

RESEARCH ARTICLE

Mechanisms of glycosylase induced genomic instability

Daniel E. Eyler¹, Kylie A. Burnham¹, Thomas E. Wilson^{2,3}, Patrick J. O'Brien^{1*}

1 Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI, United States of America, **2** Department of Pathology, University of Michigan Medical School, Ann Arbor, MI, United States of America, **3** Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, United States of America

* pjobrien@umich.edu



OPEN ACCESS

Citation: Eyler DE, Burnham KA, Wilson TE, O'Brien PJ (2017) Mechanisms of glycosylase induced genomic instability. PLoS ONE 12(3): e0174041. <https://doi.org/10.1371/journal.pone.0174041>

Editor: Robert W. Sobol, University of South Alabama Mitchell Cancer Institute, UNITED STATES

Received: January 10, 2017

Accepted: March 2, 2017

Published: March 23, 2017

Copyright: © 2017 Eyler et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Sequencing data has been deposited in the NCBI Short Read Archive under accession number SRP068914. hp_caller is available on github (https://github.com/dangenet/hp_caller).

Funding: This work was supported by grants from the American Cancer Society (RSG-11-148-01-DMC) and the National Institutes of Health (R01 GM108022). The University of Michigan DNA sequencing core receives support from the National Cancer Institute of the National Institutes of Health (P30CA046592). The funders had no role

Abstract

Human alkyladenine DNA glycosylase (AAG) initiates base excision repair (BER) to guard against mutations by excising alkylated and deaminated purines. Counterintuitively, increased expression of AAG has been implicated in increased rates of spontaneous mutation in microsatellite repeats. This microsatellite mutator phenotype is consistent with a model in which AAG excises bulged (unpaired) bases, altering repeat length. To directly test the role of base excision in AAG-induced mutagenesis, we conducted mutation accumulation experiments in yeast overexpressing different variants of AAG and detected mutations via high-depth genome resequencing. We also developed a new software tool, hp_caller, to perform accurate genotyping at homopolymeric repeat loci. Overexpression of wild-type AAG elevated indel mutations in homopolymeric sequences distributed throughout the genome. However, catalytically inactive variants (E125Q/E125A) caused equal or greater increases in frameshift mutations. These results disprove the hypothesis that base excision is the key step in mutagenesis by overexpressed wild-type AAG. Instead, our results provide additional support for the previously published model wherein overexpressed AAG interferes with the mismatch repair (MMR) pathway. In addition to the above results, we observed a dramatic mutator phenotype for N169S AAG, which has increased rates of excision of undamaged purines. This mutant caused a 10-fold increase in point mutations at G:C base pairs and a 50-fold increase in frameshifts in A:T homopolymers. These results demonstrate that it is necessary to consider the relative activities and abundance of many DNA replication and repair proteins when considering mutator phenotypes, as they are relevant to the development of cancer and its resistance to treatment.

Introduction

The base excision repair (BER) pathway is responsible for repairing a wide variety of oxidized and alkylated base lesions. In human cells, an estimated 10,000 lesions per day are processed through the BER pathway [1]. The multi-step BER pathway is initiated by DNA glycosylases, which search the genome and excise damaged bases. Alkyladenine DNA glycosylase (AAG; also known as MPG, methylpurine DNA glycosylase) is the sole glycosylase in its superfamily

in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

and is thus distinct from other human DNA glycosylases. In addition it is the principal glycosylase for a remarkably broad range of lesions, including hypoxanthine, xanthine, 1,*N*⁶-ethenoadenine (ϵ A), and 3- and 7- methyladenine and methylguanine [2]. As such, AAG plays an important role in protecting the genome against the detrimental effects of these lesions.

While base lesions are often mutagenic and sometimes replication-blocking, the downstream intermediates produced during BER are themselves mutagenic and genotoxic [3–9]. When DNA damage levels are high, or glycosylase expression is elevated, AAG may produce abasic sites faster than the downstream pathway can process them [10,11]. Indeed, overexpression of AAG in yeast was reported to cause point mutations [10] and frameshift mutations in homopolymeric repeats [12,13]. Since microsatellite instability contributes to carcinogenesis [14–17] and high AAG expression has been observed in humans [18] and in cancer cell lines [19,20], the possibility that AAG may induce mutations is relevant to human disease.

The mechanisms of glycosylase-induced mutagenesis are likely to vary depending on the class of mutations. Point mutations can arise from the replication of abasic sites [6,10,21–23], and glycosylases with increased activity toward undamaged bases (i.e., increased gratuitous repair [24]) confer point mutator phenotypes, presumably by increasing the number of abasic sites [10,25,26]. The mechanism of glycosylase-induced frameshift mutagenesis in homopolymers is less well understood. During replication, polymerase slippage gives rise to altered structures with a bulged (unpaired) base flanked by duplex DNA [27]. If undetected, these nascent frameshift mutations become fixed during subsequent rounds of replication, with deletion events being more common [27–36]. Overexpression of wild-type AAG elevated deletion frameshifts by 10-fold but only increased insertion frameshifts by 3-fold [10,12,13]. The authors of this study proposed a mismatch repair (MMR) competition model in which AAG bound to bulged, undamaged bases in homopolymers and prevented efficient repair of these errors by MMR. However, since AAG catalytic activity appeared to be required for mutagenesis, and AAG overexpression remained mutagenic in strains deficient for MMR, additional mechanisms could be operative. Given the differential effects of AAG overexpression on -1 and +1 frameshift rates [13] and the ability of AAG to remove bulged bases leading to deletion by the BER pathway [37], we hypothesized that excision of bulged bases in homopolymers would generate the observed bias. This hypothesis, which we term the bulge-excision model, is differentiated from other models for glycosylase-induced mutagenesis in microsatellites by the fact that it requires a catalytically active glycosylase.

We directly examined the relationship between the base excision activity of AAG and mutagenesis, particularly in homopolymers. Our approach was to overexpress a set of human AAG variants in haploid yeast and perform mutation accumulation experiments followed by high throughput sequencing [38]. Haploid yeast are ideally suited to investigate the mechanism of glycosylase-induced homopolymer mutagenesis because yeast lack a homolog of human AAG. This simplifies the interpretation from heterologous introduction of this enzyme, and the compact genome is also amenable to calling mutations in microsatellites. This experimental system has been experimentally validated by several previous single locus studies [10,12,13]. We chose AAG variants for study that alter its biochemical properties in order to test the proposed mechanisms of mutagenesis. The E125Q mutant binds DNA similarly to the wild-type enzyme, but it is devoid of catalytic activity [39–41] and does not protect yeast against alkylation damage [25,39,42]. The Y162A mutant retains catalytic activity once a lesion is bound, but it is defective in base flipping. This defect makes it ineffectual in finding base lesions in the context of excess DNA [43], and indeed this mutant does not protect yeast from exogenous alkylation damage [39]. Finally, the N169S mutant has altered specificity, with wild-type activity toward damaged bases but elevated activity to excise undamaged purines, especially guanine [2,25,44].

We observed homopolymer mutator phenotypes in the wild-type, E125Q, and N169S strains, as well as a point mutator phenotype at G:C base pairs in the N169S strain. The observation that catalytic activity is not required for frameshift mutagenesis disproves the bulge-excision model. Instead, this finding provides additional support for the MMR competition model proposed by Samson and colleagues. While there was no correlation between mutagenesis by wild-type AAG and its catalytic properties, the catalytic properties of the N169S mutant were correlated with an increase in mutations. The largest increases in frameshift mutations were observed in the N169S strain and can be explained by this mutant enzyme's increased excision of undamaged purines. Our data support the models of gratuitous repair by N169S AAG and competition with endogenous repair pathways. Overall, the results emphasize the importance of enzymatic specificity in restricting the deleterious consequences of inappropriate DNA repair.

Results

Experimental design for mutation accumulation

To distinguish between mechanisms of glycosylase-induced mutagenesis, we conducted mutation accumulation experiments in haploid *S. cerevisiae* under conditions similar to those previously reported for AAG-induced mutagenesis [10,13,25,44]. In cells that are fully proficient for repair, AAG overexpression has a dominant mutator phenotype. We compared cell lines containing empty vector to those that were expressing wild-type or mutant AAG (E125Q, Y162A, and N169S) at a high level. Eight lines for each construct were passaged for ~1,000 generations with bottlenecks every 20–22 generations, and unique mutations were detected by high-depth genome resequencing (~50-fold depth; see Fig A in [S1 File](#) for the analysis pipeline). Using haploid yeast allows sensitive detection of mutations even in difficult-to-sequence regions, such as homopolymers. As haploid cells are more vulnerable to negative selection against mutations, we tailored the duration of the experiment to ensure that a small number of mutations are accumulated in each line, thus minimizing the effects of negative selection.

Point mutations

Point mutation counts were clearly increased in the N169S strain but were not detectably elevated in the other strains ([Fig 1A](#)). Two mutation accumulation lines had increased mutation counts relative to the other lines within that strain, one of the E125Q lines and one of the N169S lines. These lines appear to have acquired secondary mutator phenotypes not related to AAG overexpression, because the frequency and the spectra of mutations were distinct from the related lines ([Fig D in S1 File](#)). While neither line has a mutation in an obvious DNA repair or replication gene that easily explains their respective mutator phenotypes, there are candidate mutator mutations in each line ([S2 File](#) and [S3 File](#)). Three lines in the N169S strain had lower numbers of mutations than the rest of the N169S lines. Two of these lines had acquired different inactivating mutations in the N169S AAG construct; one introduced a premature termination codon (E268X), while the other mutated an invariant and critical arginine (R182G) [39,42,45]. The remaining line duplicated the majority of its genome except for the right arm of chromosome III. These five lines (two mutators, two null mutations in AAG, and one pseudodiploid) were excluded from further analysis. The individual lines used in this study are listed in [Table B in S1 File](#).

