Hypomorphic mutation in the large subunit of replication protein A affects mutagenesis by human APOBEC cytidine deaminases in yeast.

AUTHORS

Matthew S. Dennen[1*], Zachary W. Kockler[1*], Steven A. Roberts[2], Adam B. Burkholder[3], Leszek J. Klimczak[4], Dmitry A. Gordenin[1**]

AFFILIATIONS

[1] Genome Integrity & Structural Biology Laboratory, National Institute of Environmental Health Sciences, Durham, NC 27709

[2] Department of Microbiology and Molecular Genetics, University of Vermont Cancer Center, University of Vermont, Burlington, VT 05405

[3]Office of Environmental Science Cyberinfrastructure, National Institute of Environmental Health Sciences, US National Institutes of Health, Durham, NC, 27709, USA

[4] Integrative Bioinformatics Support Group, National Institute of Environmental Health Sciences, Durham, NC, 27709

* These authors equally contributed into the study.
**To whom correspondence should be addressed. Email: gordenin@niehs.nih.gov

RUNNIING TITLE

RPA affects APOBEC mutation spectrum

KEY WORDS

APOBEC mutagenesis, single-stranded DNA, replication protein A (RPA), translesion DNA synthesis, uracil-DNA glycosylase

ABSTRACT
Human APOBEC single-strand (ss) specific DNA and RNA cytidine deaminases change cytosines to uracils and function in antiviral innate immunity, RNA editing, and can cause hypermutation in chromosomes. The resulting uracils can be directly replicated, resulting in C to T mutations, or uracil-DNA glycosylase can convert the uracils to abasic (AP) sites which are then fixed as C to T or C to G mutations by translesion DNA polymerases. We noticed that in yeast and in human cancers, contributions of C to T and C to G mutations depends on the origin of ssDNA mutagenized by APOBECs. Since ssDNA in eukaryotic genomes readily binds to replication protein A (RPA) we asked if RPA could affect APOBEC-induced mutation spectrum in yeast. For that purpose, we expressed human APOBECs in the wild-type yeast and in strains carrying a hypomorph mutation *rfa1-t33* in the large RPA subunit. We confirmed that the *rfa1-t33* allele can facilitate mutagenesis by APOBECs. We also found that the *rfa1-t33* mutation changed the ratio of APOBEC3A-induced T to C and T to G mutations in replicating yeast to resemble a ratio observed in long-persistent ssDNA in yeast and in cancers. We present the data suggesting that RPA may shield APOBEC formed uracils in ssDNA from Ung1, thereby facilitating C to T mutagenesis through the accurate copying of uracils by replicative DNA polymerases. Unexpectedly, we also found that for uracils shielded from Ung1 by wild-type RPA the mutagenic outcome is reduced in the presence of translesion DNA polymerase zeta.

INTRODUCTION

Humans have several APOBEC (Apolipoprotein B mRNA-Editing Enzyme, Catalytic polypeptide-like) cytosine deaminases that can convert cytosines to uracils in single-stranded (ss) RNA or DNA. APOBECs operate in a physiologically important manner by editing selected mRNAs and contributing into innate immunity defense against RNA and DNA viruses (BANERJEE *et al.* 2008; HARRIS AND DUDLEY 2015; LERNER *et al.* 2018). Further, APOBECs can also introduce mutations into chromosomes of human malignant tumors. In fact, APOBECs are one of the most ubiquitous and prevailing causes of mutagenesis in cancers (ROBERTS *et al.* 2013; MERTZ *et al.* 2022). Due to structural constraints, only ssDNA or RNA can be substrates for APOBEC cytidine deamination (SALTER *et al.* 2016; KOUNO *et al.* 2017). Short-lived stretches of ssDNA that are formed during replication, repair and in transient R-loops during transcription can be mutated by APOBECs. Alternatively, long persistent stretches of ssDNA formed by DNA end-resection in DSBs, at uncapped telomeres, or by break-induced replication can be deaminated simultaneously resulting in several cytosines being deaminated stretching over many kilobases, termed mutation clusters (SAINI AND GORDENIN 2020). Such clusters were observed in yeast and human cell models as well as in cancer genomes. (CHAN AND GORDENIN 2015; SAKOFSKY *et al.* 2019; SAINI AND GORDENIN 2020).

Uracils formed through APOBEC cytidine deamination in ssDNA can be accurately copied by replicative DNA polymerases resulting in C to T mutations (Figure 1; steps b2→b2.1→b2.1.1). Another path of uracils-associated mutagenesis can be triggered by uracil-DNA glycosylase (Udg), which is encoded by UNG1 gene in yeast or by UNG in humans. Udg glycosylation (U-glycosylation, step b1 on Figure 1) leaves abasic (AP) site in place of an uracil. Since AP-sites are chemically unstable they may result in ssDNA breakage, rearrangements, chromosome loss or even cell death (step (c) on Figure 1). Detrimental effects of AP-sites in ssDNA can be alleviated by several ways (BOITEUX AND JINKS-ROBERTSON 2013; KROKAN *et al.* 2014; SAINI AND GORDENIN 2020). If an AP-site is formed in ssDNA which can be re-annealed with undamaged complementary strand by replication fork regression or by re-winding of transiently unwound DNA within a R-loop, it will be fixed by templated error-free base excision repair (BER) without leaving any mutation trace (steps b1.2→b1.2.1 on Figure 1). Alternatively, DNA strands containing an AP-site can be copied directly by error-prone translesion synthesis (TLS) which will result in APOBEC-induced mutations (steps b1.1→b1.1.1 on Figure 1). All extensions from a base inserted across an AP-site require Pol zeta polymerase (catalytic subunit encoded by *REV3* gene in yeast). Pol zeta extension complex also includes non-catalytic subunits of replicative polymerase delta, Rev7 subunit and Rev1 (MARTIN AND WOOD 2019). The Rev1 serves as scaffold for the Pol zeta extension step as well as to perform TLS insertion of a cytosine across AP-site. Polymerase(s) inserting adenines or thymines are not yet defined ((CHAN *et al.* 2012; BOITEUX AND JINKS-ROBERTSON 2013; CHAN *et al.* 2013; HOOPES *et al.* 2017) and references therein). Thus, U-glycosylation followed by TLS across AP-site and subsequent DNA replication (step b1.1.1 on Figure 1) can result in any of the three possible substitutions of a cytosine, C to T, C to G, and C to A as well as in a non-mutagenic outcome. Summarizing pathways illustrated on Figure 1, the spectrum of three possible APOBEC-induced substitutions

of cytosines would be defined by the efficiency of U-glycosylation and by relative contribution of mutagenic TLS pathways.

