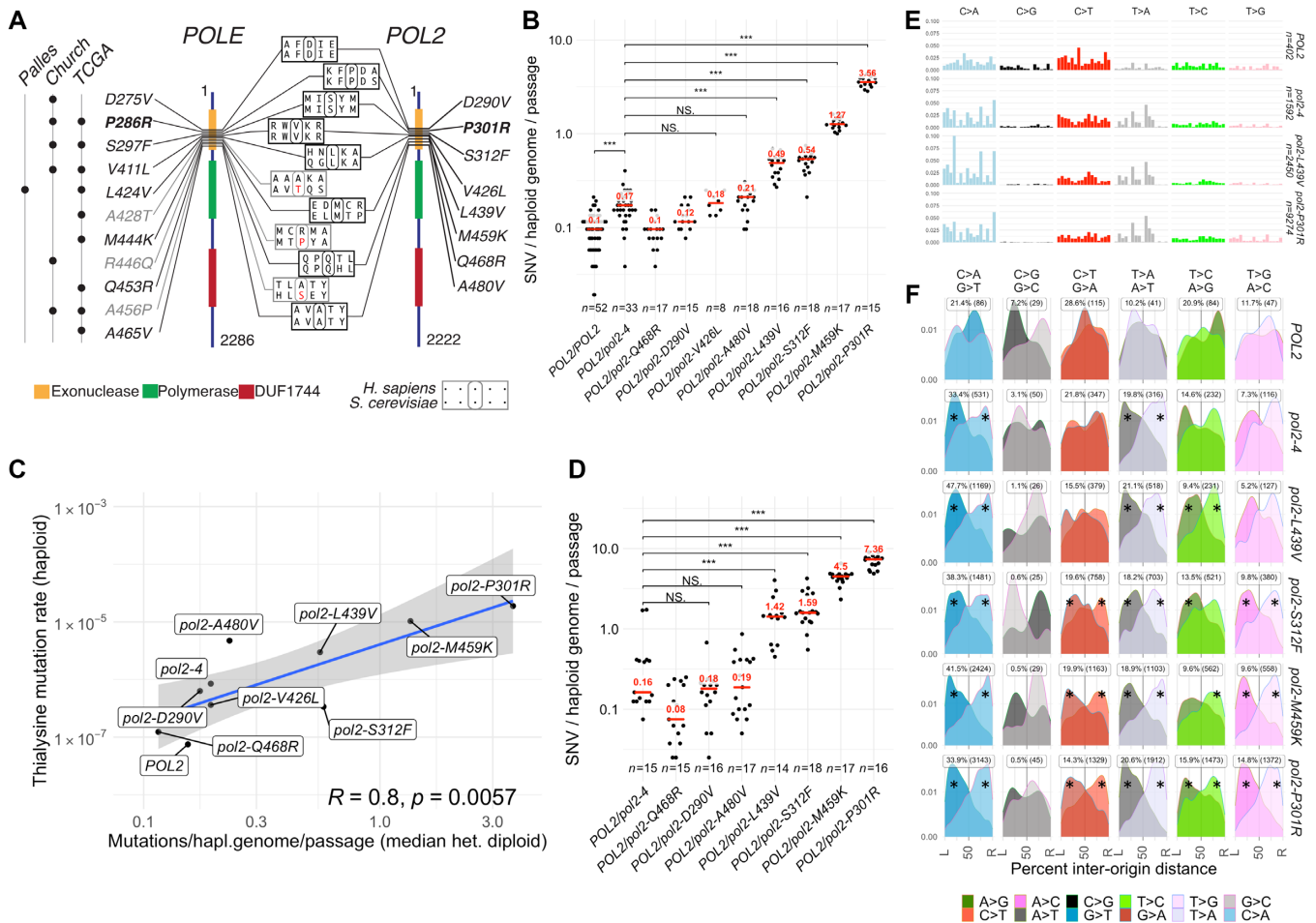
Figure 1. Analysis of *ultra*-mutator alleles of DNA polymerase ε modelled in *S. cerevisiae*. (**A**) Outline of DNA polymerase ε ultra-mutator alleles identified in cancers and corresponding mutations in *S. cerevisiae POL2*; the numbers indicate the residue position and domains are indicated with coloured boxes (DUF: domain with unknown function; Palles: mutation first reported in Palles C., *et. al*, Nat. Gen. 2013; Church: mutation first reported in Church DN, *et al*., Hum. Mol. Genet. 2013; TCGA: mutation present in The Cancer Genome Atlas). Mutations in grey occur in residues not evolutionary conserved in *S. cerevisiae*. The most studied mutation (P286R/P301R) is indicated in bold. (**B**) Rates of mutation accumulation in diploid yeast strains carrying indicated heterozygous *POL2* alleles propagated through single-cell passage bottlenecks for ∼500 generations. Each independent evolution line is indicated by a dot, while the median is indicated in red. The number of independent lines (*n*) is indicated below. Statistical test: Mann–Whitney–Wilcoxon (**C**) Comparison of mutation accumulation rates (x-axis) and loss-of-function mutation rates at *LYP1*, yielding thialysine resistance (y-axis). Blue line indicates the linear regression model; shaded area the 95% confidence interval; *R* is Pearson's correlation coefficient. (**D**) Rates of mutation accumulation in strains carrying the indicated heterozygous *POL2* alleles propagated through small-population passage bottlenecks for 350–450 generations. Different colonies derived from the final population are indicated by dots, while the median is indicated in red. The number of colonies studied (*n*) is indicated below. Statistical test: Mann–Whitney–Wilcoxon. (**E**) Frequency of mutations generated in the presence of the indicated *pol2* alleles, by type of mutation and trinucleotide context in which the mutations occur; frequencies have been corrected for the trinucleotide frequency in the yeast genome; see Figure 2C for x-axis labels. (**F**) Geographical distribution of the density of mutations by type in relation to replication origins. The origins to the left and to the right of each mutation are indicated by L and R respectively, and the distance is expressed as a percentage of the inter-origin distance of each origin pair (L→50%: leading strand; 50%→R: lagging strand). Asterisks indicate mutation types significantly deviating from uniformity ($P < 0.01$ $χ^2$ test).

servations, despite a larger number of mutations per passage, and much higher variability between different colonies (Figure 1D). These effects likely arose from the experimental settings: population expansion through ∼10^4 rather than single cell bottlenecks presumably allowed different clones to grow at different rates, and random sampling at each passage would favour the propagation and analysis of fitter, faster-growing clones, which would have completed more DNA replications and therefore accumulated more neutral or adaptive mutations than slower growing ones. The appearance and selection of anti-mutators suppressor muta-

tions (32,33) could also explain the increased variability in mutation numbers. Indeed, when we analysed the prevalence of mutations in *POL2* we found evidence of additional mutations in this gene often co-occurring with the stronger mutators (P301R, M459K, S312F, L439V) but only rarely with the weaker mutators or in wild-type cells (Supplementary Figure S3A and B). In line with recent findings (33) we also identified one spontaneous diploidization event in four isolates of the *pol2-M459K* haploid strain, which co-occurred with several additional *pol2* mutations (Supplementary Figure S3A, B; Figure S5C, blue circle).