Cells carrying empty vector had point mutation rates of 3×10^{-10} per base pair per generation ([Table 1](#)), consistent with other studies of haploid yeast [29,32,46–48]. We did not observe point mutator phenotypes in strains expressing wild-type, E125Q, and Y162A, although small (<3-fold) effects would have escaped detection in our experiment. The strain expressing

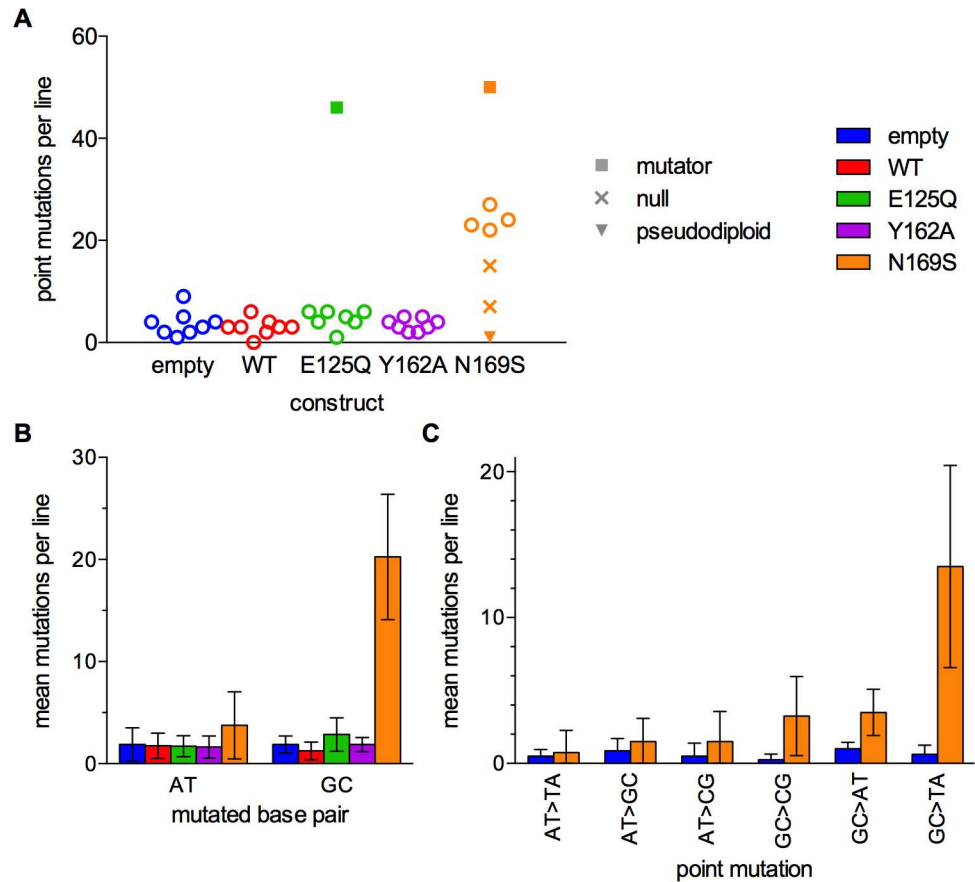


Fig 1. Point mutations in AAG overexpression lines. (A) The number of point mutations in each mutation accumulation line. Lines with additional mutator phenotypes are indicated with closed squares. Lines with inactivating mutations in the AAG gene are labeled “null” and marked with an x. The pseudodiploid line is marked with a closed triangle. (B) The mean number of mutations at A:T and G:C base pairs per strain. (C) The mean number of point mutations in each category for the empty vector strain and the N169S strain. Error bars are 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0174041.g001>

N169S AAG had a 10-fold increased point mutation rate at G:C base pairs, with a 21-fold increase in G>T transversions. No patterns were identified in the flanking sequences, demonstrating that N169S AAG causes mutations in variety of sequence contexts (Fig F in S1 File).

Table 1. Point mutation rates by strain and category ($\times 10^{10}$ per bp per generation).

	empty	WT	E125Q	Y162A	N169S
mutation	mean (95% CI)	mean (95% CI)	mean (95% CI)	mean (95% CI)	mean (95% CI)
AT>TA	0.7 (0.2–1.7)	0.3 (0.0–1.2)	0.7 (0.2–1.9)	0.8 (0.3–1.9)	1.0 (0.2–2.9)
AT>GC	1.1 (0.5–2.4)	1.1 (0.5–2.4)	1.1 (0.4–2.4)	0.5 (0.1–1.4)	2.0 (0.7–4.3)
AT>CG	0.7 (0.2–1.7)	0.8 (0.3–1.9)	0.4 (0.0–1.3)	0.8 (0.3–1.9)	2.0 (0.7–4.3)
AT	2.4 (1.4–4.0)	2.3 (1.2–3.8)	2.2 (1.2–3.9)	2.1 (1.1–3.6)	4.9 (2.7–8.1)
GC>CG	0.5 (0.1–1.9)	0.3 (0.0–1.5)	0.6 (0.1–2.2)	1.3 (0.4–3.1)	7.0 (4.0–12)
GC>AT	2.1 (0.9–4.1)	1.1 (0.3–2.7)	2.4 (1.0–4.7)	1.6 (0.6–3.4)	7.0 (4.0–12)
GC>TA	1.3 (0.4–3.1)	1.3 (0.4–3.1)	3.0 (1.4–5.5)	1.1 (0.3–2.7)	28 (21–37)
GC	3.9 (2.2–6.5)	2.6 (1.3–4.8)	6.0 (3.7–9.3)	3.7 (2.0–6.2)	43 (34–53)
total	3.0 (2.0–4.3)	2.4 (1.5–3.6)	3.7 (2.5–5.2)	2.7 (1.8–4.0)	19 (16–24)
lines	8	8	7	8	4

<https://doi.org/10.1371/journal.pone.0174041.t001>

We compared the ratio of nonsynonymous to synonymous (K_a/K_s) mutations in protein coding regions to evaluate whether there was any evidence of negative selection (Table D in [S1 File](#)). These ratios did not significantly deviate from unity, which is not surprising given the low level of mutations.

Frameshift mutations in homopolymers

Homopolymers are difficult to sequence accurately for the same reason that they are difficult for cells to replicate accurately: polymerases are prone to slippage [27]. To avoid artifacts in mutation calling, most software tools are designed to be extremely conservative in mutation calling in or near homopolymers. The samtools package called fewer indels in homopolymers than were expected on the basis of reporter assays [13,28] (Fig H in [S1 File](#)). Other callers tested were either similarly insensitive, or generated large numbers of false positives, based on visual examination of alignments. To overcome this obstacle, we developed a new homopolymer caller. This program, `hp_caller`, exploited specific characteristics of our dataset, namely the high sequencing depth in each line and the fact that most lines have the same genotype at any given locus, to enable sensitive and conservative homopolymer calling. The caller compares the distribution of observed homopolymer lengths in a single sample to the average distribution of the remaining samples, and calculates the probability that the single sample has the same modal observed length as the average distribution. Examples of the raw data analyzed by `hp_caller` are shown in [Fig 2A](#); the sample in red is called as a mutant, while the remaining lines are called as reference. The caller can identify multiple homopolymer alleles at one locus, as long there is one predominant allele present in most strains ([Fig 2B](#)). Genome-wide, samples called as mutant have the same average read depth as samples called as reference, indicating that mutation calling by `hp_caller` was not an artifact of low coverage ([Fig 2C](#)). Histograms for every mutation called by `hp_caller` are found in [S4 File](#), while the algorithm is described further in Supplementary Methods ([S1 File](#)).

Seventy-nine unique mutations were identified in homopolymers by `hp_caller`, approximately double the number of mutations called by samtools (Fig H in [S1 File](#)). Homopolymer mutation counts per line were generally low (≤ 11) and were almost exclusively in A:T homopolymers ([Fig 3A](#)). The only two mutations that were observed in G:C homopolymers were in the N169S strain (histograms are included in File S4). The mean mutation rates for A:T homopolymers with lengths between 7 and 16 nt were increased in the WT, E125Q, and N169S strains ([Fig 3B](#) and [Table 2](#)). This trend held over all lengths analyzed ([Fig 3C](#)). The majority of events were -1 frameshifts in each strain, regardless of the catalytic activity of the glycosylase or its specificity for damage ([Fig 4](#)).

Western blot analysis confirmed that the different AAG variants were expressed at similar levels (Fig B in [S1 File](#)), but we were concerned that some variants could be selected against during serial passaging. Therefore, we used fluctuation analysis to assay glycosylase-induced frameshift mutagenesis using the *LYS2* polyA reporter allele in these strains. Mutation rates were similar for unpassaged and passaged lines in each strain ([Fig 5](#)), indicating that average levels of AAG-induced mutagenesis did not change during passaging.

It was surprising that E125Q and wild-type AAG exhibited similar mutator phenotypes ([Figs 3](#) and [5](#)), because a previous study reported that the E125A variant of AAG almost completely lacked a mutator phenotype [13]. To resolve this discrepancy, we independently created the E125A mutation in AAG and used fluctuation analysis to evaluate whether or not E125A AAG is a mutator in the *LYS2* reporter strain ([Fig 5](#)). The similar mutator phenotype accompanying overexpression of either E125Q or E125A provides strong evidence that glycosylase activity is not required for frameshift mutagenesis and suggests that these two mutant