In human cancers as well as in the yeast model systems, APOBEC mutagenesis produced mostly C to T and C to G mutations with C to A mutations occurring at low frequency barely distinguishable from non-APOBEC mutation background (CHAN *et al.* 2013; CHAN AND GORDENIN 2015; MERTZ *et al.* 2022). We noticed that in UNG1 wild-type (WT) yeast the ratio of C to T and C to G mutations depended on the way ssDNA was formed (Figure 2). In yeast undergoing normal replication, APOBEC-induced C to T mutations strongly prevailed over C to G events and both rarely formed mutation clusters, i.e., were scattered over the genome. On the contrary, the spectra of APOBEC-induced clustered mutations in long persistent ssDNA formed by end-resection at uncapped telomeres or via break-induced replication (BIR) contained comparable numbers of C to G and C to T events. Mutation clusters in human cancers enriched with mutations in APOBEC mutation motif also contained nearly equal numbers of C to T and C to G events. Interestingly, spectra of scattered APOBEC motif mutations in APOBEC-hypermutated tumors were slightly, but statistically significantly, shifted towards C to T changes (Figure 2 and Figure S1). In general, the contribution of C to G changes in yeast models and in human cancers appear to increase, when long persistent ssDNA is expected to form.

The universal feature of ssDNA formed in eukaryotes is that it readily binds to Replication Protein A (RPA) - multi-subunit protein complex required for DNA replication and for DSB repair (CHEN AND WOLD 2014). Studies have shown that RPA binding can impede APOBEC cytidine deamination *in vitro* (LADA *et al.* 2011; BROWN *et al.* 2021; WONG *et al.* 2021) and reduce APOBEC mutagenesis in yeast (HOOPES *et al.* 2016). Long persistent ssDNA resulting in the formation of APOBEC mutation clusters may have a lower fraction bound by RPA at any given moment, because of RPA depletion or because of continuously ongoing exchange between ssDNA-bound RPA and RPA in solution (TOLEDO *et al.* 2013; CHEN AND WOLD 2014; TOLEDO *et al.* 2017). In order to explain the variations in the APOBEC mutation spectra in different genomic contexts (Figure 2), we propose that RPA may not only shield cytosines in ssDNA from APOBEC but can also shield APOBEC-induced uracils from uracil DNA glycosylase, thereby reducing a chance of creating AP sites (pathway (b1) on Figure 1) and consequently reducing a chance of C to G (and C to A) mutations. Another way for RPA to alter mutation spectrum of APOBEC-induced mutations could be via modifying choices of bases inserted by TLS across AP-site. Several lines of evidence suggested that RPA can play a role not only in DNA replication and repair but also in TLS by regulating PCNA sliding over ssDNA or/and by recruiting Rad6/Rad18 PCNA monoubiquitination essential for TLS (HEDGLIN AND BENKOVIC 2017b; HEDGLIN AND BENKOVIC 2017a; HEDGLIN *et al.* 2019). It may turn out that RPA modulation of TLS affects the TLS choice of inserting one of three possible bases across AP-site (step (b1.1) on Figure 1), thereby modulating spectrum of base substitutions in concert with ssDNA accessibility to RPA. Altogether, the RPA ssDNA binding and/or its' interference with TLS may affect choice between C to T and C to G APOBEC mutagenesis. Therefore, we have explored the effects of a RPA hypomorph allele in the presence and in the absence of TLS capacity on the spectrum of APOBEC-induced mutations in a yeast *Saccharomyces cerevisiae* model. We present here results favoring the hypothesis that during DNA replication RPA shields

uracils in ssDNA formed by APOBEC cytidine deamination from uracil DNA glycosylase Ung1, thereby allowing the direct copying of uracils by replicative DNA polymerases. Unexpectedly, we also found that for uracils shielded from Ung1 by wild-type RPA the mutagenic outcome is reduced by wild-type Pol zeta.

MATERIALS AND METHODS

Mutation reporter for accessing mutagenesis by APOBEC cytidine deaminases.
The reporter design and rationale were described in (Figure 3) and in (ROBERTS *et al.* 2012; HOOPES *et al.* 2016). An additional single-base substitution reporter *ura3-29* (SHCHERBAKOVA AND PAVLOV 1996; ELANGO *et al.* 2019) contains an A:T to G:C mutation. This construct allows selection of the range of forward mutations inactivating *CAN1* gene by canavanine-resistance (Can-R). *CAN1* reporter was used to compare mutagenic effects of an APOBEC enzyme in strains of different genotypes. The reporter also enables selection of mutations in the *ura3-29* mutant cytosine base to one of the three possible bases, A, T, or G, because each of these substitutions and no other change in a yest genome can result in reversions of a strain to Ura^+. Importantly, mutant cytosine is located within the trinucleotide context tCt (mutated base capitalized) preferred by most of APOBEC3 cytosine deaminases. The reporter cassette was placed on one of the two sides around the strong replication origin ARS216 thereby allowing the *ura3-29* cytosine to be mostly present in either lagging or in the leading strand template (left and right positions, respectively).

Yeast strains.
The *Saccharomyces cerevisiae* yeast strains of CG379 genetic background used in this study were derivatives of the ySR128 (*MAT alpha ura3Δ can1Δ ade2Δ leu2-3,112 trp1-289 lys2::ADE2-URA3-CAN1*) in which triple-gene *ADE2-URA3-CAN1* was inserted into LYS2 in its normal chromosome II location to the left of ARS216 or of ySR366 (*MAT a ura3Δ can1Δ ade2Δ leu2-3,112 trp1-289 lys2Δ Chr.II 488694::lys2::ADE2-URA3-CAN1*) in which the entire ADE2-URA3-CAN1 cassette was moved on the left side of strong replication origin ARS216 (Figure 3 and (ROBERTS *et al.* 2012; HOOPES *et al.* 2016)). Standard methods were used to handle, grow, cross and dissect tetrads of yeast strains were used (SHERMAN *et al.* 1981). Introducing and PCR verification of the mutant *rfa1-t33* allele was described in (HOOPES *et al.* 2016). Single-base substitution *ura3-29* was introduced by integration and pop-out as described in (SHCHERBAKOVA AND PAVLOV 1996). Deletions of *UNG1* and *REV3* were generated by replacing their ORFs with natMX3 cassette, conferring resistance to nourseothricin or with the kanMX4 cassette, conferring resistance to geneticin, respectively (GOLDSTEIN AND MCCUSKER 1999). Combination of different genetic defects was done by mating type switching and crossing isogenic strains followed by tetrad dissection (HOOPES *et al.* 2016). Haploid progeny with a desired combination of generic defects was identified by phenotype and by PCR -verification. Yeast strains used in this work and their genotypes are listed in Supplementary Table 4.

APOBEC plasmids.