Exonuclease deficient and *pol2-ultra* alleles have similar mutational spectra

Analysis of the mutational spectra generated in the absence of Pol ε exonuclease activity (*pol2–4*) revealed a relative increase in the frequency of T>A/A>T transversions, which is further expanded in *pol2-ultra* cells, while C>G transversions—the rarest class in wild-type strains—becomes relatively rarer (Figure 1E). Analysis of the trinucleotide context in which mutations occur also revealed a very strong resemblance between the mutational patterns generated by the different *pol2-ultra* mutants and by *pol2–4* cells (Figure 1E and Supplementary Figure S4A and B; cosine similarity 0.93–0.98). *POLE*–mutated tumours display very specific mutation signatures (SBS 10 and 28) in humans and mice (34,35), and these signatures have been recapitulated in human cells engineered to express mutant POLE (36). When we compared the mutational patterns generated by *pol2-ultra* mutants with cancer signatures we however found only moderate or weak similarities to SBS 14 (Supplementary Figure S4C, cosine similarity 0.66–0.75) or SBS 8 (0.64–0.68).

Most variants generated in the presence of Pol ε mutants are likely introduced on the replication leading strand, where the activity of this DNA polymerase is confined (37,38). To measure the replication-strand bias of the observed mutations, we calculated the relative distance of each mutation from the replication origins located to its immediate left and right, and then calculated the mutational density of each complementary mutation pair (e.g. G>A and C>T) as a function of the distance (Figure 1F). This analysis revealed a strong asymmetry (Cohen's $w = 0.43$–0.51) in the distribution of A>C, A>T and G>T in *pol2-P301R* cells, and a weaker asymmetry ($w = 0.17$–0.29) for A>G and C>T mutations. In particular A>C, A>G, A>T, G>A and G>T were observed more frequently on the leading strand than their complementary counterparts (Figure 1F and Supplementary Figure S4D; the low mutation counts for C>G/G>C transversions did not permit establishment of whether a bias is present in this channel). Strikingly, this pattern of mutagenesis was observed in cells carrying both stronger and weaker *ultra*-mutator alleles, and in *pol2–4* cells as well, with the degree of asymmetry increasing with the total number of mutations available for analysis (Supplementary Figure S4E). Taken together, these results strongly suggest a common mechanistic origin for mutations observed in cells lacking Pol ε exonuclease activity and in cells carrying *ultra*-mutator Pol ε variants. They also indicate that SNV accumulation in *pol2–4* cells does not originate from the mere lack of exonucleolytic activity.

Synergism of Pol ε exonuclease domain mutants with MMR deficiency

Mismatch-repair (MMR) recognises different DNA duplex mis-pairs with different efficiencies (39), thereby distorting the frequencies of different mutation classes from the frequencies generated by DNA polymerases. As an approach to determine the mutational patterns as they are generated by Pol ε *ultra*-mutators (those Pol ε alleles driving mutagenesis at a higher rate than *pol2–4*), we attempted to delete *MSH2*, which encodes a mismatch binding ATPase that

is required for all branches of MMR. As previously observed for several other DNA polymerase mutator alleles (16,24,32), sporulation of most *pol2-ultra msh2Δ* heterozygous diploids did not yield viable double mutant strains for the strongest *ultra*-mutators (*pol2-P301R*, *pol2-M459K* and *pol2-S312F*). Microscopic observation of these spores revealed that they germinated to vegetative cells but ceased to divide after a few generations, a phenotype consistent with extreme mutational burden leading to 'error-induced extinction' (32,33). By contrast, double mutants for the hypermutator *pol2–4* in combination with *msh2Δ* were obtained from corresponding heterozygous diploids but displayed reduced colony size compared to MMR-proficient controls. We also managed to obtain a viable MMR-deficient version of *pol2–L439V*—a relatively weak *ultra*-mutator. MA analyses of these strains confirmed that, similarly to heterozygous diploids, *pol2–4* and *pol2-L439V* haploid cells accumulated SNVs at a faster rate compared to wild-type strains in the presence of functional MMR (∼4× and ∼12× respectively; Figure 2A and Supplementary Figure S2B), while disruption of *MSH2* alone resulted in a ∼16-fold increase (or ∼7.4 × 10^{-9} SNV/bp/generation, similar to a previous estimate of 4.8 × 10^{-9} SNV/bp/generation (40)). In contrast, a dramatic increase in SNV accumulation was evident when mismatch repair inactivation was combined with *pol2–4* or *pol2-L439V* (∼421× and ∼874×, respectively; Figure 2A). These numbers are well above the expected mutation rate increases in the double mutants under an additive model (∼21X and ∼28X, respectively) or a multiplicative model (41) (∼69X and ∼196X, respectively), demonstrating a synergistic interaction. Additionally, *pol2-L439V* accumulated more mutations than *pol2–4* even in the absence of MMR, indicating that the ultra-mutator phenotype does not arise from a differential mismatch repair efficiency. Analysis of allele frequencies and status of the mating type (*MAT*/*HM*) loci also revealed that both *msh2Δpol2–4* and *msh2Δpol2-L439V* strains were actually diploid. Both strains were homozygous for *MSH2* deletion and *pol2* mutations, but the former was a *MAT***a**/*MAT***a** diploid—possibly originating from a whole-genome duplication event caused by failed cytokinesis (33)—and the latter a *MAT***a**/*MAT***alpha** diploid—possibly originating from homothallic mating after a rare mating-type switch event (Supplementary Figure S5A and B). These results are in line with recent findings in yeast *pol2-P301R* cells (33) and in *POLE P286R*-driven mouse cancer models (42), and suggest that a transition to the diploid or polyploid state facilitates survival in the face of extreme mutagenesis as expected from the fact that deleterious mutations are frequently recessive and often masked in a heterozygotic diploid state. Similar to what was previously observed for cells carrying the strongest Pol ε hypermutators, we also found that both *msh2Δ*/*msh2Δ pol2–4*/*pol2–4* and *msh2Δ*/*msh2 pol2-L439V*/*pol2-L439V* accumulated additional mutations in *POL2* (Supplementary Figure S3C and D).