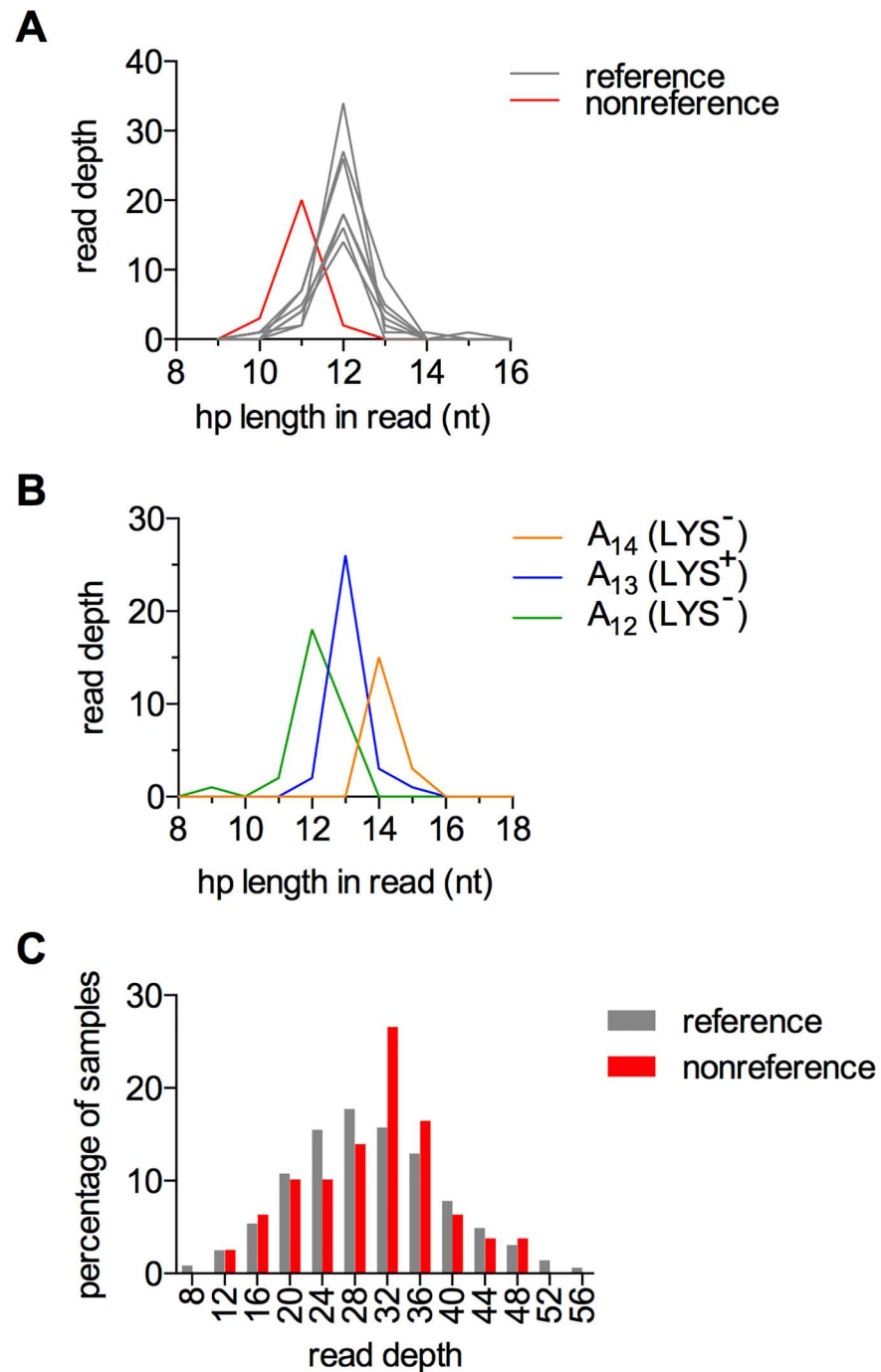


Fig 2. High-confidence genotyping at homopolymer loci with hp_caller. A) The hp_caller program discriminates between a mutant sample (red) and the other seven non-mutant samples in the same strain. Histograms plotting read depth at each observed homopolymer length are shown for the wild-type AAG strain for the T_{12} homopolymer at chromosome VIII:50981–50993. The mutation is in line 22556 (see Table B in [S1 File](#) for line identification). B) Genotype calls at the *LYS2* polyA reporter locus are consistent with the lysine requirements of the three strains shown ($A_{14} = 22537$, $A_{13} = 22538$, $A_{12} = 22539$). C) Read depth distributions at all loci with mutations were comparable for strains called as reference (gray) and non-reference (red).

<https://doi.org/10.1371/journal.pone.0174041.g002>

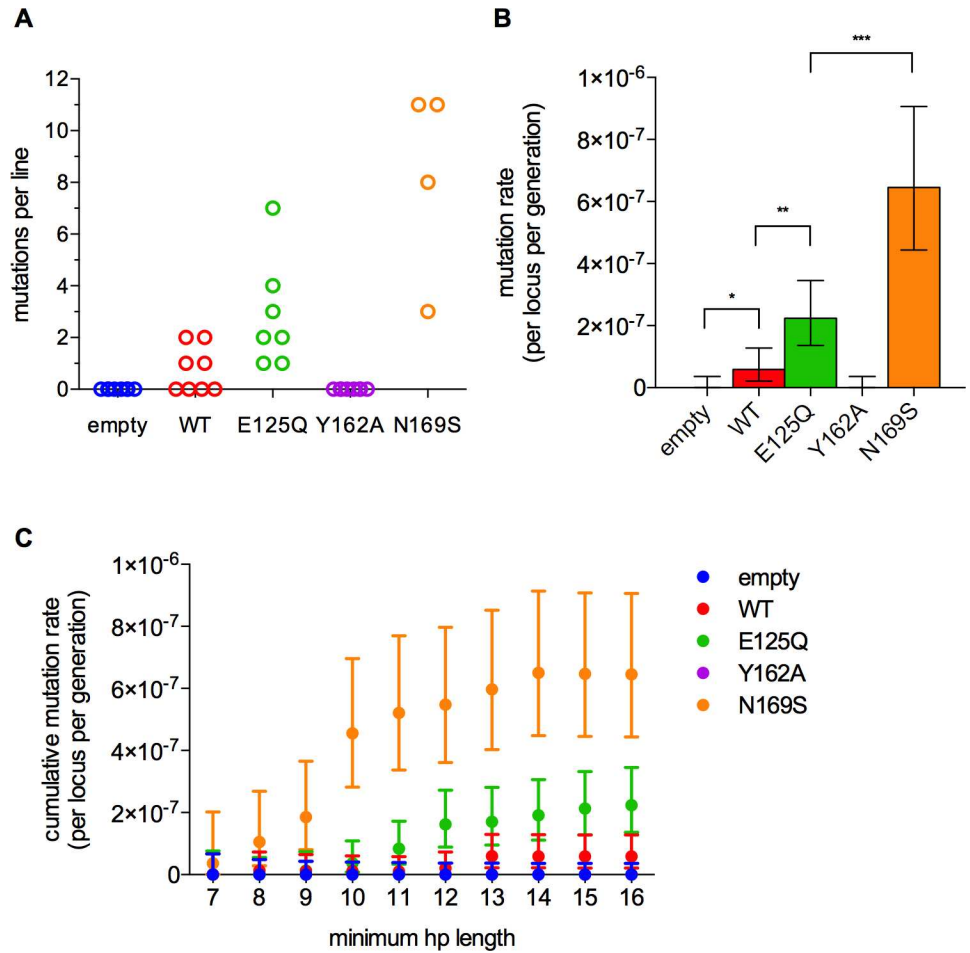


Fig 3. Frameshift mutations in A:T homopolymers are increased in the wild-type, E125Q, and N169S strains. (A) The number of frameshift mutations in A:T homopolymers in each line. (B) The average rate of frameshift mutations at all A:T homopolymers with lengths between 7 and 16 nt. Error bars are 95% confidence intervals calculated by the Clopper-Pearson method. Asterisks indicate p-values: *, p < 0.05; **, p < 0.005; ***, p < 0.0005; Fisher's exact test for indicated pairwise comparisons. (C) Mutation rates in A:T homopolymers as a function of homopolymer length.

<https://doi.org/10.1371/journal.pone.0174041.g003>

proteins behave similarly. Whereas both are catalytically inactive [41], the E125A mutant was reported to be unable to bind to bulged, undamaged bases [13].

To directly compare the DNA binding affinity of the E125Q and E125A mutant proteins, and to compare this affinity to the other AAG variants, we used electrophoretic mobility shift assays (EMSAs). Under the conditions tested, the wild-type and mutant AAG proteins bound

Table 2. Mutation rates in A:T homopolymers with lengths between 7 and 16 nt (per locus per generation × 10⁸).

construct	indels	deletions	insertions
	mean (95% CI)	mean (95% CI)	mean (95% CI)
empty	0.0 (0.0–3.6)	0.0 (0.0–3.6)	0.0 (0.0–3.6)
WT	5.9 (2.2–13)	3.9 (1.1–10)	2.0 (0.2–7.1)
E125Q	23 (15–36)	21 (13–33)	2.2 (0.3–8.1)
Y162A	1.0 (0.0–5.5)	0.0 (0.0–3.6)	1.0 (0.0–5.5)
N169S	65 (44–91)	51 (33–75)	14 (5.5–28)

<https://doi.org/10.1371/journal.pone.0174041.t002>

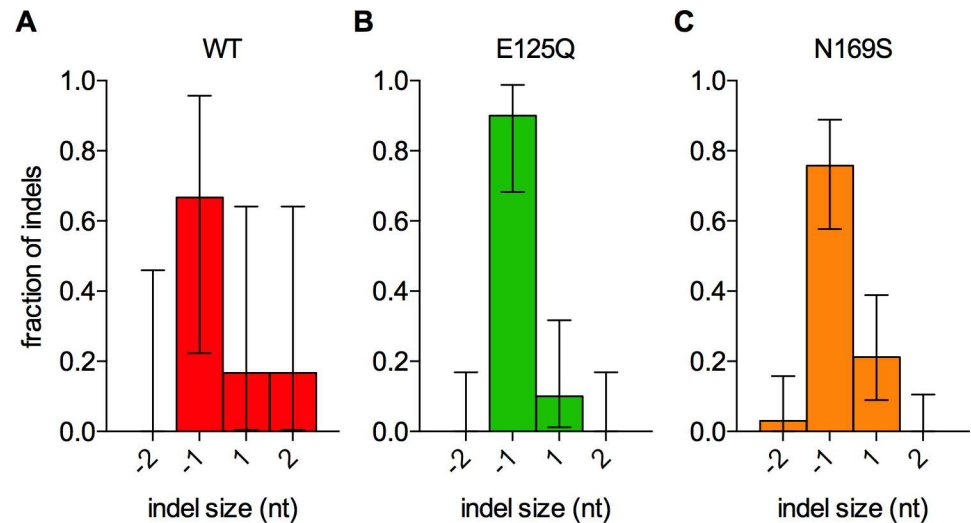


Fig 4. Glycosylase-induced indels in A:T homopolymers are predominantly -1 frameshifts. The fraction of indels of a given size is plotted for the (A) wild-type, (B) E125Q, and (C) N169S AAG strains. Error bars are 95% confidence intervals. The empty vector and Y162A strains are not shown because they have no indel mutations.

<https://doi.org/10.1371/journal.pone.0174041.g004>

to the A-bulge-containing DNA, which mimics a single nucleotide polymerase slipping event, with an observed affinity that is approximately 3-fold tighter for the bulge than for perfectly matched duplex DNA of the same sequence (Fig 6). As there are many overlapping binding sites on the undamaged oligonucleotide duplex, this equilibrium binding result indicates that the microscopic K_d for binding to a bulged site is much tighter than for binding to a typical undamaged site. Nevertheless, this assay allows us to conclude that the E125Q and E125A mutant proteins have almost identical affinity for bulged DNA. The N169S and wild-type AAG also bound specifically to bulged DNA, albeit with weaker preference for the bulged DNA (Fig 6). In contrast, Y162A AAG was unable to recognize the bulged DNA. The observation that Y162A binds with the same affinity to bulged and duplex DNA is explained by non-specific duplex binding with both oligonucleotides.

No strong patterns in the locations of glycosylase-induced mutations

We investigated the possibility that glycosylase-induced mutagenesis might be biased towards loci with specific characteristics. We analyzed flanking sequences (Fig F in S1 File) [49], annotated genomic features (Fig J in S1 File) [50], replication timing (Fig K in S1 File) [51], and transcription (Fig L in S1 File) [52,53]. These analyses were restricted to strains with appreciable numbers of each class of mutation (96 point mutations in the N169S strain and 53 frame-shift mutations in the E125Q and N169S strains). We did not observe any correlations between mutated loci and the characteristics we tested, indicating that glycosylase-induced mutagenesis is not strongly dependent on any of these features. The absence of strong correlations is consistent with the idea that AAG has broad access to the genome, but given the small number of mutations that were observed, it is possible that there are some biases that were not detected.