Vectors for the codon optimized expression of A3A, A3B, and A3G in yeast (i.e. pSR435, pSR440, and pCM-252-A3G, respectively) and the corresponding empty vector control (pySR419) were previously described (CHAN *et al.* 2012; HOOPES *et al.* 2016). A1 (pSR433), A3C (pSR469), A3DE (pSR470), A3F (pSR471), and A3H (pSR472) yeast expression vectors were created by digesting gene-blocks containing codon optimized cDNAs (ordered from DNA2.0) with StuI and ClaI and ligating the resulting restriction fragment into the corresponding restriction sites of pySR419. Accurate cloning of the respective cDNAs was confirmed by Sanger sequencing of the inserted cDNA and DNA flanking the insertion site. Plasmid sequences in GenBank format are provided in Supplementary Dataset 3.

Determining APOBEC-induced mutation rates.
Yeast strains carrying an APOBEC plasmid or empty vector (EV) control plasmid were taken from genotype and phenotype-verified stocks (20% glycerol-80°C), patched on YPDA supplemented with 300 mg/L hygromycin (HYG) and grown for 2-3 days at 23 ^{o}C. These yeast strains were then streaked for single colonies on YPDA supplemented with 300 mg/L HYG with an addition of 20 mg/L doxycycline hyclate (DOX) to facilitate expression of an APOBEC ORF for 3 days at 30°C and to allow for large single colonies to form. The entire colony was then picked using a toothpick or pipette tip and suspended in 300uL of dd$H_2$0 in a centrifuge tubes (dilution tube 0) followed by ten-fold serial dilutions. 100uL of 10,000x and 100,000x dilutions were plated onto COM media and grown for three days at 23°C to determine numbers of viable cells in a suspension. Non-diluted and 10x diluted suspensions were plated onto COM without arginine supplemented with 60 mg/L of L-canavanine or onto complete media lacking uracil and grown for 7 days at 23 ^{o}C. After the period of growth, the plates were stored at 4°C and the colonies were counted using the automated colony counter Protocol3. Mutation rates were calculated as described in (SAINI *et al.* 2017).

Collecting independent APOBEC-induced and spontaneous Ura+ revertants for determining spectra of base substitutions in *ura3-29* site and across the yeast genome.
Independent Ura+ revertants were collected from small independent yeast culture grown on solid media as described in ((JIN *et al.* 2003) and references therein). Briefly small, about 1 uL, drops containing from 10^3 to 10^4 yeast cells were picked from yeast 10^5 to 10^6 cells/mL suspension with 121-pin multiprong device to solid YPDA media supplemented with HYG (300 mg/L) and 20 mg/L DOX for plasmid selection and for APOBEC induction, respectively, or directly to COM media without uracil to confirm that there was no or very little number of preexisting Ura+ mutants in the initial suspension. YPDA +HYG +DOX plates were incubated for 2-3 days at 30°C and then replica plated to COM media without uracil. By the time of replica plating each small prong imprint contained 2-5x10^6 cells, so vast majority of Ura+ revertants on each prong occurred during growth of initially plated cells and therefore were independent from revertants on other prongs. One Ura+ revertant was picked from a single prong and streaked for single colonies. URA3 ORF was PCR-amplified, and Sanger sequenced or used for whole-genome sequencing. Primers used to amplify and sequence URA3 ORF from revertants are listed in Supplementary Table 4.

Whole genome sequencing, mutation calling and mutation spectra analyses.

Genomic DNA preparation, library preparation and Illumina sequencing and mutation calling was performed as described in (HUDSON *et al.* 2023) with the exception that Illumina reads were mapped against yeast reference ySR128 (ROBERTS *et al.* 2012). Densities of APOBEC-induced mutations were presented as an average of mutation counts per genome for a group of APOBEC-expressing strain with the same genotype minus average density in the group of strains with the same genotype carrying the empty vector control plasmid. If the difference was negative, density of induced mutations was considered as zero. The average densities of induced mutations were used only for illustration purposes only. Set of whole-genome mutation counts in individual samples of the same genotypes were compared with the set of mutation counts in the group of another genotype using two-sided Mann-Whitney test. Mutation spectra were compared by two-sided chi-square of by Fisher's exact test.

RESULTS

Increased APOBEC mutagenesis in yeast strains with hypomorph RPA allele *rfa1-t33*. As presented in the Introduction, we proposed that the increased fraction of APOBEC induced C→G changes in persistent long ssDNA as compared with ssDNA formed during normal replication could be explained by different access of Replication Protein A (RPA) to ssDNA formed in different genomic contexts. We employed *rfa1-t33* (S373P), a hypomorphic allele of the yeast RPA large subunit *RFA1*, to assess this hypothesis. The *rfa1-t33* was previously shown to reduce RPA DNA binding to ssDNA and to increase mutagenesis by APOBEC3A (A3A) and APOBEC3B (A3B) (DENG *et al.* 2014; DENG *et al.* 2015; HOOPES *et al.* 2016; RUFF *et al.* 2016). In extension of that finding we assessed effects of *rfa1-t33* on mutagenesis in *CAN1* gene by all human APOBECs for which deaminase activity has been reported (REFSLAND AND HARRIS 2013; SALTER *et al.* 2016; MERTZ *et al.* 2022) (Figure 4). Mutagenesis was assessed in *ung1Δ* strains lacking uracil DNA glycosylase, therefore allowing all (or most of all) uracils generated by an APOBEC enzyme to be fixed into mutations (see Figure 1). The hypomorph *rfa1-t33* caused increased *CAN1* mutation rates even in the absence of APOBEC expression (EV, empty vector control) which could be explained by higher exposure of ssDNA to uncontrolled base-damaging factors and/or by increase in gross chromosome rearrangements encompassing *CAN1* (BANERJEE *et al.* 2008). All APOBECs except APOBEC1 (A1) caused statistically significant increases of *CAN1* mutation rates over the empty vector control in *RFA1-WT* background, however, the increases in APOBEC3DE (A3DE), APOBEC3F (A3F), and APOBEC3H (A3H) expressing strains were only moderate, within 2-fold (Figure S2A). The same three APOBECs, A3DE, A3F, and A3H, did not result in statistically significant increases over the empty vector control in the strains carrying the hypomorph *rfa1-t33* (Figure S2B). On the contrary the *rfa1-t33* strains carrying A1, A3A, A3B, A3C, and A3G, showed statistically significant increases over the empty vector (Figure S2B), indicating that the increase in mutation rates was due to APOBEC-induced mutagenesis. As anticipated based on prior studies (see Introduction), the comparison of *CAN1* mutation rates between *RFA1-WT* and *rfa1-t33* strains also revealed the increase in APOBEC-induced mutagenesis caused by a the hypomorph allele in yeast carrying A1, A3B, and A3C (Figure 4). The strongest increase in rfa1-t33 over RFA1-WT background was observed for A1. This agreed with biochemical experiments that showed the strongest inhibition of A1 deaminase activity by human RPA *in vitro* (WONG *et al.* 2021). The