Pol ε–L439V and Pol ε exo⁻ yield replication-strand biased Signature 14

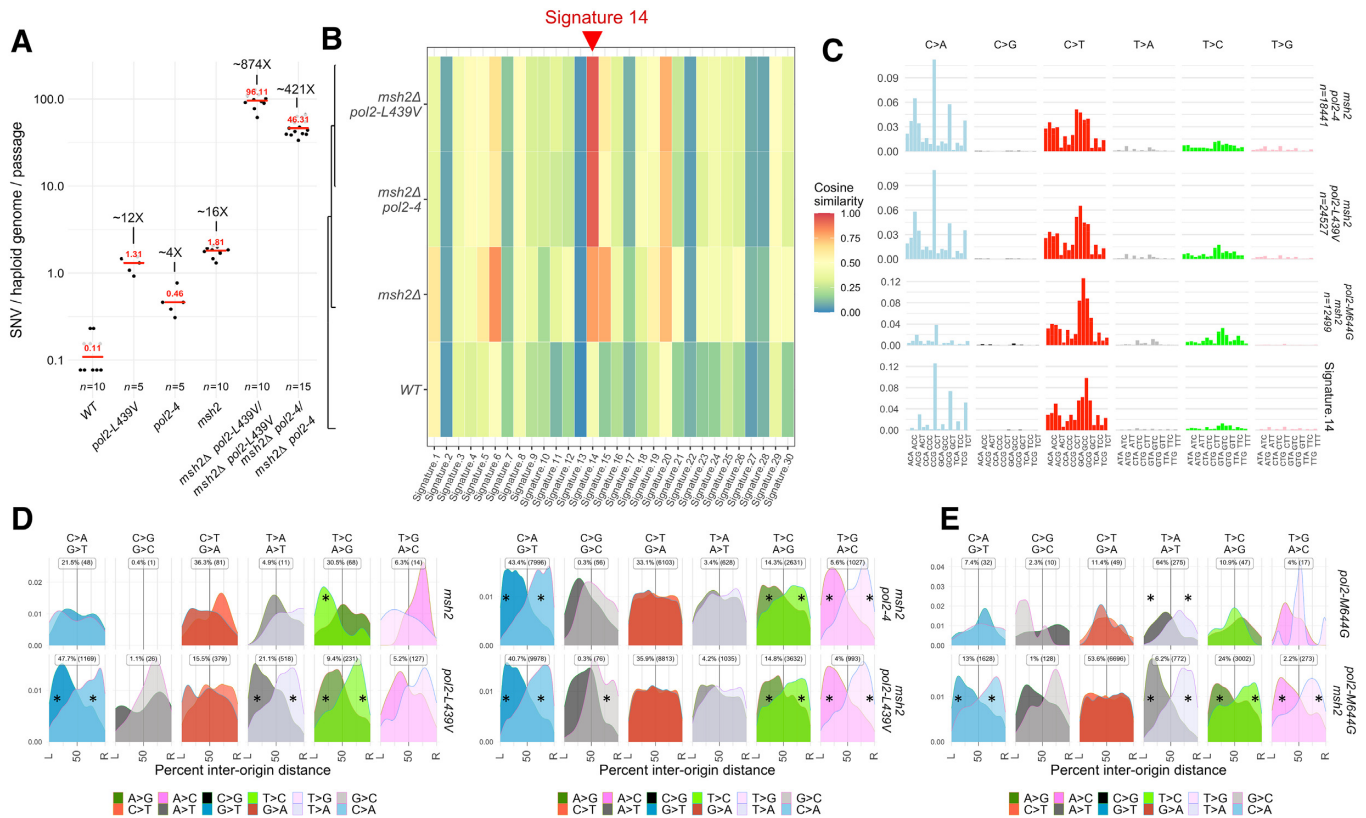Analysis of the trinucleotide context in which mutations are introduced by Pol ε–L439V and Pol ε exo⁻, before MMR

Figure 2. Mutation signatures generated by Pol ε exonuclease domain alleles in the absence of mismatch repair. (**A**) Number of SNVs generated during single-cell bottleneck propagation by Pol2 exo(*pol2–4*) or the *pol2-ultra* mutant *pol2-L439V* in the presence or in the absence of functional mismatch repair. The median is indicated in red. (**B**) Similarity between the SNV patterns generated by *pol2–4* and *pol2-L439V* in the absence of functional mismatch repair and the 30 mutation signatures identified in cancers. Colours indicate cosine similarity (Alexandrov, 2013). (**C**) Frequency of mutations generated by Pol2 exo⁻ (*pol2–4*), the *pol2-ultra* mutant *pol2-L439V*, or *pol2-M644G* in the absence of mismatch repair, by type of mutation and trinucleotide context in which the mutations occur; frequencies have been corrected for the trinucleotide frequency in the yeast genome; the most similar COSMIC signature (SBS14) is also represented. (**D**) Geographical distribution of the density of mutations by type in relation to replication origins. The origins to the left and to the right of each mutation are indicated by L and R, and the distance is expressed as a percentage of the inter-origin distance of each origin pair. The data for *pol2-L439V* have been carried over from Figure 1 for comparison purposes. The intensity of each mutation channel is proportional to the intensity of the colour. Asterisks indicate mutation types significantly deviating from uniformity ($P < 0.01$, χ^2 test). (**E**) Geographical distribution of the density of mutations by type in relation to replication origins generated by the *pol2-M644G* allele in the presence or in the absence of functional MMR.

correction, revealed that these are nearly identical (cosine similarity = 0.992). Comparison of these profiles with the profiles of 30 mutational signatures originally identified in cancers ([43]), detected a very high similarity to COSMIC Signature SBS 14 (Figure 2B and C, cosine similarity 0.93–0.94) and a moderate similarity to SBS20 (0.73–0.76). In agreement with this finding, SBS 14 was originally identified in uterine cancers and low grade gliomas ([2]), and is also observed in cancers carrying both *POLE* mutations and microsatellite instability—the latter a feature of MMR inactivation ([44]). These findings, together with similar results obtained for *pol2-P301R/POL2 msh2Δ/msh2Δ* diploid cells ([33]), further support the notion that SBS 14 is the product of mutagenesis driven by Pol ε exonuclease domain mutants, before MMR correction. Interestingly, SBS 20 has been proposed to arise from mutations in DNA polymerase δ in conjunction with MMR inactivation ([44]).

Replication-strand bias analysis revealed a strong preference for A>C, A>G and G>T mutations on the leading strand, as it was observed in the presence of functional MMR (Figure 2D and Supplementary Figure S4D and F).

Conversely, the strong preference for A>T and the weak preference for G>A mutations on the leading strand was reduced or disappeared when MMR was inactivated (Figure 2D and Supplementary Figure S4D and F), suggesting that while Pol2 mutants equally introduced A>T and T>A mutations, MMR corrected one (T>A) more efficiently than the other (A>T). This pattern of mutagenesis is reminiscent of a yeast Pol ε mutant, *pol2-M644G* ([45]), which carries a mutation in the polymerase domain of Pol ε that creates a 'looser' active site that allows mis-incorporation of dNTPs and rNTPs ([37,46]). To identify similarities and differences between the mutational patterns generated by Pol ε exonuclease and polymerase-domain mutator alleles, we re-analysed previously published mutation accumulation experiments for *pol2-M644G* in the presence or absence of *MSH2* ([45]), and compared these to our results.

In the absence of confounding MMR effects, we found that the activity of each mutation channel was substantially different between exonuclease and polymerase domain mutants, with a lower frequency of C>A mutations and a higher frequency of C>T mutations (Figure 2C and E).

However, the overall replication-strand bias was strikingly similar, with only one major difference in the A>T/T>A channel: while polymerase domain alleles, before (or after) mismatch correction, were more likely to produce A>T mutations by mispairing T:dT more frequently than A:dA (37,45), exonuclease domain alleles produced both types of mis-pair essentially equally (Figure 2E, Supplementary Figure S4G). The striking similarity between mutations introduced in the genome by Pol ε exonuclease and polymerase domain mutants suggest that, with some minor differences, a similar mutagenic process is active in cells carrying either mutant.

A unique signature generated by Pol ε exonuclease domain alleles

We next compared the context in which every class of mutation was observed on the two replication strands. To do this, we pooled all mutations from *pol2–4 msh2Δ* and *pol2-L439V msh2Δ* strains, given their near identity (cosine similarity = 0.99, Supplementary Figure S6A and B) and apparent common aetiology, and compared them with mutations generated by the polymerase domain allele *pol2-M644G* in the absence of *MSH2* (45). This indicated that alteration of either the exonuclease or polymerase domain leads to the mis-insertion of dCTP opposite to the second T of a –TT– dimer template; less frequently the inverse is also observed (dTTP mis-insertion opposite to the C of a –TC– dimer template, Figure 3A, B red boxes, and Supplementary Figure S6C). Overall, these two classes of mutations were more prevalent in exonuclease than in polymerase-domain mutator strains (~34% versus ~10% of all mutations, respectively; $P < 0.01$, χ^2 z-test of given proportions) suggesting that a similar mutagenic mechanism occurs with different intensity in different mutants.