Variation in plasmid copy numbers

Lastly, we analyzed the copy numbers of the endogenous 2-micron circle and the pYES2 expression vector. The copy number of the pYES2 expression vector appeared to be lower in

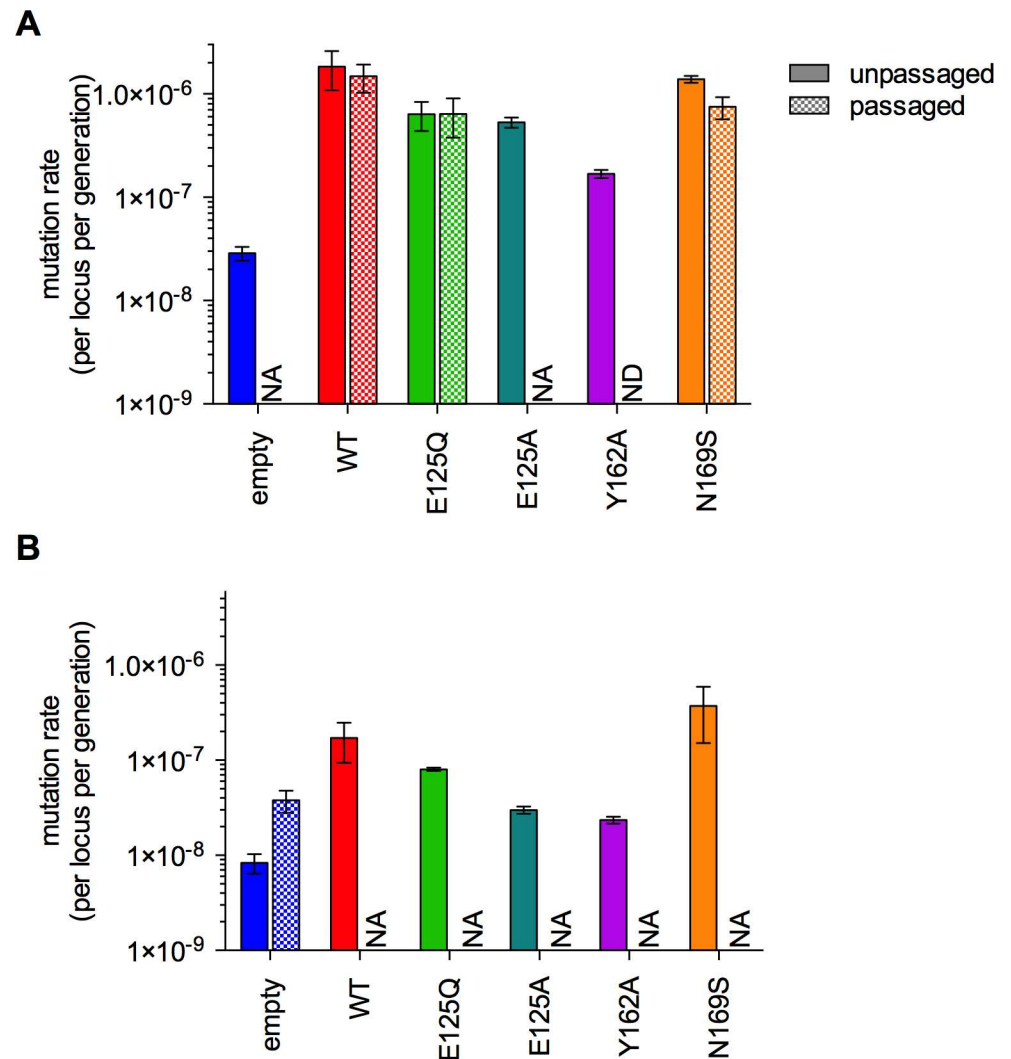


Fig 5. Fluctuation analysis reveals similar mutator phenotypes for E125Q and E125A AAG and preservation of mutator phenotypes during passaging. Frameshift mutation rates were measured using the *LYS2* poly-A reporter allele for -1 frameshift events in E134-derived lines (A) and +1 frameshift events in E133-derived lines (B). Mutation rates were measured by fluctuation analysis in unpassaged strains and for a subset of the lines in each strain after passaging, specifically, the passaged lines with the fewest number of mutations. Error bars indicate the standard error of the mean. Fluctuation analysis could not be performed in ending lines marked with “NA” (not applicable), because these lines did not contain the appropriate *LYS2* reporter allele (see Table B in S1 File). Fluctuation analysis was not performed (“ND”) in the passaged Y162A lines since this construct did not induce a mutator phenotype in the mutation accumulation experiments.

<https://doi.org/10.1371/journal.pone.0174041.g005>

the E125Q and N169S strains, which had the strongest mutator phenotypes (Fig 7A). In comparison, the endogenous 2-micron circle copy number did not seem to be affected (Fig 7B). To test this hypothesis, we grouped all the strains with low or non-existent mutator phenotypes (empty vector, wild-type, and Y162A) and the two strains with mutator phenotypes (E125Q and N169S). The copy number of the pYES2 expression vector was significantly decreased in the mutator strains relative to the non-mutator strains, while the endogenous 2-micron circle was unaffected (Fig 7C). The inverse correlation between expression plasmid copy number and mutator severity suggests that passaged cells experienced negative selection if they harbored a mutator glycosylase.

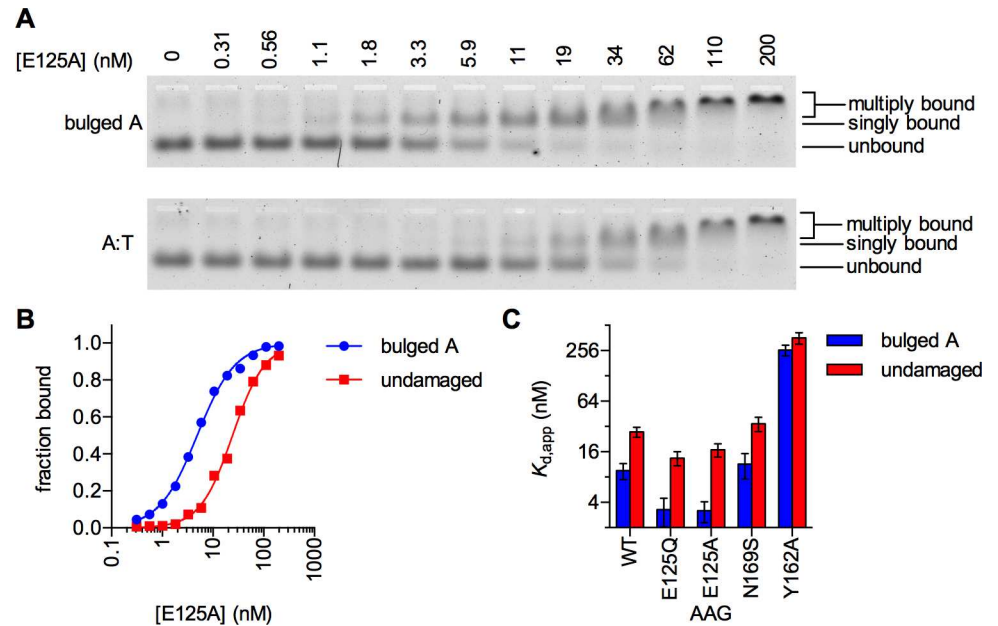


Fig 6. The wild-type and mutant AAG enzymes exhibit different affinities for bulged and duplex DNA. (A) Sample EMSA gel image for E125A. Singly and multiply bound complexes were combined in calculating the fraction of DNA bound, because these species are not cleanly separated. (B) Example binding curves for the E125A mutant. (C) Apparent K_D values were determined by EMSA for the indicated substrates, containing a centrally located, bulged undamaged adenosine, or fully duplex DNA (undamaged). $n = 3$; error bars are SEM. E125Q and E125A both exhibit tighter binding than wild-type to a bulged site, whereas the Y162A exhibits much weaker binding and does not distinguish between bulged and intact duplexes. Binding data for all constructs is shown in Fig M in S1 File.

<https://doi.org/10.1371/journal.pone.0174041.g006>

Discussion

This study was designed to examine the contributions of catalytic activity and specificity to glycosylase-induced mutagenesis in an unbiased manner throughout the genome. Several AAG variants induced frameshift mutations in homopolymers, while N169S AAG induced point and frameshift mutations. We consider first the mechanisms for frameshift mutagenesis that apply to most AAG variants and second the mechanisms that are specific to the N169S mutant.

Mechanisms of AAG-induced frameshift mutagenesis

The prevailing model for glycosylase-induced frameshift mutagenesis is the MMR competition model (Fig 8A; [13]) in which glycosylase binding to bulged nucleotides blocks access to MMR enzymes. This mechanism predicts that inactive glycosylases will be mutagenic. In contrast to MMR competition, the alternative bulge-excision model requires catalytically active glycosylases. We tested this model, which involves excision and deletion of the bulged base to preferentially generate -1 frameshift events, using wild-type and inactive (E125Q) AAG (Fig 8B; [37,54]).

To test the bulge-excision model, we used previously characterized variants of AAG that lacked catalytic activity but retained ability to bind to bulges (E125Q/E125A), or that retained glycosylase activity but lost the ability to bind specifically to bulges (Y162A), and compared these to the wild-type AAG. Consistent with the previous report [13], overexpression of wild-type AAG caused an increase in frameshift mutations in A:T homopolymeric regions. In contrast to the earlier study [13], we found that inactive AAG also acted as a frameshift mutator in A:T homopolymers.