weakest *rfa1-t33*-associated increase was for A3B which has the highest APOBEC-induced mutagenesis in both, *RFA1-WT* and in *rfa1-t33* backgrounds. Moreover, two other highly active enzymes, A3A and A3G did not show difference between *RFA1-WT* and in *rfa1-t33.* This minimal or non-existent difference in mutagenesis with the highly active APOBECs can be explained by the excess of enzymatic activity over the available ssDNA substrate. Altogether, our results indicate that *rfa1-33* hypomorph can facilitate access of an APOBEC enzyme to ssDNA in our strain background and therefore can be used to test our hypothesis about differences in RPA-ssDNA interactions accounting for differences in the spectra of APOBEC-induced mutation spectra in different genomic contexts (Introduction and Figure 2). We chose for this purpose the yeast strains expressing A3A, because this enzyme showed robust mutagenesis in both *RFA1-WT* as well as in *rfa1-t33* strains (Figure 4). Also, A3A is the most potent source of APOBEC-induced scattered and clustered mutations in APOBEC-hypermutated human tumors (CHAN *et al.* 2015; PETLJAK AND MACIEJOWSKI 2020; PETLJAK *et al.* 2022).

Uracil DNA glycosylase Ung1 deletion and *rfa1-t33* hypomorph can affect spectrum of base substitutions induced by APOBEC3A in *ura3-29* single-nucleotide reporter.
As outlined in the Introduction, we hypothesized that the excess of C to G changes in APOBEC-induced mutation clusters as compared with mutations induced in the course of normal replication could be due to RPA shielding ssDNA in replicating cells from Ung or due to differences in choices of TLS across AP-sites in different genomic contexts. In order to explore these hypotheses, we compared mutation spectra induced APOBEC3A in glycosylase-proficient (*UNG1-WT*) replicating yeast between strains carrying *RFA1-WT* and *rfa1-t33* hypomorph allele in the presence and in the absence of TLS function (*REV3-WT* and *rev3Δ*, respectively). We also studied mutagenesis in *ung1Δ* strains, where no APOBEC-induced AP-sites were expected.

Mutagenesis was assessed with *ura3-2*9 single-base reporter allowing the determination of the mutation spectrum by sequencing *ura3-29* mutation site in the revertants. As expected, the Ura^+ mutation rates in *ung1Δ* strains, where all mutations originated from mere copying of uracils, were generally higher than in *UNG1-WT*, where mutations can originate from uracils as well as from AP-sites (compare Figures 5A and 5B). In agreement with (HOOPES *et al.* 2016), there was a small but statistically significant increase in reversion rates of *ura3-29* positioned to the left of strong replication origin with the mutant cytosine would be mostly present in the lagging strand template. This bias was observed in wild-type strains with or without Ung1 as well as in *ung1Δ RFA1-WT rev3Δ* strains. Because of orientation dependence of ura3-29 reversion rates, both orientations were included into analysis of the mutation spectrum. Since reversion rates in in *UNG1-WT rfa1-t33 rev3Δ* strains were very low and did not produce enough revertants, *ura3-29* mutation spectrum was assessed only in *REV3-WT* strains (Figure 6). Mutation spectra were very similar between two orientations (Left and Right sides of the replication origin) therefore they will be considered as biological repeats in spectra analyses. As expected, all changes in *ung1Δ* strains were C to T regardless of *RFA1* genotype (Figure 6A). On the contrary, the spectra in UNG1-WT strains contained C to T as well as C to G and C to A mutations (Figure 6B). Mutation spectra in *UNG1-WT RFA1-WT* were mostly C to T with a very small fraction of C

to G. This was similar to previously published whole-genome and reporter-based spectra of APOBEC-induced mutations in replicating yeast (Figure 2). Strikingly, the spectrum in replicating yeast strains of *UNG1-WT rfa1-t33* genotype contained nearly equal amounts of C to G and C to T mutations also with small presence of C to A. Thus, hypomorph *rfa1-t33* allele shifted the spectrum of APOBEC-induced mutations in replicating yeast strains to resemble the spectra observed in mutation clusters formed in long persistent ssDNA in yeast as well as in human cancers (Figure 2).

Hypomorph *rfa1-t33* allele increases probability of APOBEC-induced uracils being converted into AP-sites by Ung1 glycosylase.

As outlined in the Introduction, the increase in C to G (and C to A) mutations of *ura3-29* in *rfa1-t33* strains could be explained by either increased probability of AP-site generation or by altering TLS preference to more frequently inserting cytosines across AP-sites. This question can be resolved by assessing mutation spectra in yeast *rev3Δ* lacking TLS capacity. However, *ura3-29* reversion rate in *UNG1-WT rfa1-t33 rev3Δ* strain was very low (Figure 5B). In order to assess the spectrum and frequency of APOBEC-induced mutations in all strains, including infrequently mutating *UNG1-WT rfa1-t33 rev3Δ*, we have sequenced genomes of 30-48 isolates of *UNG1-WT* strains belonging to each of the four genotypes in which we initially measured rates in *ura3-29* mutation reporter. Whole-genome sequencing allowed to build representative mutation catalogs for all strains, regardless of mutation rate. It also alleviated the possible impact of genomic context and of sequence composition in reporters. After subtracting empty vector background from mutation spectrum in all four genotypes we found that over 95% of remaining substitutions were in C:G pairs (Figure S3), which indicated that these are indeed A3A-induced mutations. Consistent with observations for *ura3-29* single-base substitution reporter in *UNG1-WT* strains (Figure 6A), whole-genome mutation spectra of *UNG1-WT rfa1-t33* contained more C to G mutations than the spectrum of *UNG1-WT RFA1-WT* strain (Figure 7A, B). (We note that all genotypes contained a tiny fraction of C to A mutations).

While C to G (and C to A) mutations in UNG1-WT yeast can occur only from TLS of AP-sites generated by Ung excision of A3A-created uracils, C to T mutations can stem from either TLS of AP-sites or from error-free replication of uracils (Figure 1). When all branches of TLS across AP site were eliminated by *rev3Δ* in *rfa1-t33* background, densities of both C to G (and C to A) as well as C to T mutations were reduced more than 10-fold (Figure 7A). Such a severe reduction in densities of all types of APOBEC-induced substitutions suggested that bulk of A3A mutagenesis in *UNG1-WT rfa1-t33* was caused by AP-sites. There was smaller, albeit detectable presence of C to G (and C to A) mutations in *UNG1-WT RFA1-WT* strain, however the major fraction of APOBEC-induced mutations was due to C to T changes (Figure 7A, B). Elimination of TLS by *rev3Δ* in *RFA1-WT* strain left only C to T mutations in the spectrum but did not reduce mutation density (Figure 7A), therefore indicating that the majority of APOBEC-induced mutations in *RFA1-WT* replicating yeast resulted from copying of uracils that escaped Ung glycosylase processing into AP-site. Since most mutations in *rfa1-t33* had come from AP-

sites, we conclude that the wild-type RPA can protect ssDNA of replicating yeast cells not only from APOBEC deaminases (Figure 4) but also from uracil DNA-glycosylase.