A second shared pattern of mutagenesis between polymerase and exonuclease domain mutants of Pol ε is the mis-insertion of dTTP in front of G or T, often occurring after a pyrimidine in the template (Figure 3C and D, blue boxes). Notably, these classes of mutations were more frequently observed in MA lines with a polymerase-domain mutator allele (~33% of all mutations) than in strains with exonuclease-domain mutator alleles (~19% of all mutations; $P < 0.01$). Other classes of mutation showed very little or no sequence-context specificity, despite their overall relatively high prevalence (Figure 3C and E; unboxed patterns). A notable exception to this was insertions of A in front of the first A after a long T homopolymer that is followed by an AA dimer (TTTTTAA for example; Figure 3D, green boxes). Despite being a relatively uncommon event (~2% of all mutations), this signature was unique to exonuclease-domain mutators, being completely absent from the mutational spectra of *pol2-M644G msh2Δ* cells, and accounted for the observed lack of bias in the A>T/T>A channel that is characteristic of Pol ε exonuclease-domain mutators (Figure 2D).

Pol ε proofreading domain mutations increase the frequency of insertions

Analysis of the number of insertions/deletions accumulated in the absence of MMR, which would otherwise efficiently repair them, revealed that Pol ε exo⁻ introduces insertions ~4.5 times more frequently than wild-type Pol ε does, and that this is further increased two-fold in the presence of Pol ε–L439V (Figure 4A). Analysis of the type of base inserted revealed that in both cases, mutant Pol ε is mainly responsible for the introduction of +T and +A, which overall represent 70–80% of all insertions (Figure 4B, as opposed to ~31% for *msh2Δ* strains). Since these insertions likely arise from the role of Pol ε as the leading strand replicase (38), replication-strand bias analysis suggests that Pol ε–L439V is mostly responsible for +A insertions (Figure 4C). Moreover, these +A insertions on the leading strand tend to occur in the context of a 3–6 nucleotide T homopolymer in the template strand (Figure 4D), suggesting that they originate from polymerase slippage.

Pol ε proofreading activity generates deletions

Differently from what we have observed for insertions, MA analysis revealed that, before MMR correction, the production of short, mostly single-base, deletions was ~50% lower in *pol2–4* cells compared to wild-type *POL2*. This suggests that Pol ε exonuclease activity is directly responsible for essentially half of the deletions produced during normal DNA replication in the absence of MMR which would otherwise repair them. The Pol ε–L439V mutant showed an intermediate deletion rate phenotype, possibly because this mutant could retain partial exonuclease activity, as it was shown for Pol ε–P301R. Analysis of the spectrum of deletions implied that a wild-type replisome largely introduces –T, –A, –TT and –AA deletions, and that Pol ε exonuclease activity contributes to roughly half of these (Figure 4E). Analysis of replication-strand bias also revealed that the frequent – A and –T deletions do not normally show any discernible strand preference. Inactivation of Pol ε exonuclease activity, however, led to a strong bias for –A deletions on the leading strand and corresponding –T deletions on the lagging strand (Figure 4F). These results suggest that both Pol ε exonuclease activity and an unidentified process on the lagging strand (possibly Pol δ exonuclease activity) produce frequent –T deletions, giving rise to no overall replication-strand bias. In *pol2–4* cells, however, it appears that the leading strand branch of this mutagenic pathway is inactivated, generating a –A deletion bias that is the reflection of the –T deletions produced on the lagging strand (Figure 4G).

DISCUSSION

DNA-replication associated hypermutation is a key risk factor for the accumulation of mutations in genes driving colorectal and endometrial cancers, whether arising from mismatch repair (MMR) inactivation or from DNA polymerase ε or δ exonuclease domain mutator (EDM) alleles. While it is clear how MMR inactivation increases mutagenesis, establishing the source of mutations generated by EDM alleles has proven more difficult. The yeast benchmark for EDM alleles, *pol2-P301R*, generates mutations at a much higher rate than the corresponding exonuclease-dead strain (an *ultra*-mutator phenotype) (14), while retaining part of the wild-type exonuclease activity (16). Our results now show that the *ultra*-mutator phenotype is shared
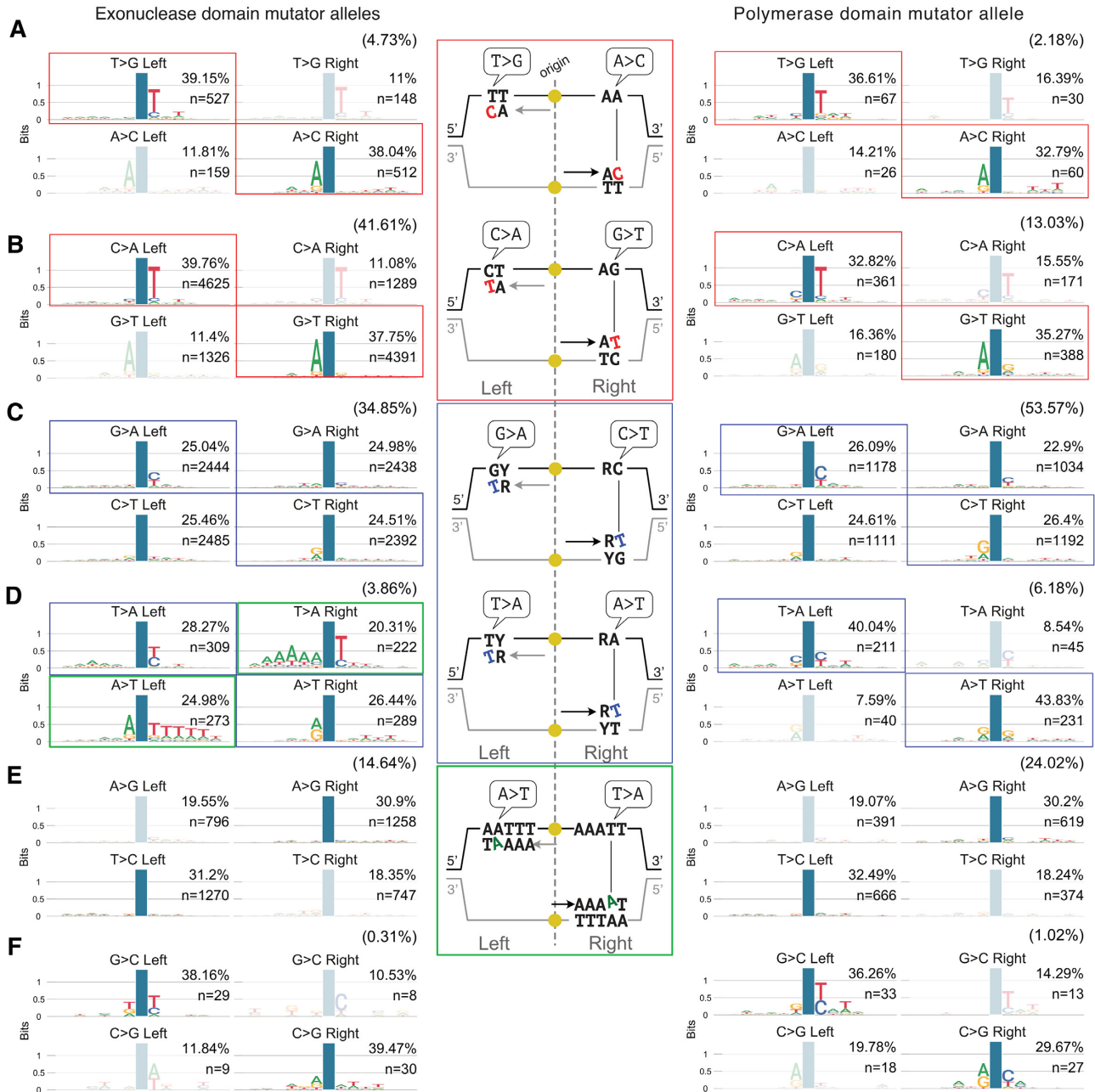
Figure 3. Sequence context of the *de novo* mutations generated in cells with Pol ε exonuclease or polymerase domain mutant alleles before MMR correction. (**A-F**) Sequence context in which each *de novo* mutation class occurs in cells carrying exonuclease domain mutator alleles (*pol2–4* or *pol2-L439V*) or polymerase domain mutator alleles (*pol2-M644G*), in the absence of functional MMR (*msh2Δ*). Mutations, as indicated by the nucleotide change observed on the Watson strand, were categorised by class (**A–F**) with the frequency of each class indicated in brackets. Within each class mutations were further categorised as the two possible types generating that class and whether each mutation type occurred to the right or to the left of the closest replication origin. The position of the mutation is indicated by a blue column. Transparent diagrams indicate less frequent types/positions in each class, which could arise from artefacts created by leading strand replication not precisely terminating at the inter-origin midpoint or by some origins being passively replicated. In the middle, a diagram describing the mis-pairs at the origin of the indicated mutations and contexts is shown.