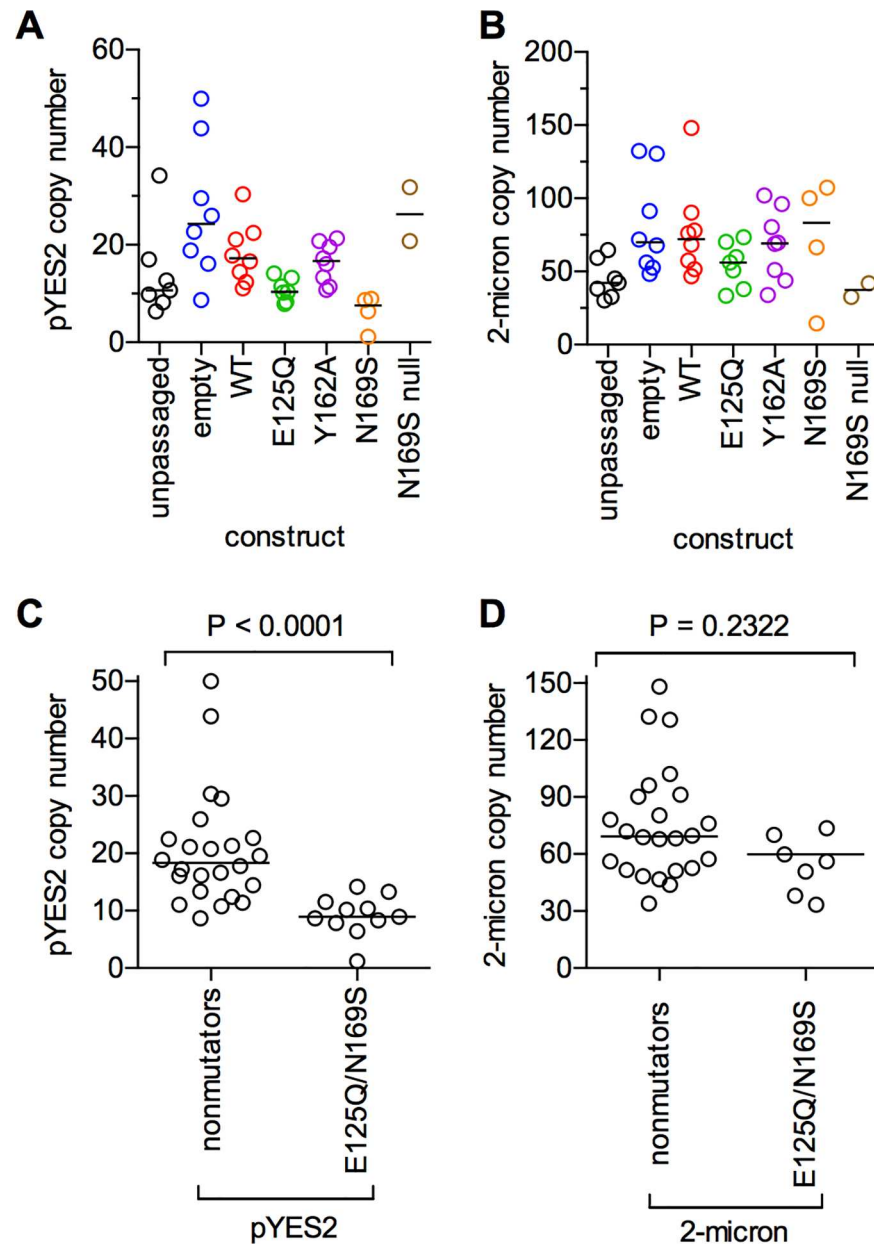
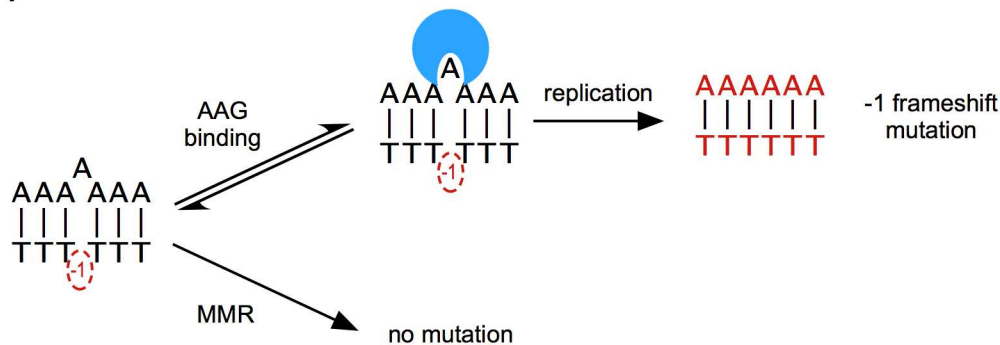


Fig 7. Strains expressing mutator glycosylase alleles have decreased pYES2 expression plasmid copy numbers. (A) pYES2 copy numbers for the parental (unpassaged) strains, as well as each of the ending strains. The two lines with inactivating mutations in the AAG gene are shown as the strain “N169S null”. (B) Endogenous 2-micron circle copy numbers for the parental strains and ending strains. (C) U-test comparing the distribution of plasmid copy numbers between the mutator strains (E125Q + N169S) and the nonmutator strains (empty + WT + Y162A) for the pYES2 expression vector. (D) U-test comparing the copy numbers of the endogenous 2-micron circle for mutator and nonmutator strains. Horizontal lines in all panels indicate the median of each distribution.

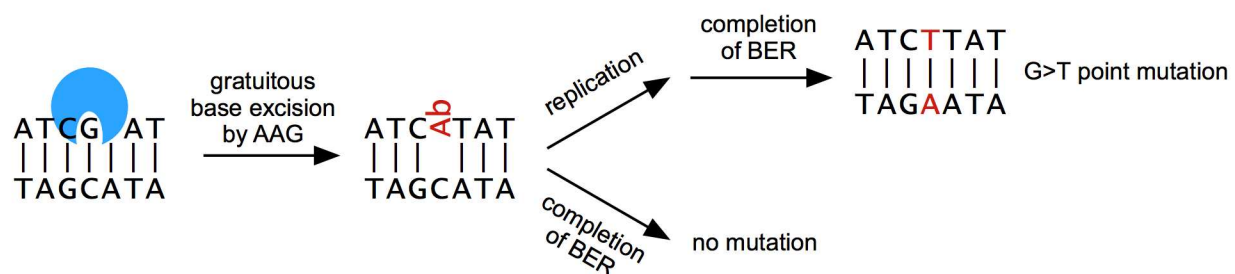
<https://doi.org/10.1371/journal.pone.0174041.g007>

Since the previous study utilized E125A AAG and we used E125Q AAG, we directly compared the biochemical and cellular properties of these two mutants. The two mutants had similar affinities for single A-bulges (Fig 6A), similarly impaired (10^5 -fold) glycosylase activity [41], were similarly expressed (Fig B in S1 File), and conferred similar increases in -1 frameshifts in vivo (Fig

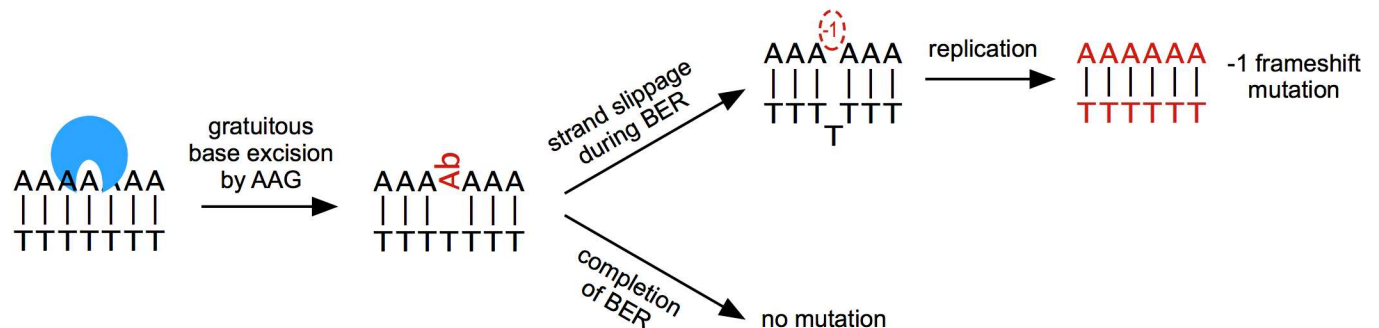
A) MMR inhibition model for frameshift mutations.



B) Gratuitous repair model for point mutations.



C) Gratuitous repair model for frameshift mutations.



D) Bulge/excision model for frameshift mutations.

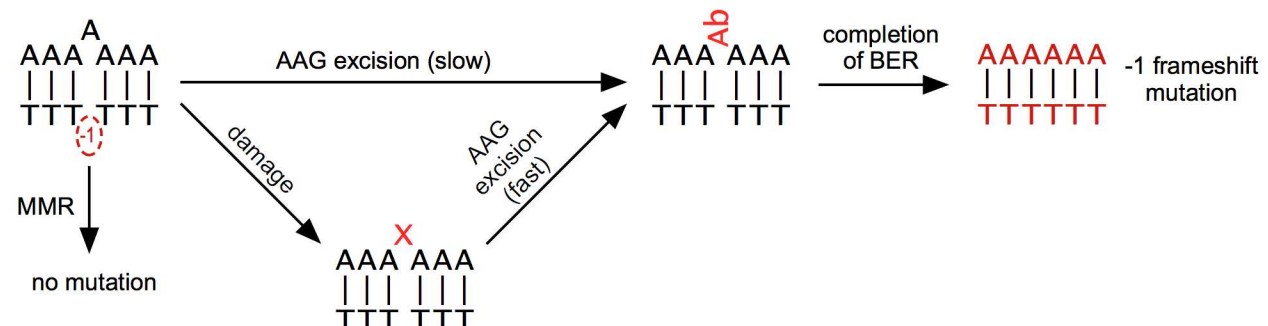


Fig 8. Models for glycosylase-induced mutagenesis. For simplicity, only the most relevant pathways and intermediates are shown. Replication (not shown) occasionally gives rise to bulged species in models A and B (see Fig O in S1 File for more detailed models). (A) The MMR interference model

proposes that glycosylase binding to bulged, undamaged bases prevents MMR from repairing the bulged base. The pathway shown is for -1 frameshift events, but +1 frameshifts can be stabilized by the same mechanism. (B) The bulge-excision model proposes that bulged bases are deleted by the BER pathway and could be elevated by base damage. This pathway produces only -1 frameshifts, because nascent +1 frameshifts (not shown) would be faithfully repaired. (C) The gratuitous excision model postulates that removal of undamaged bases is mutagenic. In the case of the N169S mutant, most of the gratuitous excision occurs at undamaged guanines. In yeast, abasic sites are paired with A and C during replication. Pairing with C is non-mutagenic in this case, and is not shown. (D) Gratuitous excision in homopolymers effectively increases the number of times that a homopolymer is replicated, thus increasing frameshift rates. Both -1 and +1 frameshift events would occur by this mechanism, but only a -1 frameshift is shown.

<https://doi.org/10.1371/journal.pone.0174041.g008>

5A). Since the E125Q and E125A mutants behaved similarly in four different experiments, we conclude that inactive AAG confers a frameshift mutator phenotype.

The Y162A variant was investigated as an additional test of the role of bulge binding in mutagenesis. This mutation destabilizes the flipped-out complex by a factor of greater than 100-fold, without changing the rate of ϵ A excision [43]. Here we show that the Y162A mutant is incapable of binding to a single nucleotide bulge site (Fig 6A). Weakened bulge binding predicts that Y162A would not interfere with other DNA repair processes, and indeed, overexpression of Y162A AAG was less mutagenic than other AAG constructs in both mutation accumulation (Fig 3) and fluctuation analysis (Fig 5) experiments.

Since the strong bias towards deletions proved to be an intrinsic property of DNA replication in yeast [28–32,34–36,55], and mutagenesis was not dependent on catalytic activity, the bulge-excision model can be rejected. Frameshift mutagenesis by inactive AAG is instead consistent with a broad class of models in which AAG binding to nascent bulged nucleotides during replication prevents their repair. Nascent bulged nucleotides can be repaired by a number of pathways, including polymerase proofreading [55] and MMR. However, previous work demonstrated that competition with MMR is the most frequent event, because the number of AAG-induced frameshift mutations is greatly decreased in MMR deficient backgrounds [13]. This interpretation is also supported by the fact that AAG and the MSH2-MSH6 complex have similar K_D values for bulged bases (8 and 28 nM [56], respectively).