Unexpectedly, the density and fraction of C to T mutations increased in *UNG1-WT rev3Δ* strain as compared with *UNG1-WT REV3-WT* (Figure 7A). Since in *rev3Δ* all C to T mutations should originate from replication of APOBEC-induced uracils, the observed increase indicates that more uracils are formed and/or retained until replication. More studies are needed to understand the causes of the observed stimulation of C to T mutagenesis in the absence of Rev3 protein. Speculations about possible mechanisms are presented in Discussion.

DISCUSSION

Our study revealed that in replicating yeast the contribution of two major base substitution types, C to T and C to G, into spectrum of APOBEC-induced mutations relies on capacity of RPA. Measurement of reporter-based and whole-genome mutagenesis in yeast carrying combinations of defects, hypomorph mutation in RPA large subunit *rfa1-t33*, uracil DNA glycosylase *UNG1*, and a deletion of a gene for the catalytic subunit of TLS polymerase Pol zeta (*rev3Δ*) indicated that the impeded functionality of RPA increases a chance to generate AP-sites from uracils formed by APOBEC cytosine deaminase in transient ssDNA (path (b1) on Figure 1). Replication across AP-sites by TLS can generate C to G as well as C to T mutations. Our results suggest a testable hypothesis about RPA counteracting Ung1 activity on ssDNA. Such a counteraction to Ung1 agrees with previously established inhibition of APOBEC cytosine deamination by RPA (Figure 4, and Introduction). Normal RPA could shield uracils in ssDNA from Ung1 the same manner that it shields cytosines from APOBECs. Thus, with normal RPA, most of APOBEC-induced uracils would be directed to pathway (b2) depicted in Figure 1, where only C to T mutations can occur.

Prevalence of the C to T mutations over C to G in wild type yeast replication (published data summarized in Figure 2 and the data of this study (Figures 6, 7) suggest that the latter base substitutions comprised only a minor fraction because RPA is protecting APOBEC-induced uracils from Ung in short-lived ssDNA within replication forks. Unlike replication forks, long, multi-kilobase stretches of ssDNA formed by end-resection of uncapped telomeres or in the course of BIR may exist for a much longer time and thus may give a greater chance for Ung acting on APOBEC-induced uracils. In addition, there could be some depletion of RPA in cells forming long persistent ssDNA (TOLEDO *et al.* 2013; CHEN AND WOLD 2014; TOLEDO *et al.* 2017), which would further enable Ung access to uracils. Altogether, enhanced Ung function in long persistent ssDNA formed by end resection or in BIR would explain, why in these locations APOBEC induces approximately equal amounts of C to T and C to G mutations (Figure 2).

Our work also revealed unexpected result of increased C to T mutagenesis in *RFA1-WT* replicating yeast carrying deletion *rev3Δ* of the yeast Pol zeta essential for all branches of yeast TLS. In the absence of Pol zeta, even if AP-sites are formed by Ung (Figure 1 pathway (b1)) they could result in a base substitution. Some AP-sites could result in a loss of the DNA

molecule (pathway (c)) and thus, no mutations are produced. Alternatively, AP-sites could be channeled into either error-free BER via lesion bypass or into reannealing of R-loop (Figure 1, pathway (b1.2), which would not produce mutations as well. Thus, changes within pathways generating AP-sites cannot explain increase in numbers of C to T mutations in the absence of Pol zeta. The only way to generate C to T mutations in the absence of Rev3 would be through accurate replication of uracils (Figure 1, pathway (b→b2→b2.1→b2.1.1)). At the moment there are no mechanistic knowledge suggesting how this pathway can be facilitated in the absence of Rev3 TLS. In principle, wild-type Rev3 could enhance channeling of uracils into error-free lesion bypass (Figure 1 (b2.2)) thereby reducing a chance for C to T mutations via accurate replication of uracils (b→b2→b2 →b2.2.1). It is well known that uracils impede archaebacterial DNA polymerases (GREAGG *et al.* 1999; RICHARDSON *et al.* 2013). On the contrary, uracil is readily copied by prokaryotic as well as eukaryotic replicases *in vitro* (WARDLE *et al.* 2007). It remains to determine whether there is still some *in vivo* impediment to yeast replicases caused by uracils in DNA template. In support of possible direct effects of uracils in eukaryotes, recent studies indicated that uracils can trigger DNA replication stress in cancer cell lines even after siRNA suppression of Ung activity (SAXENA *et al.* 2024).

Long and persistent stretches of ssDNA formed in cancer cells are the substrate for APOBEC mutagenesis resulting in clusters of multiple mutations in cytosines. In agreement with mutation spectrum observed for long persistent ssDNA formed in yeast APOBEC-hypermutated clusters in cancers contained nearly equal numbers of C to T and C to G mutations (Figure 2 and Supplementary Figure S1). While an unknown fraction of scattered (non-clustered) APOBEC-induced mutations in APOBEC hypermutated tumors may also be generated in persistent ssDNA, another part of scattered mutations would be coming from short-lived ssDNA in the replication fork. Indeed, a fraction of APOBEC motif C to T substitutions among scattered mutations was greater than the C to T fraction within APOBEC-hypermutated clusters. We note that the fraction of C to G substitutions mutations within the scattered mutation group in APOBEC hypermutated clusters is still higher than in replicating yeast. It remains to establish if the high fraction of C to T mutations is due to better access of human uracil DNA glycosylase (hUdg) to short lived ssDNA in replication fork or to other differences between yeast and cancer cells. Regardless of specific reasons, our work indicates that the ratio between C to T and C to G APOBEC-induced mutations can serve as a potential readout of the interplay between RPA protection of ssDNA and uracil DNA glycosylase action on uracils in ssDNA. The relative differences between C to T and C to G ratios in yeast and scattered replication associated APOBEC-induced mutations in human cancers highlights key differences in the dynamics of replication in these systems. The higher amounts of C to G substitutions in human cancers could indicate that the human RPA complex has a lower strength of ssDNA binding compared to the yeast complex. Alternatively, greater C to G could indicate the extent of RPA exhaustion and replication stress conditions present within tumor cells.