by other yeast Pol ε EDM alleles orthologous to cancer-associated Pol ε mutations, most notably *pol2-M459K* (human *POLE M444K*), and that increased mutation accumulation in *pol2-ultra* cells compared to *pol2–4* cells also occurs in the absence of MMR activity, strongly suggesting that differential repair of mismatches is not a major source of hypermutation.

Studies of Pol ε protein structure have revealed that the P301R mutation creates a positively charged surface that likely hinders a 3′ mismatched primer end from properly accessing the proofreading domain and could, thus, favour its extension (16,17). Strikingly, *pol2-M459K*—the second strongest mutator in our dataset—also introduces a positive charge in the same area (Supplementary Figure S7), while
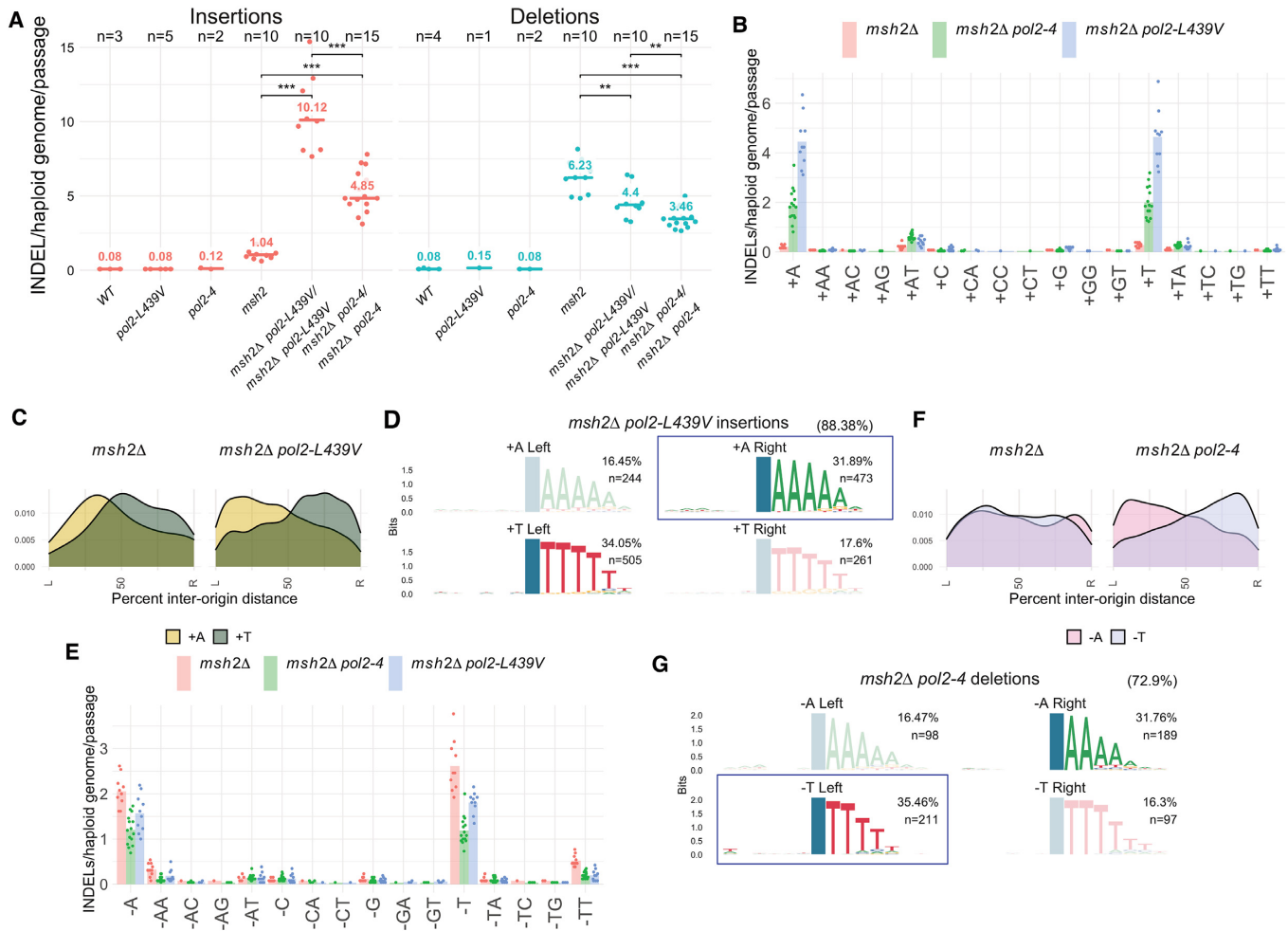
Figure 4. Pol ε *ultra*-mutators introduce +A insertions opposite T homopolymers. (**A**) Rate of insertions and deletions introduced by Polε exo⁻ (*pol2–4*) and the *ultra*-mutator Polε L439V (*pol2-L439V*) in the presence or absence of functional mismatch repair (*msh2Δ*). (**B**) Rate of insertion of different mono- or di-nucleotides. (**C**) Geographical distribution of the density of the two most common types of insertions in relation to replication origins. (**D**) Sequence context in which +A/+T insertions occur in relation to wether they occurred to the right or to the left of the closest replication origin; the position of the mutation is indicated by a blue column; the numbers in brackets indicate the prevalence of this class compared to all insertions; the conserved homopolymer context always appears 3′ of the mutation because INDELs were left-aligned during variant normalization. A box indicates the insertions most frequently observed on the leading strand. (**E**) Rate of deletion of different mono- or di-nucleotides. (**F**) Geographical distribution of the density of the two most common types of deletion in relation to replication origins. (**G**) Sequence context in which –A/–T deletions occur in relation to wether they occurred to the right or to the left of the closest replication origin; the position of the mutation is indicated by a blue column; the numbers in brackets indicate the prevalence of this class compared to all deletions. A box indicates deletions most frequently observed on the lagging strand.

the weaker *ultra*-mutator alleles could also hinder proper DNA strand placement, perhaps by altering Pol ε structure more subtly.