Mutagenesis associated with gratuitous repair

Wild-type AAG has low, but detectable activity to excise normal, undamaged purines from DNA [57], and it has been reported that overexpression confers a modest increase in point mutations [10]. The gratuitous excision of undamaged guanine is suppressed by a steric clash with N169 in the active site, and mutation to N169S opens up the active site to increase glycosylase activity by 3-fold and 30-fold for excision of A and G, respectively [2,25,44]. The N169S variant increased point mutations in G:C pairs by 10-fold and in A:T pairs by 2-fold (Fig 1 and Table 1). The largest class of point mutations observed in this case was G→T transversions, which could be explained by replication of an abasic site to insert an A following gratuitous repair (Fig 8C; [23,58–60]). This specific mechanism of mutagenesis by abasic sites has been previously invoked to explain the mutator phenotypes induced by other glycosylases with high activity on undamaged bases [10,26,61].

The N169S variant also induced high levels of frameshift mutations in A:T homopolymers. Unlike other variants tested, the N169S frameshift mutation spectrum included a significant increase in +1 frameshift events. This result was surprising because base excision by a glycosylase cannot insert a base. To explain this result, we propose that N169S gratuitously initiates BER in homopolymers, as it does in heterogeneous sequences. Since completion of BER requires DNA polymerization, there is an opportunity for slippage and realignment with concomitant introduction of a bulged (unpaired) base (Fig 8D). In addition to A:T frameshift mutations, the N169S strains also had two -1 deletions in G:C homopolymers, while all other strains had no mutations in G:C homopolymers. This small number of events is indicative of a

large increase in N169S-induced frameshifts in G:C homopolymers, because there are many fewer sites in the genome (Fig G in [S1 File](#)).

Role of selection in mutation accumulation experiments

In comparison to mutator phenotypes in microsatellite-unstable human cancers [62], the mutator phenotype induced by overexpression of AAG was fairly mild. Nevertheless, there are two mechanisms by which selective pressure could influence the calculated mutation rates. The first would be if deleterious mutations were selected against, and therefore the number of mutations observed would be an underestimate. We mitigated the effects of negative selection by growing yeast on plates and keeping the mutation load low in each passaged strain. Comparison of synonymous to nonsynonymous mutations suggested that the point mutations that were observed in the genome were not selected against (Table D in [S1 File](#)). However, the second type of selective pressure would be if the cells were to reduce or eliminate the expression of the glycosylase variant. This would also lead to an underestimate of the mutator phenotype. There was clear evidence for selection against the AAG expression vectors. Both E125Q and N169S AAG showed significant reduction in the copy number of the multicopy expression plasmid (Fig 7). The selection against the N169S construct was particularly striking, because the enzyme was inactivated by homoplasmic point mutations in 25% (2/8) of the independently passaged cell lines. These observations are a cautionary tale for researchers performing mutation accumulation experiments or overexpressing putatively innocuous proteins (e.g., DNA repair proteins) from yeast high-copy plasmids.

Implications of glycosylase-induced mutagenesis

This work used unbiased mutation accumulation experiments to validate previous observations of mutator phenotypes induced by overexpression of DNA glycosylases. Galactose-driven expression of AAG in yeast yields approximately 13,000 molecules per cell [42], which is similar to the higher end of the range estimated for normal human cells (500–10,000 molecules per cell) [63,64]. Although the total mutations per cell would therefore be comparable, the absolute mutation rate is expected to be smaller in human cells because the genome is much larger than the yeast genome. The finding that wild-type AAG expression is barely mutagenic suggests that AAG has evolved strategies to avoid causing mutations even when highly expressed. This is in stark contrast to the much stronger mutator phenotypes that have been observed for overexpression of yeast [10] and bacterial alkylation repair enzymes [57].

The E125Q and N169S variants were more mutagenic than the wild-type AAG, underscoring the potential of single amino acid changes to confer mutator phenotypes. These mutator effects were dominant, because the yeast were not DNA repair deficient. Whereas many loss-of-function alleles are tolerated due to redundancies between repair pathways, the DNA repair proteins are susceptible to gain-of-function mutator phenotypes. We attribute this to their privileged access to genomic DNA. Given that humans have a multitude of DNA repair enzymes, there are many potential mutations that could confer a mutator phenotype. Indeed, a growing number of mutator glycosylase alleles have been discovered. For AAG, the Y127I allele also confers a strong frameshift mutator phenotype [13]. The D239Y allele of endonuclease III homolog (NTHL1) stably binds, but fails to excise damaged bases [65], and the G199S allele in thymine DNA glycosylase stably binds abasic sites, blocking repair [66]. In addition, the N204D and Y147A mutants of uracil DNA glycosylase allow for gratuitous excision of undamaged pyrimidines [26], and the A145G and H151A/Q mutations allow thymine DNA glycosylase to initiate gratuitous repair at normal A:T sites in DNA [61]. Mutator alleles of pol β have also been observed in human cancers [9,67].

While the mutator phenotypes we observed in our experiment were relatively minor [62], with increases in mutation rates ranging from 5- to 50-fold, they may still be sufficient to increase an individual's lifetime cancer risk. In humans, alleles that confer strong mutator phenotypes are readily detected in genome-wide association studies of cancer patients. However, because of the many different alleles that generate less dramatic mutator phenotypes, these studies may not identify all mutator alleles that contribute to carcinogenesis. Continued biochemical and genetic study of allelic variants in DNA repair proteins is thus important to comprehensively identify sources of genetic instability. This information may ultimately be useful for assessing individual lifetime cancer risk and for guiding cancer treatment choices.

Materials and methods

Yeast strains and media

The yeast reporter strains (E133 and E134) were provided by D. Gordenin [28]. The pYES2 constructs for expression of N169S and Y162A AAG were obtained from M. Wyatt [25,44]. The construct for expression of wild-type AAG was previously described [42]. Site-directed mutagenesis was used to introduce the following point mutations: G373C for E125Q; T484G and A485C for Y162A; and A506G for N169S. Each construct was verified by sequencing of the entire AAG coding sequence. Plasmids were transformed into yeast by the standard lithium acetate procedure. Single clones were picked and maintained as patches until the start of the mutation accumulation experiment. At the start of the mutation accumulation experiment, each starting strain was streaked for isolation on CSM-URA with 2% galactose. A single colony was picked and restreaked on CSM-URA with 2% galactose (and used to prepare a glycerol stock). From this plate, eight independent colonies were picked and streaked separately on CSM-URA with 2% galactose. Each line was passaged independently, in parallel, with bottle-necks every 20–22 generations, for a total of 1000 generations. The number of generations at each plating was verified by independently resuspending 4–5 colonies in water and determining the number of cells per colony by measuring OD₆₀₀. At the end of passaging, each line was patched onto CSM-URA with 2% dextrose. From this plate, samples were frozen in YPD with 15% glycerol at -80°C for preservation. Overnight cultures were grown for each line and used to inoculate 25 mL of CSM-URA with 2% dextrose at OD₆₀₀ 0.1; these cultures were grown to an OD₆₀₀ of 0.5 before harvesting. DNA was purified with a Qiagen DNeasy Blood and Tissue Kit (#69504). Library preparation (mean insert size of 500 bp) and Illumina sequencing (100 bp paired end) were performed by the University of Michigan DNA Sequencing Core Facility.

Alignment and mutation calling

An overview of data processing is shown in Fig A in S1 File. Similar to other pipelines [29,47,48], reads were aligned to the yeast genome (SacCer3) using bowtie2 [68] (parameters in Table A in S1 File). Alignments were marked as not unique if the next-best alignment had a score within seven points of the best alignment (based on the distribution of XS scores calculated by bowtie and the physical meaning of a score greater than -7). Duplicates were marked with MarkDuplicates in picard version 1.110 (<http://broadinstitute.github.io/picard>). Reads were realigned around indels using the IndelRealigner tool in GATK version 3.1-1-g07a4bf8 [69]. Samtools v0.1.19+ was used to construct a pileup and call mutations, assuming haploid genomes [70] (parameters in Table A in S1 File). Calls were filtered for unique mutations using custom Perl scripts. Unique mutations were defined as those occurring in no more than three lines out of the 47 lines sequenced (see the allele frequency spectrum in Fig C, part A, S1 File). Additionally, loci with unique mutations were required to have data in at least 44 out of the 47 total lines to avoid false positives. We also filtered out clusters of point mutations

occurring within 20 bp of one another in a single line. Finally, point mutations were then filtered on QUAL score with a threshold of 100 based on the distribution of scores and visual examination of mutations in each QUAL score bin. Indels in homopolymers were called with the custom software package `hp_caller`, which is described in more detail below and in the Supplemental Methods in [S1 File](#). At the end of processing, mutations were evenly distributed across chromosomes and show no evidence of tightly spaced clusters (Fig E in [S1 File](#)).

Mutation calling at homopolymers

We did not achieve the desired level of sensitivity and accuracy for mutations at homopolymers using `samtools` or other tools. Hence, we developed a set of tools that utilize the characteristics of our experiment to call homopolymers with as much sensitivity and accuracy as possible. The programs are written in Perl and are available at https://github.com/dangenet/hp_caller.

Briefly, at each homopolymer locus, `hp_caller` compares the distribution of homopolymer lengths in reads from a given sample to the average distribution of homopolymer lengths in all the remaining samples. The assumption is that the majority of samples share the same genotype at any given locus. `Hp_caller` performs a binomial test of the hypothesis that the sample distribution has a different mode than the locus distribution. If that p-value is below the threshold set for mutation calling, then the sample is called as a mutant. In addition to this calculation, there are a number of other requirements for calling, such as minimum and maximum read depths in the sample and for all samples at the locus, and a minimum locus quality (LQ) score. A more detailed description may be found in the Supplemental Methods section of [S1 File](#).

Unique mutations were identified using custom scripts as described for point mutations; the allele frequency spectrum for homopolymer mutations is shown in Fig C, part B in [S1 File](#). Custom scripts were also used to compare `samtools` homopolymer calls to `hp_caller` mutation calls. Correlations with genomic features, transcription, and replication timing were analyzed using `bedtools` [71] and custom scripts.

Calculation of mutation rates in homopolymers

The mean mutation frequency is the number of mutations divided by the number of callable loci, and the mean mutation rate is the mutation frequency divided by the number of generations (~1000). The Clopper-Pearson method was used to calculate 95% confidence intervals assuming a binomial distribution. In many cases, there were no events observed at a particular length in a particular construct. For this reason, we typically performed analyses on length windows, such as A:T homopolymers with lengths between 7 and 16 nt.

Fluctuation analysis

Fluctuation analysis was performed as described, using 10 cultures per strain per experiment [13]. Mutation rates were estimated for each experiment using the maximum likelihood method as implemented in `FALCOR` [72]. The reported mutation rates and errors are the mean and standard error of mutation rate estimates from at least three independent experiments for each strain. A comparison of mutation rates determined by mutation accumulation and fluctuation analysis experiments is shown in Fig I in [S1 File](#).