DATA AVAILABILITY

All numeric and mutation call data necessary for confirming conclusions of this paper are presented in the text, in main and in supplementary display items, in supplementary tables and

in supplementary datasets. Supplementary tables also contain numeric source data for each graph on display items. Illumina sequence reads were submitted to NCBI Sequence Read Archive under BioProject PRJNA1128109

ACKNOWLEDGEMENTS
We thank Drs. Natalya P. Degtyareva and Rajula Elango for advice on the manuscript. This work was supported by US National Institutes of Health Intramural Research Program Project Z1AES103266 to D.A.G.; National Institutes of Health grants R01 CA218112 and R01 CA269784 (to S.A.R.).

REFERENCES

Banerjee, S., S. Smith, J. H. Oum, H. J. Liaw, J. Y. Hwang *et al.*, 2008 Mph1p promotes gross chromosomal rearrangement through partial inhibition of homologous recombination. J Cell Biol 181**:** 1083-1093.

Boiteux, S., and S. Jinks-Robertson, 2013 DNA repair mechanisms and the bypass of DNA damage in Saccharomyces cerevisiae. Genetics 193**:** 1025-1064.

Brown, A. L., C. D. Collins, S. Thompson, M. Coxon, T. M. Mertz *et al.*, 2021 Single-stranded DNA binding proteins influence APOBEC3A substrate preference. Sci Rep 11**:** 21008.

Chan, K., and D. A. Gordenin, 2015 Clusters of Multiple Mutations: Incidence and Molecular Mechanisms. Annu Rev Genet 49**:** 243-267.

Chan, K., M. A. Resnick and D. A. Gordenin, 2013 The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. DNA Repair (Amst) 12**:** 878-889.

Chan, K., S. A. Roberts, L. J. Klimczak, J. F. Sterling, N. Saini *et al.*, 2015 An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet 47**:** 1067-1072.

Chan, K., J. F. Sterling, S. A. Roberts, A. S. Bhagwat, M. A. Resnick *et al.*, 2012 Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. PLoS Genet 8**:** e1003149.

Chen, R., and M. S. Wold, 2014 Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. Bioessays 36**:** 1156-1161.

Consortium, I. T. P.-C. A. o. W. G., 2020 Pan-cancer analysis of whole genomes. Nature 578**:** 82-93.

Deng, S. K., H. Chen and L. S. Symington, 2015 Replication protein A prevents promiscuous annealing between short sequence homologies: Implications for genome integrity. Bioessays 37**:** 305-313.

Deng, S. K., B. Gibb, M. J. de Almeida, E. C. Greene and L. S. Symington, 2014 RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. Nat Struct Mol Biol 21**:** 405-412.

Elango, R., B. Osia, V. Harcy, E. Malc, P. A. Mieczkowski *et al.*, 2019 Repair of base damage within break-induced replication intermediates promotes kataegis associated with chromosome rearrangements. Nucleic Acids Res 47**:** 9666-9684.

Goldstein, A. L., and J. H. McCusker, 1999 Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast 15**:** 1541-1553.

Greagg, M. A., M. J. Fogg, G. Panayotou, S. J. Evans, B. A. Connolly *et al.*, 1999 A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. Proc Natl Acad Sci U S A 96**:** 9045-9050.

Harris, R. S., and J. P. Dudley, 2015 APOBECs and virus restriction. Virology 479-480**:** 131-145.

Hedglin, M., M. Aitha, A. Pedley and S. J. Benkovic, 2019 Replication protein A dynamically regulates monoubiquitination of proliferating cell nuclear antigen. J Biol Chem 294**:** 5157-5168.

Hedglin, M., and S. J. Benkovic, 2017a Eukaryotic Translesion DNA Synthesis on the Leading and Lagging Strands: Unique Detours around the Same Obstacle. Chem Rev 117**:** 7857-7877.

Hedglin, M., and S. J. Benkovic, 2017b Replication Protein A Prohibits Diffusion of the PCNA Sliding Clamp along Single-Stranded DNA. Biochemistry 56**:** 1824-1835.

Hoopes, J. I., L. M. Cortez, T. M. Mertz, E. P. Malc, P. A. Mieczkowski *et al.*, 2016 APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. Cell Rep 14**:** 1273-1282.

Hoopes, J. I., A. L. Hughes, L. A. Hobson, L. M. Cortez, A. J. Brown *et al.*, 2017 Avoidance of APOBEC3B-induced mutation by error-free lesion bypass. Nucleic Acids Res 45**:** 5243-5254.

Hudson, K. M., L. J. Klimczak, J. F. Sterling, A. B. Burkholder, M. D. Kazanov *et al.*, 2023 Glycidamide-induced hypermutation in yeast single-stranded DNA reveals a ubiquitous clock-like mutational motif in humans. Nucleic Acids Res 51**:** 9075-9100.

Jin, Y. H., A. B. Clark, R. J. Slebos, H. Al-Refai, J. A. Taylor *et al.*, 2003 Cadmium is a mutagen that acts by inhibiting mismatch repair. Nat Genet 34**:** 326-329.

Kouno, T., T. V. Silvas, B. J. Hilbert, S. M. D. Shandilya, M. F. Bohn *et al.*, 2017 Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. Nat Commun 8**:** 15024.

Krokan, H. E., P. Saetrom, P. A. Aas, H. S. Pettersen, B. Kavli *et al.*, 2014 Error-free versus mutagenic processing of genomic uracil--relevance to cancer. DNA Repair (Amst) 19**:** 38-47.

Lada, A. G., I. S. Waisertreiger, C. E. Grabow, A. Prakash, G. E. Borgstahl *et al.*, 2011 Replication protein A (RPA) hampers the processive action of APOBEC3G cytosine deaminase on single-stranded DNA. PLoS One 6**:** e24848.

Lerner, T., F. N. Papavasiliou and R. Pecori, 2018 RNA Editors, Cofactors, and mRNA Targets: An Overview of the C-to-U RNA Editing Machinery and Its Implication in Human Disease. Genes (Basel) 10.

Martin, S. K., and R. D. Wood, 2019 DNA polymerase ζ in DNA replication and repair. Nucleic Acids Research 47**:** 8348-8361.

Mertz, T. M., C. D. Collins, M. Dennis, M. Coxon and S. A. Roberts, 2022 APOBEC-Induced Mutagenesis in Cancer. Annu Rev Genet.

Petljak, M., A. Dananberg, K. Chu, E. N. Bergstrom, J. Striepen *et al.*, 2022 Mechanisms of APOBEC3 mutagenesis in human cancer cells. Nature 607**:** 799-807.

Petljak, M., and J. Maciejowski, 2020 Molecular Origins of APOBEC-Associated Mutations in Cancer. DNA Repair**:** 102905.

Refsland, E. W., and R. S. Harris, 2013 The APOBEC3 family of retroelement restriction factors. Curr Top Microbiol Immunol 371**:** 1-27.

Richardson, T. T., L. Gilroy, Y. Ishino, B. A. Connolly and G. Henneke, 2013 Novel inhibition of archaeal family-D DNA polymerase by uracil. Nucleic Acids Research 41**:** 4207-4218.