The puzzling *ultra*-mutator phenotype could be explained if most of the mis-pairs observed in cells lacking Pol ε exonuclease activity are not just caused by the inability of this mutant polymerase to remove misincorporated bases, but rather arise from an active mutagenic process whose intensity is heightened by Pol ε EDM alleles. A prediction of this model is that Pol ε *ultra* and Pol ε exo⁻ would produce the same pattern of mutations; and, indeed, the mutational profiles, strand bias, and sequence contexts in which mispairs occur in *pol2-ultra* and *pol2–4* cells are virtually indistinguishable from each other, especially before MMR correction. In this scenario, Pol ε could contain a 'mutagenic activity', normally suppressed by the exonuclease activity,

and activated by EDM alleles. The aminoacid changes encoded by these alleles could, for instance, unmask a translesion synthesis activity of Pol ε, as it has been shown for the increased ability of exonuclease-deficient Pol δ to bypass AP sites (47,48).

We suggest, however, that at least part of the mutations introduced in *pol2-ultra* cells arise from the backtracking of the mutant Pol ε while attempting proofreading. In a wild-type polymerase, this movement melts the primer-template junction until the nascent 3′ end has been inserted in the exonuclease domain for hydrolysis (Figure 5A). In the absence of hydrolysis—or even more so when access to the exonuclease active site has been blocked by *pol2-ultra* mutations—the nascent strand would prevent this movement. In these situations, backtracking could still occur if the nascent strand were to shift backwards and extrude a
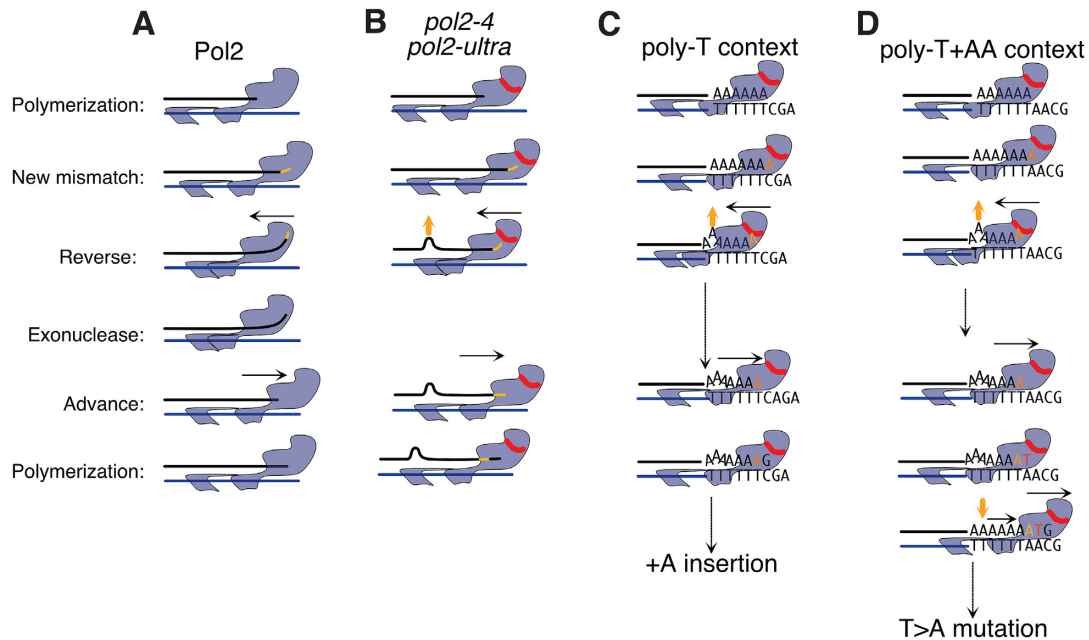
Figure 5. Model for the origin of mutations specifically introduced by Pol ε exonuclease domain mutants. (**A**) In the presence of a terminal mismatch (yellow segment), Pol ε backtracks (black arrow) and repositions the terminal-mismatched primer in the proofreading domain; after hydrolysis the primer is re-annealed and extended. (**B**) In the presence of exonuclease-inactive mutants or *ultra*-mutator mutants backtracking is blocked unless the newly synthesised strand melts, shifts back and extrudes one base (yellow arrow). (**C**) Melting and extrusion can only occur on homopolymer templates (TTTTT), where the shift does not alter correct base-base pairing between template and nascent strand. If the mis-inserted base is an adenine (yellow A), this can pair with the last T of the template and reconstitute a proper base pair that can be extended, generating insertions. (**D**) If a T homopolymer is followed by two A's, then a T would be inserted in front of the first A (red T), but would then move in front of the second, when the base extruded in the homopolymer is re-annealed.

base further upstream (Figure 5B), an activity that would result in the generation of insertions, especially after A:T homopolymers that are easier to melt than G:C ones because of their weaker bonding. At this point, further extension would require the mis-inserted base to form a proper pair with the T template, and thus would occur only when an adenine was mis-inserted in the first place (Figure 5C). In agreement with this model, we found that, in the absence of MMR, *pol2–4* cells and to a greater extent *pol2-L439V* cells, frequently accumulated +A insertions when replicating through T-homopolymers. Furthermore, when two adenines follow the T-homopolymer template, the first one becomes a hotspot for A:A mis-pairs. This would arise from the newly synthesised strand sliding forward after the first base post-mismatch has been introduced, in a model analogous to dislocation mutagenesis (49). This would restore full base pairing, converting the +A insertion into a A:A mispair and generating the observed T>A transversions (Figure 5D). Given the sequence and mis-insertion requirements needed, the above-described events should be comparatively rare; and indeed, insertions represented only ∼9.3% of all the mutations we observed in MMR-deficient cells carrying EDM alleles, while T>A transversions accounted for less than 2%.

With the exception of insertions and T>A transversions the sequence context and strand bias of the remaining mutations, which account for the vast majority (∼89%), closely resembled the context of mutations deposited by a low-fidelity Pol ε mutant carrying a mutation in the polymerase domain (*pol2-M644G*). This strong similarity suggests that

the mutations generated by these two Pol ε variants arise from the same set of mutagenic mechanisms. At the same time, differences in the activity of the single mutation channels between exonuclease and polymerase domain mutants, could point to a different balance of mutagenic processes. In this regard, it has been shown that Pol δ can proofread the errors introduced by Pol ε (29), and does proofread the errors introduced by *pol2-P301R* (50) and *pol2-M644G* (30): a different *in trans* proofreading efficiency in different Pol ε mutants could possibly contribute to the observed differences. Variation in the intensity of different mutation channels could equally arise from the increased dNTP levels, which have been shown to contribute to the mutator phenotype of *pol2-M644G* cells (51), but not that of *pol2-P301R* cells (16). While these data suggest that the majority of mutations introduced in *ultra*-mutator cells are directly introduced by Pol ε, we cannot however exclude that another polymerase is responsible for introducing mis-pairs on the leading strand after Pol ε stalling, for example through the activity of translesion polymerases or Pol δ on the leading strand (52).