Electrophoretic mobility shift assays

Oligonucleotide sequences were as follows: forward ϵ A oligo, 5' -DyLight647-TAGCAT CCT ϵ A CCTTCTCTC; forward A oligo, 5' -DyLight647-TAGCATCCTACCTTCTCTC;

reverse_T oligo, GAGAGAAGG**T**AGGATGCTA; reverse_b1 5' -GAGAGAAGGAGGATGCTA. The abbreviation ϵ A indicates the nucleotide containing the base 1,N⁶-ethenoadenine. Oligonucleotides were ordered from either Glen Research or IDT. All oligonucleotides were gel purified before use; concentrations were determined by absorbance at 260 nm. The following substrates were prepared from the above-listed oligonucleotides: forward_εA:reverse_T; forward_A:reverse_T; forward_A:reverse_b1. Oligonucleotides were mixed at a ratio of 1:1.5 forward:reverse and annealed by slow cooling from 95°C to 4°C in 10 mM MES, 50 mM NaCl, pH 6.5.

All purified AAG proteins were the Δ80 construct lacking the amino-terminal 79 amino acids, and were purified as previously described [41]. The active concentration of enzyme was determined by titration of AAG with tight binding εA-DNA using the EMSA assay. Active AAG (0–1000 nM) was incubated with 200 nM εA:T substrate in binding buffer (25 mM HEPES pH 7.5, 200 mM NaCl, 5% glycerol, 0.2 mg/mL BSA, 1 mM DTT, 1 mM EDTA) at 4°C for one minute, and then immediately loaded onto a 2% agarose gel in 1X TBE running at 3 V/cm in the cold room. After electrophoresis for 60 minutes, gels were scanned on a Typhoon Trio+ imager using the 633 nm laser, the 670BP30 emission filter, and the +3 mm focal plane. Bands were quantitated using ImageQuant TL (GE Healthcare) and the fraction of substrate bound was calculated according to the equation:

$$f_{bound} = \frac{\sum(\text{signal in shifted bands})}{\sum(\text{signal in all bands})}$$

Binding affinities for undamaged duplex DNA and undamaged bulged adenines were determined by titrating active enzyme. We fit the fraction of substrate bound to a one-site binding model with a Hill coefficient:

$$y = \frac{x^h}{K_d^h + x^h}$$

The K_d determined by this method is an apparent K_d that reflects nonspecific binding of AAG to the multiple sites on undamaged duplex DNA as well as specific binding to the bulged adenine for the A:b1 substrate. In this application, the Hill coefficients do not have a physical meaning [73]; rather they are correction factors that compensate for the simplistic mathematical model. We observed Hill coefficients between 1.3 and 1.4 for the bulged DNA and between 1.2 and 1.8 for the undamaged DNA with standard errors of ~0.2 for all fits.

Supporting information

S1 File. Supporting text, figures, and tables. This file contains supplementary methods describing hp_caller, supporting Figs A through O, and tables A through D which are referred to in the main text.

(PDF)

S2 File. Excel spreadsheet of unique mutations in strain 22561, the E125Q mutator strain.

(XLSX)

S3 File. Excel spreadsheet of unique mutations in strain 22579, the N169S mutator strain.

(XLSX)

S4 File. Histograms for homopolymer loci called as having mutations by hp_caller. Each histogram shows the distribution of homopolymer lengths in all reads at the locus (“LOCUS”) and the distribution of homopolymer lengths in reads from samples called as mutants (indicated by sample numbers). The read depth of the locus distribution is normalized to the read

depth in the sample with the fewest reads at the locus, which is not necessarily the mutant sample. The read depths for the mutant sample distributions are not normalized. The title for each panel indicates the chromosome and start position of the homopolymer, as well as the hp_caller LQ score for the locus. The two mutant calls at G:C homopolymers have “GC” appended to their titles; all other loci are A:T homopolymers. The legend indicates the sample name and the hp_caller LQ score for mutant samples.

(PDF)

Acknowledgments

We thank D. Gordenin for the kind gift of the E133 and E134 reporter strains and M. Wyatt for the pYES2 plasmids for expression of N169S and Y162A AAG. We thank T. Anens and D. Lyons for preliminary fluctuation analysis experiments.

Author Contributions

Conceptualization: DE TW PO.

Funding acquisition: PO.

Investigation: DE KB.

Methodology: DE TW PO.

Software: DE.

Supervision: PO.

Visualization: DE.

Writing – original draft: DE.

Writing – review & editing: DE TW PO.

References

1. Barnes DE, Lindahl T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu Rev Genet.* 2004; 38: 445–76. <https://doi.org/10.1146/annurev.genet.38.072902.092448> PMID: 15568983
2. O'Brien PJ, Ellenberger TE. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J Biol Chem.* 2004; 279: 9750–7. <https://doi.org/10.1074/jbc.M312232200> PMID: 14688248
3. Ma W, Resnick M, Gordenin DA. Apn1 and Apn2 endonucleases prevent accumulation of repair-associated DNA breaks in budding yeast as revealed by direct chromosomal analysis. *Nucleic Acids Res.* 2008; 36: 1836–1846. <https://doi.org/10.1093/nar/gkm1148> PMID: 18267974
4. Ma W, Panduri V, Sterling JF, Van Houten B, Gordenin DA, Resnick MA. The transition of closely opposed lesions to double-strand breaks during long-patch base excision repair is prevented by the coordinated action of DNA polymerase delta and Rad27/Fen1. *Mol Cell Biol.* 2009; 29: 1212–1221. <https://doi.org/10.1128/MCB.01499-08> PMID: 19075004
5. Ma W, Westmoreland JW, Gordenin DA, Resnick MA. Alkylation base damage is converted into repairable double-strand breaks and complex intermediates in G2 cells lacking AP endonuclease. *PLoS Genet.* 2011; 7.
6. Xiao W, Chow BL, Hanna M, Doetsch PW. Deletion of the MAG1 DNA glycosylase gene suppresses alkylation-induced killing and mutagenesis in yeast cells lacking AP endonucleases. *Mutat Res—DNA Repair.* 2001; 487: 137–147. PMID: 11738940
7. O'Brien PJ. Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem Rev.* 2006; 106: 720–52. <https://doi.org/10.1021/cr040481v> PMID: 16464022

8. Sobol RW, Kartalou M, Almeida KH, Joyce DF, Engelward BP, Horton JK, et al. Base excision repair intermediates induce p53-independent cytotoxic and genotoxic responses. *J Biol Chem.* 2003; 278: 39951–9. <https://doi.org/10.1074/jbc.M306592200> PMID: 12882965
9. Nemecek AA, Wallace SS, Sweasy JB. Variant base excision repair proteins: contributors to genomic instability. *Semin Cancer Biol.* 2010; 20: 320–8. <https://doi.org/10.1016/j.semcancer.2010.10.010> PMID: 20955798
10. Glassner BJ, Rasmussen LJ, Najarian MT, Posnick LM, Samson LD. Generation of a strong mutator phenotype in yeast by imbalanced base excision repair. *Proc Natl Acad Sci.* 1998; 95: 9997–10002. PMID: 9707589
11. Calvo JA, Meira LB, Lee C-YI, Moroski-Erkul CA, Abolhassani N, Taghizadeh K, et al. DNA repair is indispensable for survival after acute inflammation. *J Clin Invest.* 2012; 122: 2680–9. <https://doi.org/10.1172/JCI63338> PMID: 22684101
12. Hofseth LJ, Khan MA, Ambrose M, Nikolayeva O, Xu-Welliver M, Kartalou M, et al. The adaptive imbalance in base excision–repair enzymes generates microsatellite instability in chronic inflammation. *J Clin Invest.* 2003; 112: 1887–1894. <https://doi.org/10.1172/JCI119757> PMID: 14679184
13. Klapacz J, Lingaraju GM, Guo HH, Shah D, Moar-Shoshani A, Loeb LA, et al. Frameshift mutagenesis and microsatellite instability induced by human alkyladenine DNA glycosylase. *Mol Cell.* 2010; 37: 843–53. <https://doi.org/10.1016/j.molcel.2010.01.038> PMID: 20347426
14. Loeb LA. Microsatellite instability: marker of a mutator phenotype in cancer. *Cancer Res.* 1994; 54: 5059–63. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7923117> PMID: 7923117
15. Loeb LA. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer.* 2011; 11: 450–457. <https://doi.org/10.1038/nrc3063> PMID: 21593786
16. Grady WM. Genomic instability and colon cancer. *Cancer Metastasis Rev.* 2004; 23: 11–27. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15000146> PMID: 15000146
17. Boland CR, Lynch HT. The History of Lynch Syndrome. *Fam Cancer.* 2013;
18. Calvo JA, Moroski-Erkul CA, Lake A, Eichinger LW, Shah D, Jhun I, et al. Aag DNA glycosylase promotes alkylation-induced tissue damage mediated by Parp1. *PLoS Genet.* 2013; 9: e1003413. <https://doi.org/10.1371/journal.pgen.1003413> PMID: 23593019
19. Vickers MA, Vyas P, Harris PC, Simmonst DL, Higgs DR. Structure of the human 3-methyladenine DNA glycosylase gene and localization close to the 16p telomere. *Proc Nat Acad Sci USA.* 1993; 90: 3437–3441. PMID: 8475094
20. Cerda SR, Turk PW, Thor AD, Weitzman SA. Altered expression of the DNA repair protein, N-methyl-purine-DNA glycosylase (MPG) in breast cancer. *FEBS Lett.* 1998; 431: 12–8. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9684856> PMID: 9684856
21. Chan K, Sterling JF, Roberts SA, Bhagwat AS, Resnick MA, Gordenin DA. Base Damage within Single-Strand DNA Underlies In Vivo Hypermutability Induced by a Ubiquitous Environmental Agent. *PLoS Genet.* 2012; 8.
22. Auerbach P, Bennet RA, Bailey EA, Krokhan HE, Demple B. Mutagenic specificity of endogenously generated abasic sites in *Saccharomyces cerevisiae* chromosomal DNA. *PNAS.* 2005; 102: 17771–17776.
23. Gibbs PE, Lawrence CW. Novel mutagenic properties of abasic sites in *Saccharomyces cerevisiae*. *J Mol Biol.* 1995; 251: 229–236. <https://doi.org/10.1006/jmbi.1995.0430> PMID: 7643399
24. Branum ME, Reardon JT, Sancar A. DNA repair excision nuclease attacks undamaged DNA: A potential source of spontaneous mutations. *J Biol Chem.* 2001; 276: 25421–25426. <https://doi.org/10.1074/jbc.M101032200> PMID: 11353769
25. Connor EE, Wyatt MD. Active-site clashes prevent the human 3-methyladenine DNA glycosylase from improperly removing bases. *Chem Biol.* 2002; 9: 1033–41. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12323378> PMID: 12323378
26. Kavli B, Slupphaug G, Mol CD, Arvai AS, Peterson SB, Tainer JA, et al. Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J.* 1996; 15: 3442–7. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8670846> PMID: 8670846
27. Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, et al. Frameshift Mutations and the Genetic Code. *Cold Spring Harb Symp Quant Biol.* 1966; 31: 77–84. PMID: 5237214
28. Tran HT, Keen JD, Krickler M, Resnick MA, Gordenin DA. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol Cell Biol.* 1997; 17: 2859–65. PMID: 9111358
29. Lujan SA, Clausen AR, Clark AB, MacAlpine HK, MacAlpine DM, Malc EP, et al. Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.* 2014; 24: 1751–1764. <https://doi.org/10.1101/gr.178335.114> PMID: 25217194