Roberts, S. A., M. S. Lawrence, L. J. Klimczak, S. A. Grimm, D. Fargo *et al.*, 2013 An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet 45**:** 970-976.

Roberts, S. A., J. Sterling, C. Thompson, S. Harris, D. Mav *et al.*, 2012 Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell 46**:** 424-435.

Ruff, P., R. A. Donnianni, E. Glancy, J. Oh and L. S. Symington, 2016 RPA Stabilization of Single-Stranded DNA Is Critical for Break-Induced Replication. Cell Rep 17**:** 3359-3368.

Saini, N., and D. A. Gordenin, 2020 Hypermutation in single-stranded DNA. DNA Repair (Amst) 91-92**:** 102868.

Saini, N., S. A. Roberts, J. F. Sterling, E. P. Malc, P. A. Mieczkowski *et al.*, 2017 APOBEC3B cytidine deaminase targets the non-transcribed strand of tRNA genes in yeast. DNA Repair (Amst) 53**:** 4-14.

Sakofsky, C. J., N. Saini, L. J. Klimczak, K. Chan, E. P. Malc *et al.*, 2019 Repair of multiple simultaneous double-strand breaks causes bursts of genome-wide clustered hypermutation. PLoS Biol 17**:** e3000464.

Salter, J. D., R. P. Bennett and H. C. Smith, 2016 The APOBEC Protein Family: United by Structure, Divergent in Function. Trends Biochem Sci 41**:** 578-594.

Saxena, S., P. S. Patel, C. S. Nabel, A. S. Kawale, C. R. Crosby *et al.*, 2024 Unprocessed Genomic Uracil as a Source of DNA Replication Stress in Cancer Cells. bioRxiv**:** 2024.2002.2005.578390.

Shcherbakova, P. V., and Y. I. Pavlov, 1996 3'-->5' exonucleases of DNA polymerases epsilon and delta correct base analog induced DNA replication errors on opposite DNA strands in Saccharomyces cerevisiae. Genetics 142**:** 717-726.

Sherman, F., G. R. Fink, J. B. Hicks and L. Cold Spring Harbor, 1981 *Methods in yeast genetics : Laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Taylor, B. J., S. Nik-Zainal, Y. L. Wu, L. A. Stebbings, K. Raine *et al.*, 2013 DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. Elife 2**:** e00534.

Toledo, L., K. J. Neelsen and J. Lukas, 2017 Replication Catastrophe: When a Checkpoint Fails because of Exhaustion. Mol Cell 66**:** 735-749.

Toledo, L. I., M. Altmeyer, M. B. Rask, C. Lukas, D. H. Larsen *et al.*, 2013 ATR prohibits replication catastrophe by preventing global exhaustion of RPA. Cell 155**:** 1088-1103.

Wardle, J., P. M. J. Burgers, I. K. O. Cann, K. Darley, P. Heslop *et al.*, 2007 Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. Nucleic Acids Research 36**:** 705-711.

Wong, L., F. S. Vizeacoumar, F. J. Vizeacoumar and L. Chelico, 2021 APOBEC1 cytosine deaminase activity on single-stranded DNA is suppressed by replication protein A. Nucleic Acids Res 49**:** 322-339.

Figure Legends.

Figure 1. Channeling a uracil (U) created by APOBEC deamination of cytosine (C) in transient single-stranded (ss) DNA via error-prone and error-free replication and repair pathways. Pathways and steps are identified by the combined letter and legal numbering styles. **(a)** Transient ssDNA intermediates can be formed through a range of DNA replication, transcription, and repair events (see Figures 1 and 2 in (SAINI AND GORDENIN 2020) and references therein). **(b)** ssDNA-specific APOBEC cytosine deaminase converts C to U and then is processed via sub-pathways **(b1)** and **(b2)**, which are color-coded for distinction. Sub-pathway **(b1)** starts from abasic site (AP, shown as a small ball) created by uracil-DNA glycosylase (Ung1 in yeast). AP-site can lead to DNA breakage, rearrangements to chromosome loss - box **(c)**. It can also result in base-substitutions. Another sub-pathway involves unchanged U (sub-pathway **(b2**)). Restoration of transient ssDNA containing an AP site to double-stranded (ds) form can be performed by DNA polymerase(s), including by Pol-zeta-dependent error prone translesion synthesis (TLS) **(b.1.1)**, Next round of DNA replication **(b.1.1.1)** will fix a base substitution or a non-mutant sequence. Restoration to of ssDNA to dsDNA can also occur by re-annealing (sub-pathways **(b.1.2)**) and **(b.1.2)**. Re-annealing of ssDNA can involve a complementary strand of the same DNA molecule when ssDNA was formed by transient unwinding or within the R-loop transcription intermediate. It can also involve a complimentary strand of a sister DNA molecule, if re-annealing occurs via replication fork regression. Re-annealed of AP-containing dsDNA can be repaired by base-excision repair (BER) utilizing a complimentary strand with wild type sequence as a template **(b.1.2)**, restoring wild type dsDNA sequence. Similarly, re-annealing involving U-containing ssDNA **(b.2.1)** would be a subject to BER fixing the wild type sequence in resulting dsDNA. On the contrary, two rounds of accurate replication of U-containing ssDNA **(b.2.2)** and **(b.2.2.1)** would always generate a C→T mutation.

Figure 2. APOBEC-induced C to T and C to G mutagenesis in the data from PCAWG cancers and from yeast model studies. Yeast data include all C to T and C to G mutations. Counts for sections of yeast genome rendered single-stranded (subtelomeric 5'→3' resection in uncapped telomeres or DSB-induced BIR (CHAN *et al.* 2012; CHAN *et al.* 2013; HOOPES *et al.* 2017; ELANGO *et al.* 2019)) were totaled into "yeast ssDNA" category. Counts for whole-genome sequenced yeast cultures studies (TAYLOR *et al.* 2013; HOOPES *et al.* 2016; HOOPES *et al.* 2017; SAINI *et al.* 2017) were totaled into "yeast replication" category. Counts for individual yeast studies can be found in Supplementary Table 1. Cancer data include only C to T and C to G mutations found in tCw mutational motifs. Cancer data are totaled for all cancer types. Counts for each cancer type can be found in Supplementary Table 2, Figure S1 and in Supplementary Dataset 1. **** represents two-sided chi-squared P-value<0.0001

Figure 3. Yeast strains with *ura3-29* single-base mutation reporter for selection of three possible base substitutions resulting from TLS across AP-site. Details of all constructs of mid-chromosome triple-gene reporter were described in (ROBERTS *et al.* 2012; HOOPES *et al.* 2016) and in Materials and Methods. **A.** *ura3-29* mutation was placed within the triple-gene (*ADE2*, *URA3*, *CAN1*) mutation reporter, which was inserted into the LYS2 native ORF. Each of the three genes were deleted in their native positions. In addition to single base *ura3-29* reversion, *CAN1* gene allows to select forward mutations in the entire 1.6 kB ORF. **B.** Two orientations of triple-gene reporter around strong replication origin ARS216 place the mutant cytosine of the *ura3-29* in the lagging ("Left" construct; native position of LYS2 as in ySR128) or

in the leading ("Right" construct originating from ySR366) strand template. Strain's complete genotypes are listed in Supplementary Table 4.