In conclusion, our findings have provided further insights into how cells normally guard against mutagenesis during DNA replication, and how specific point mutations in replicative polymerases affect their function to heighten mutation rates and lead to distinctive mutational signatures. Given that DNA replication and replicative polymerases are highly conserved throughout evolution, the effects and mechanisms that we have described likely also operate in cancers with orthologous polymerase alleles.

DATA AVAILABILITY

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Campbell,B.B., Light,N., Fabrizio,D., Zatzman,M., Fuligni,F., de Borja,R., Davidson,S., Edwards,M., Elvin,J.A., Hodel,K.P. *et al.* (2017) Comprehensive analysis of hypermutation in human cancer. *Cell*, **171**, 1042–1056.
2. Alexandrov,L.B., Nik-Zainal,S., Wedge,D.C., Aparicio,S.A.J.R., Behjati,S., Biankin,A.V., Bignell,G.R., Bolli,N., Borg,Å., Børresen-Dale,A.-L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.
3. Eshleman,J.R. and Markowitz,S.D. (1996) Mismatch repair defects in human carcinogenesis. *Hum. Mol. Genet.*, **5**, 1489–1494.
4. Palles,C., Cazier,J.-B., Howarth,K.M., Domingo,E., Jones,A.M., Broderick,P., Kemp,Z., Spain,S.L., Guarino,E., Guarino Almeida,E. *et al.* (2013) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.*, **45**, 136–144.
5. Church,D.N., Briggs,S.E.W., Palles,C., Domingo,E., Kearsey,S.J., Grimes,J.M., Gorman,M., Martin,L., Howarth,K.M., Hodgson,S.V. *et al.* (2013) DNA polymerase ε and δ exonuclease domain mutations in endometrial cancer. *Hum. Mol. Genet.*, **22**, 2820–2828.
6. Cancer Genome Atlas Research Network, Kandoth,C., Schultz,N., Cherniack,A.D., Akbani,R., Liu,Y., Shen,H., Robertson,A.G., Pashtan,I., Shen,R. *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**, 67–73.
7. Brutlag,D. and Kornberg,A. (1972) Enzymatic synthesis of deoxyribonucleic acid. 36. A proofreading function for the 3′ leads to 5′ exonuclease activity in deoxyribonucleic acid polymerases. *J. Biol. Chem.*, **247**, 241–248.
8. Khare,V. and Eckert,K.A. (2001) The 3′ → 5′ exonuclease of T4 DNA polymerase removes premutagenic alkyl mispairs and contributes to futile cycling at O6-methylguanine lesions. *J. Biol. Chem.*, **276**, 24286–24292.
9. Bernad,A., Blanco,L., Lázaro,J.M., Martín,G. and Salas,M. (1989) A conserved 3′—-5′ exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell*, **59**, 219–228.
10. Hogg,M., Osterman,P., Bylund,G.O., Ganai,R.A., Lundström,E.-B., Sauer-Eriksson,A.E. and Johansson,E. (2014) Structural basis for processive DNA synthesis by yeast DNA polymerase ε. *Nat. Struct. Mol. Biol.*, **21**, 49–55.
11. Jain,R., Rajashankar,K.R., Buku,A., Johnson,R.E., Prakash,L., Prakash,S. and Aggarwal,A.K. (2014) Crystal structure of yeast DNA polymerase ε catalytic domain. *PLoS One*, **9**, e94835.
12. Reddy,M.K., Weitzel,S.E. and Hippel,von, P.H. (1992) Processive proofreading is intrinsic to T4 DNA polymerase. *J. Biol. Chem.*, **267**, 14157–14166.
13. Capson,T.L., Peliska,J.A., Kaboord,B.F., Frey,M.W., Lively,C., Dahlberg,M. and Benkovic,S.J. (1992) Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. *Biochemistry*, **31**, 10984–10994.
14. Kane,D.P. and Shcherbakova,P.V. (2014) A common cancer-associated DNA polymerase ε mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. *Cancer Res.*, **74**, 1895–1901.
15. Morrison,A., Bell,J.B., Kunkel,T.A. and Sugino,A. (1991) Eukaryotic DNA polymerase amino acid sequence required for 3′—-5′ exonuclease activity. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 9473–9477.
16. Xing,X., Kane,D.P., Bulock,C.R., Moore,E.A., Sharma,S., Chabes,A. and Shcherbakova,P.V. (2019) A recurrent cancer-associated substitution in DNA polymerase ε produces a hyperactive enzyme. *Nat. Commun.*, **10**, 374.
17. Parkash,V., Kulkarni,Y., Beek,Ter, J., Shcherbakova,P.V., Kamerlin,S.C.L. and Johansson,E. (2019) Structural consequence of the most frequently recurring cancer-associated substitution in DNA polymerase ε. *Nat. Commun.*, **10**, 373.
18. Putnam,C.D. and Kolodner,R.D. (2010) Determination of gross chromosomal rearrangement rates. *Cold Spring Harb. Protoc.*, **2010**, doi:10.1101/pdb.prot5492.
19. Zackrisson,M., Hallin,J., Ottosson,L.-G., Dahl,P., Fernandez-Parada,E., Ländström,E., Fernandez-Ricaud,L., Kaferle,P., Skyman,A., Stenberg,S. *et al.* (2016) Scan-o-matic: High-Resolution microbial phenomics at a massive scale. *G3 (Bethesda)*, **6**, 3003–3014.
20. Puddu,F., Herzog,M., Selivanova,A., Wang,S., Zhu,J., Klein-Lavi,S., Gordon,M., Meirman,R., Millan-Zambrano,G., Ayestaran,I.X.I. *et al.* (2019) Genome architecture and stability in the Saccharomyces cerevisiae knockout collection. *Nature*, **573**, 416–420.
21. Alexandrov,L.B., Nik-Zainal,S., Wedge,D.C., Campbell,P.J. and Stratton,M.R. (2013) Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.*, **3**, 246–259.
22. Blokzijl,F., Janssen,R., van Boxtel,R. and Cuppen,E. (2018) MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med*, **10**, 33–11.
23. Wagih,O. (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*, **33**, 3645–3647.
24. Morrison,A., Johnson,A.L., Johnston,L.H. and Sugino,A. (1993) Pathway correcting DNA replication errors in Saccharomyces cerevisiae. *EMBO J.*, **12**, 1467–1473.
25. Zhu,Y.O., Siegal,M.L., Hall,D.W. and Petrov,D.A. (2014) Precise estimates of mutation rate and spectrum in yeast. *PNAS*, **111**, E2310–E2318.
26. Lynch,M., Sung,W., Morris,K., Coffey,N., Landry,C.R., Dopman,E.B., Dickinson,W.J., Okamoto,K., Kulkarni,S., Hartl,D.L. *et al.* (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9272–9277.
27. Nishant,K.T., Wei,W., Mancera,E., Argueso,J.L., Schlattl,A., Delhomme,N., Ma,X., Bustamante,C.D., Korbel,J.O., Gu,Z. *et al.* (2010) The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. *PLoS Genet.*, **6**, e1001109.