30. Gragg H, Harfe BD, Jinks-Robertson S. Base Composition of Mononucleotide Runs Affects DNA Polymerase Slippage and Removal of Frameshift Intermediates by Mismatch Repair in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 2002; 22: 8756–8762. <https://doi.org/10.1128/MCB.22.24.8756-8762.2002> PMID: 12446792
31. Greene C, Jinks-Robertson S. Frameshift intermediates in homopolymer runs are removed efficiently by yeast mismatch repair proteins. *Mol Cell Biol*. 1997; 17. Available: <http://mcb.asm.org/content/17/5/2844.short>
32. Lang GI, Parsons L, Gammie AE. Mutation Rates, Spectra, and Genome-Wide Distribution of Spontaneous Mutations in Mismatch Repair Deficient Yeast. G3 (Bethesda). 2013;
33. Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PMJ, Kunkel TA. Division of Labor at the Eukaryotic Replication Fork. *Mol Cell*. 2008; 30: 137–144. <https://doi.org/10.1016/j.molcel.2008.02.022> PMID: 18439893
34. Fortune JM, Pavlov YI, Welch CM, Johansson E, Burgers PMJ, Kunkel TA. *Saccharomyces cerevisiae* DNA Polymerase δ : High fidelity for base substitutions but lower fidelity for single- and multi-base deletions. *J Biol Chem*. 2005; 280: 29980–29987. <https://doi.org/10.1074/jbc.M505236200> PMID: 15964835
35. Lujan SA, Williams JS, Pursell ZF, Abdulovic-Cui AA, Clark AB, Nick McElhinny SA, et al. Mismatch repair balances leading and lagging strand DNA replication fidelity. *PLoS Genet*. 2012; 8: e1003016. <https://doi.org/10.1371/journal.pgen.1003016> PMID: 23071460
36. Tran HT, Degtyareva NP, Gordenin DA, Resnick MA. Genetic factors affecting the impact of DNA polymerase δ proofreading activity on mutation avoidance in yeast. *Genetics*. 1999; 152: 47–59. PMID: 10224242
37. Lyons DM, O'Brien PJ. Human base excision repair creates a bias toward -1 frameshift mutations. *J Biol Chem*. 2010; 285: 25203–12. <https://doi.org/10.1074/jbc.M110.118596> PMID: 20547483
38. Nishant KT, Singh ND, Alani E. Genomic mutation rates: What high-throughput methods can tell us. *BioEssays*. 2009; 31: 912–920. <https://doi.org/10.1002/bies.200900017> PMID: 19644920
39. Lau AY, Wyatt MD, Glassner BJ, Samson LD, Ellenberger T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc Natl Acad Sci*. 2000; 97: 13573–13578. <https://doi.org/10.1073/pnas.97.25.13573> PMID: 11106395
40. Abner CW, Lau AY, Ellenberger TE, Bloom LB. Base Excision and DNA Binding Activities of Human Alkyladenine DNA Glycosylase Are Sensitive to the Base Paired with a Lesion. *J Biol Chem*. 2001; 276: 13379–13387. <https://doi.org/10.1074/jbc.M010641200> PMID: 11278716
41. O'Brien PJ, Ellenberger TE. Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines. *Biochemistry*. 2003; 42: 12418–29. <https://doi.org/10.1021/bi035177v> PMID: 14567703
42. Zhang Y, O'Brien PJ. Repair of Alkylation Damage in Eukaryotic Chromatin Depends on Searching Ability of Alkyladenine DNA Glycosylase. *ACS Chem Biol*. 2015; 10: 2606–2615. <https://doi.org/10.1021/acscchembio.5b00409> PMID: 26317160
43. Hendershot JM, O'Brien PJ. Critical role of DNA intercalation in enzyme-catalyzed nucleotide flipping. *Nucleic Acids Res*. 2014; 42: 12681–12690. <https://doi.org/10.1093/nar/gku919> PMID: 25324304
44. Connor EE, Wilson JJ, Wyatt MD. Effects of substrate specificity on initiating the base excision repair of N-methylpurines by variant human 3-methyladenine DNA glycosylases. *Chem Res Toxicol*. 2005; 18: 87–94. <https://doi.org/10.1021/tx049822q> PMID: 15651853
45. Guo HH, Choe J, Loeb LA. Protein tolerance to random amino acid change. *Proc Natl Acad Sci U S A*. 2004; 101: 9205–10. <https://doi.org/10.1073/pnas.0403255101> PMID: 15197260
46. Lang GI, Murray AW. Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics*. 2008; 178: 67–82. <https://doi.org/10.1534/genetics.107.071506> PMID: 18202359
47. Zhu YO, Siegal ML, Hall DW, Petrov DA. Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci U S A*. 2014; 111: E2310–8. <https://doi.org/10.1073/pnas.1323011111> PMID: 24847077
48. Serero A, Jubin C, Loeillet S, Legoix-Ne P, Nicolas AG. Mutational landscape of yeast mutator strains. *Proc Natl Acad Sci*. 2014; 111.
49. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004; 14: 1188–90. <https://doi.org/10.1101/gr.849004> PMID: 15173120
50. SGD Project. *saccharomyces_cerevisiae.gff* [Internet]. [cited 15 Jul 2014]. Available: http://downloads.yeastgenome.org/curation/chromosomal_feature/saccharomyces_cerevisiae.gff
51. McGuffee SR, Smith DJ, Whitehouse I. Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination. *Mol Cell*. 2013; 50: 123–135. <https://doi.org/10.1016/j.molcel.2013.03.004> PMID: 23562327

52. Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*. 2013; 497: 127–131. <https://doi.org/10.1038/nature12121> PMID: 23615609
53. Knott SRV, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, et al. Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in *S. cerevisiae*. *Cell*. 2012; 148: 99–111. <https://doi.org/10.1016/j.cell.2011.12.012> PMID: 22265405
54. Lyons DM O'Brien PJ. Efficient recognition of an unpaired lesion by a DNA repair glycosylase. *J Am Chem Soc*. 2009; 131: 17742–3. <https://doi.org/10.1021/ja908378y> PMID: 19924854
55. Lujan SA, Clark AB, Kunkel TA. Differences in genome-wide repeat sequence instability conferred by proofreading and mismatch repair defects. *Nucleic Acids Res*. 2015; 43: 4067–4074. <https://doi.org/10.1093/nar/gkv271> PMID: 25824945
56. Zhai J, Hingorani MM. *Saccharomyces cerevisiae* Msh2-Msh6 DNA binding kinetics reveal a mechanism of targeting sites for DNA mismatch repair. *Proc Natl Acad Sci U S A*. 2010; 107: 680–685. <https://doi.org/10.1073/pnas.0908302107> PMID: 20080735
57. Berdal KG, Johansen RF, Seeberg E. Release of normal bases from intact DNA by a native DNA repair enzyme. *EMBO J*. 1998; 17: 363–367. <https://doi.org/10.1093/emboj/17.2.363> PMID: 9430628
58. Chan K, Resnick M, Gordenin D. The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. *DNA Repair (Amst)*. 2013; 12: 878–889.
59. Auerbach P, Demple B. Roles of Rev1, Pol ζ , Pol32 and Pol η in the bypass of chromosomal abasic sites in *Saccharomyces cerevisiae*. *Mutagenesis*. 2010; 25: 63–69. <https://doi.org/10.1093/mutage/geb045> PMID: 19901007
60. Pagès V, Johnson RE, Prakash L, Prakash S. Mutational specificity and genetic control of replicative bypass of an abasic site in yeast. *Proc Natl Acad Sci U S A*. 2008; 105: 1170–1175. <https://doi.org/10.1073/pnas.0711227105> PMID: 18202176
61. Maiti A, Noon MS, MacKerell AD, Pozharski E, Drohat AC. Lesion processing by a repair enzyme is severely curtailed by residues needed to prevent aberrant activity on undamaged DNA. *Proc Natl Acad Sci U S A*. 2012; 109: 8091–6. <https://doi.org/10.1073/pnas.1201010109> PMID: 22573813
62. Fox EJ, Prindle MJ, Loeb LA. Do mutator mutations fuel tumorigenesis? *Cancer Metastasis Rev*. 2013; 32: 353–61. <https://doi.org/10.1007/s10555-013-9426-8> PMID: 23592419
63. Madhukar NS, Warmoes MO, Locasale JW. Organization of enzyme concentration across the metabolic network in cancer cells. *PLoS One*. 2015; 10: 1–14.
64. Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, et al. The quantitative proteome of a human cell line. *Mol Syst Biol*. 2011; 7: 549. <https://doi.org/10.1038/msb.2011.82> PMID: 22068332
65. Galick HA, Kathe S, Liu M, Robey-Bond S, Kidane D, Wallace SS, et al. Germ-line variant of human NTH1 DNA glycosylase induces genomic instability and cellular transformation. *Proc Natl Acad Sci U S A*. 2013; 110: 14314–9. <https://doi.org/10.1073/pnas.1306752110> PMID: 23940330
66. Sjolund A, Nemeč AA, Paquet N, Prakash A, Sung P, Doublé S, et al. A germline polymorphism of thymine DNA glycosylase induces genomic instability and cellular transformation. *PLoS Genet*. 2014; 10: e1004753. <https://doi.org/10.1371/journal.pgen.1004753> PMID: 25375110
67. Yamtich J, Nemeč AA, Keh A, Sweasy JB. A germline polymorphism of DNA polymerase beta induces genomic instability and cellular transformation. *PLoS Genet*. 2012; 8: e1003052. <https://doi.org/10.1371/journal.pgen.1003052> PMID: 23144635
68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9: 357–9. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
69. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20: 1297–1303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
70. Li H. A statistical framework for SNP calling mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011; 1–7.
71. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
72. Hall BM, Ma C-X, Liang P, Singh KK. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics*. 2009; 25: 1564–5. <https://doi.org/10.1093/bioinformatics/btp253> PMID: 19369502
73. Senear DF, Brenowitz M. Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. *J Biol Chem*. 1991; 266: 13661–13671. PMID: 1856200