28. Zwolshen,J.H. and Bhattacharjee,J.K. (1981) Genetic and biochemical properties of thialysine-resistant mutants of Saccharomyces cerevisiae. *J. Gen. Microbiol.*, **122**, 281–287.

29. Flood,C.L., Rodriguez,G.P., Bao,G., Shockley,A.H., Kow,Y.W. and Crouse,G.F. (2015) Replicative DNA polymerase δ but not ε proofreads errors in Cis and in Trans. *PLos Genet.*, **11**, e1005049.

30. Bulock,C.R., Xing,X. and Shcherbakova,P.V. (2020) DNA polymerase δ proofreads errors made by DNA polymerase ε. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 6035–6041.

31. Barbari,S.R., Kane,D.P., Moore,E.A. and Shcherbakova,P.V. (2018) Functional analysis of cancer-associated DNA polymerase ε variants in *Saccharomyces cerevisiae*. *G3 (Bethesda)*, **8**, 1019–1029.

32. Williams,L.N., Herr,A.J. and Preston,B.D. (2013) Emergence of DNA polymerase ε antimutators that escape error-induced extinction in yeast. *Genetics*, **193**, 751–770.

33. Tracy,M.A., Lee,M.B., Hearn,B.L., Dowsett,I.T., Thurber,L.C., Loo,J., Loeb,A.M., Preston,K., Tuncel,M.I., Ghodsian,N. *et al.* (2020) Spontaneous polyploids and antimutators compete during the evolution of saccharomyces cerevisiae mutator cells. *Genetics*, **215**, 959–974.

34. Galati,M.A., Hodel,K.P., Gams,M.S., Sudhaman,S., Bridge,T., Zahurancik,W.J., Ungerleider,N.A., Park,V.S., Ercan,A.B., Joksimovic,L. *et al.* (2020) Cancers from novel pole-mutant mouse models provide insights into polymerase-mediated hypermutagenesis and immune checkpoint blockade. *Cancer Res.*, **80**, 5606–5618.

35. Li,H.-D., Cuevas,I., Zhang,M., Lu,C., Alam,M.M., Fu,Y.-X., You,M.J., Akbay,E.A., Zhang,H. and Castrillon,D.H. (2018) Polymerase-mediated ultramutagenesis in mice produces diverse cancers with high mutational load. *J. Clin. Invest.*, **128**, 4179–4191.

36. Hodel,K.P., Sun,M.J.S., Ungerleider,N., Park,V.S., Williams,L.G., Bauer,D.L., Immethun,V.E., Wang,J., Suo,Z., Lu,H. *et al.* (2020) POLE mutation spectra are shaped by the mutant allele identity, its abundance, and mismatch repair status. *Mol. Cell*, **78**, 1166–1177.

37. Pursell,Z.F., Isoz,I., Lundström,E.-B., Johansson,E. and Kunkel,T.A. (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science (New York, N.Y.)*, **317**, 127–130.

38. Nick McElhinny,S.A., Gordenin,D.A., Stith,C.M., Burgers,P.M.J. and Kunkel,T.A. (2008) Division of labor at the eukaryotic replication fork. *Mol. Cell*, **30**, 137–144.

39. Marsischky,G.T. and Kolodner,R.D. (1999) Biochemical characterization of the interaction between the Saccharomyces cerevisiae MSH2-MSH6 complex and mispaired bases in DNA. *J. Biol. Chem.*, **274**, 26668–26682.

40. Lang,G.I., Parsons,L. and Gammie,A.E. (2013) Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3 (Bethesda)*, **3**, 1453–1465.

41. Morrison,A. and Sugino,A. (1994) The $3' \rightarrow 5'$ exonucleases of both DNA polymerases delta and epsilon participate in correcting errors of DNA replication in Saccharomyces cerevisiae. *Mol. Gen. Genet.*, **242**, 289–296.

42. Li,H.-D., Lu,C., Zhang,H., Hu,Q., Zhang,J., Cuevas,I.C., Sahoo,S.S., Aguilar,M., Maurais,E.G., Zhang,S. *et al.* (2020) A PoleP286R mouse model of endometrial cancer recapitulates high mutational burden and immunotherapy response. *JCI Insight*, **5**, 2311.

43. Tate,J.G., Bamford,S., Jubb,H.C., Sondka,Z., Beare,D.M., Bindal,N., Boutselakis,H., Cole,C.G., Creatore,C., Dawson,E. *et al.* (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **47**, D941–D947.

44. Haradhvala,N.J., Kim,J., Maruvka,Y.E., Polak,P., Rosebrock,D., Livitz,D., Hess,J.M., Leshchiner,I., Kamburov,A., Mouw,K.W. *et al.* (2018) Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. *Nat. Commun.*, **9**, 1746–1749.

45. Lujan,S.A., Clausen,A.R., Clark,A.B., MacAlpine,H.K., MacAlpine,D.M., Malc,E.P., Mieczkowski,P.A., Burkholder,A.B., Fargo,D.C., Gordenin,D.A. *et al.* (2014) Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.*, **24**, 1751–1764.

46. Nick McElhinny,S.A., Kumar,D., Clark,A.B., Watt,D.L., Watts,B.E., Lundström,E.-B., Johansson,E., Chabes,A. and Kunkel,T.A. (2010) Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.*, **6**, 774–781.

47. Meng,X., Zhou,Y., Zhang,S., Lee,E.Y.C., Frick,D.N. and Lee,M.Y.W.T. (2009) DNA damage alters DNA polymerase delta to a form that exhibits increased discrimination against modified template bases and mismatched primers. *Nucleic Acids Res.*, **37**, 647–657.

48. Hirota,K., Yoshikiyo,K., Guilbaud,G., Tsurimoto,T., Murai,J., Tsuda,M., Phillips,L.G., Narita,T., Nishihara,K., Kobayashi,K. *et al.* (2015) The POLD3 subunit of DNA polymerase δ can promote translesion synthesis independently of DNA polymerase ζ. *Nucleic Acids Res.*, **43**, 1671–1683.

49. Kunkel,T.A. and Soni,A. (1988) Mutagenesis by transient misalignment. *J. Biol. Chem.*, **263**, 14784–14789.

50. Bulock,C.R., Xing,X. and Shcherbakova,P.V. (2020) Mismatch repair and DNA polymerase δ proofreading prevent catastrophic accumulation of leading strand errors in cells expressing a cancer-associated DNA polymerase ε variant. *Nucleic Acids Res.*, **48**, 9124–9134.

51. Williams,L.N., Marjavaara,L., Knowels,G.M., Schultz,E.M., Fox,E.J., Chabes,A. and Herr,A.J. (2015) dNTP pool levels modulate mutator phenotypes of error-prone DNA polymerase ε variants. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E2457–E2466.

52. Garbacz,M.A., Cox,P.B., Sharma,S., Lujan,S.A., Chabes,A. and Kunkel,T.A. (2019) The absence of the catalytic domains of Saccharomyces cerevisiae DNA polymerase ε strongly reduces DNA replication fidelity. *Nucleic Acids Res.*, **47**, 3986–3995.