Figure 4. Stimulation of APOBEC mutagenesis in yeast strains carrying hypomorphic allele of RFA1. CanR mutation rates were measured in *RFA1* and *rfa1-t33* strains, all carrying a deletion of *ung1*. Shown are median values for mutation rates measured in 6-12 independent cultures and 95% confidence intervals for the medians. Statistically significant difference with P<0.05 of the two-tailed Mann-Whitney test is shown by brackets. Strains expressing A3B showed statistical significance only in one-tailed Mann-Whitney test. Source data and statistical analyses can be found in **Supplementary Table 4**.

Figure 5. Effects of replication fork direction and strain genotype on rates of APOBEC3A-induced mutagenesis in *ura3-29* reporter. For all strains mutation rates were determined in 6 independent cultures. Rate of induced mutagenesis in each strain was calculated by subtracting median mutation rate in empty vector strain from a median mutation rate in A3A-expressing strain of the same genotype. Statistical comparison was performed by comparing sets of rates in a pair of strains by Mann-Whitney two-sided test. Pairs of strains showing statistically significant differences between mutation rates are connected with brackets. P-values are indicated as: * <0.05; ** < 0.005. **A, B** – *ura3-29* APOBEC3A-induced reversion rates in ung1- and in UNG1+ strains, respectively. Source data for panels **A** and **B** including details of statistical comparisons for P-values<0.05 can be found in Supplementary Table 5.

Figure 6. Base substitution spectra of APOBEC3A-induced ura3-29 changes in Ura$^+$ reversion events. A. Mutation spectrum in *ung1Δ* strains. **B.** Mutation spectrum on UNG1-WT strains. Shown are percentages and numbers of base substitutions in ura3-29 mutation position of various genotypes. A single isolate without base substitution in ura3-29 position also included. For comparisons between spectra C to G and C to A changes were pulled to a single category. The only event with no changes in ura3-29 site was not included. Pairwise comparisons were performed using two-sided Fisher's exact test. There were no statistically significant differences in spectra between strains of the same genotype carrying ura3-29 to the left (L) or to the right (R) of ARS216 (see Figure 3). Spectra in *RFA1-WT* strains showed strong difference from spectra in *rfa1-t33* strains (P<0.0001). Source data and complete outputs of statistical analyses of differences between spectra in different genotypes can be found in Supplementary Table 6.

Figure 7. Whole-genome APOBEC3A-induced mutagenesis in yeast strains carrying wild-type allele of *UNG1*.
Mutation calls can be found in Supplementary Dataset 2.
A. Densities of APOBEC3A-unduced C:G base pair substitutions per sequenced genome (see Materials and Methods). Shown are substitutions in cytosines (C) of both DNA strands. Pairwise statistical comparisons between genotypes were performed by two-sided Mann-Whitney test comparing sets of C:G pair substitution counts in the sets of sequenced genomes of each genotype. All pairs with one exception produced P<0.05. For the comparison between sequenced isolates of *RFA1 REV3* and *rfa1-t33 REV3* strains, the Mann-Whitney test did not show statistically significance, however statistically significant differences of the mean values were confirmed by t-test which assumed that standard deviations of mutation calls were the same for the sets of sequenced isolates of these strains. Source data and complete outputs of

statistical analyses of differences between different genotypes can be found in Supplementary Table 7.
B. Total counts and fractions of total mutation counts for different substitutions in C:G base pairs of all genomes of a given genotype. All pairwise comparison between all pairs of genotypes by Chi-square showed strong differences (P<0.0001). Source data and complete outputs of statistical analyses of differences between different genotypes can be found in Supplementary Table 7.

Figure S1. C to T and C to G mutations in APOBEC-hypermutated cancers. Shown are counts of mutation in APOBEC motifs tCw to T and tCw to G.
Source data for this figure can be found in Supplementary Table 2. Cancer mutation catalogues and APOBEC analyses results can be found in PCAWG project data repository (CONSORTIUM 2020). Relevant summaries can be found in Supplementary Dataset 1. All mutation counts were taken from a row "Total" in columns 'APOBEC_to_T' or 'APOBEC_to_G'of a cancer type. Separately calculated were values for mutations in C- or G-strand-coordinated clust ers (files * _sum_G_C_clusters_fisher_Pcorr.txt in Suplementary Dataset 1) or for mutations not assigned to any cluster (scattered mutations; files * _sum_scattered_fisher_Pcorr.txt). Cancer types are abbreviated as in TCGA (https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations).

Figure S2. Comparisons of CanR mutation rates in yeast strains carrying vectors expressing different APOBEC enzymes with CanR rates in strains with empty vector. Filled bars correspond to APOBEC-expressing strains showing statistically significant (P<0.05) difference from the empty vector strain of the same genotype; numbers above filled bars indicate fold increases over empty vector. Source data and statistical analyses can be found in Supplementary Table 4. **A.** *ung1Δ RFA1-WT* strains. **B.** *ung1Δ rfa1-t33* strains.

Figure S3. Calculated densities of APOBEC3A-unduced substitutions in C:G and A:T base pairs per sequenced genome.

Supplementary Items

Supplementary Table 1. APOBEC-induced C to T and C to G mutations in different yeast studies.

Supplementary Table 2. APOBEC enrichments and APOBEC mutation loads in APOBEC-hypermutated cancer types from PCAWG dataset.

Supplementary Table 3. Source data and statistical analysis for Figure 2

Supplementary Table 4. Source data and statistical analyses for Figure 3, Figure 4 and Figure S2.

Supplementary Table 5. Source data for Figure 5 A, B.

Supplementary Table 6. Source data for Figure 6

Supplementary Table 7. Source data for Figure 7 and Figure S3

Supplementary Dataset 1. Summaries for APOBEC mutagenesis analyses in APOBEC-hypermutated cancers (see details in the Figure S1 legend).

Supplementary Dataset 2. Mutation calls in yeast clones analyzed by whole-genome sequencing. Each file contains mutation calls in the independent clones of a strain identified in the filename. LT and RT folders contain mutation calls in strains with triple reporter located to the left or to the right of ARS216, respectively.

Supplementary Dataset 3. APOBEC-expression vectors and empty vector plasmids in GenBank format.

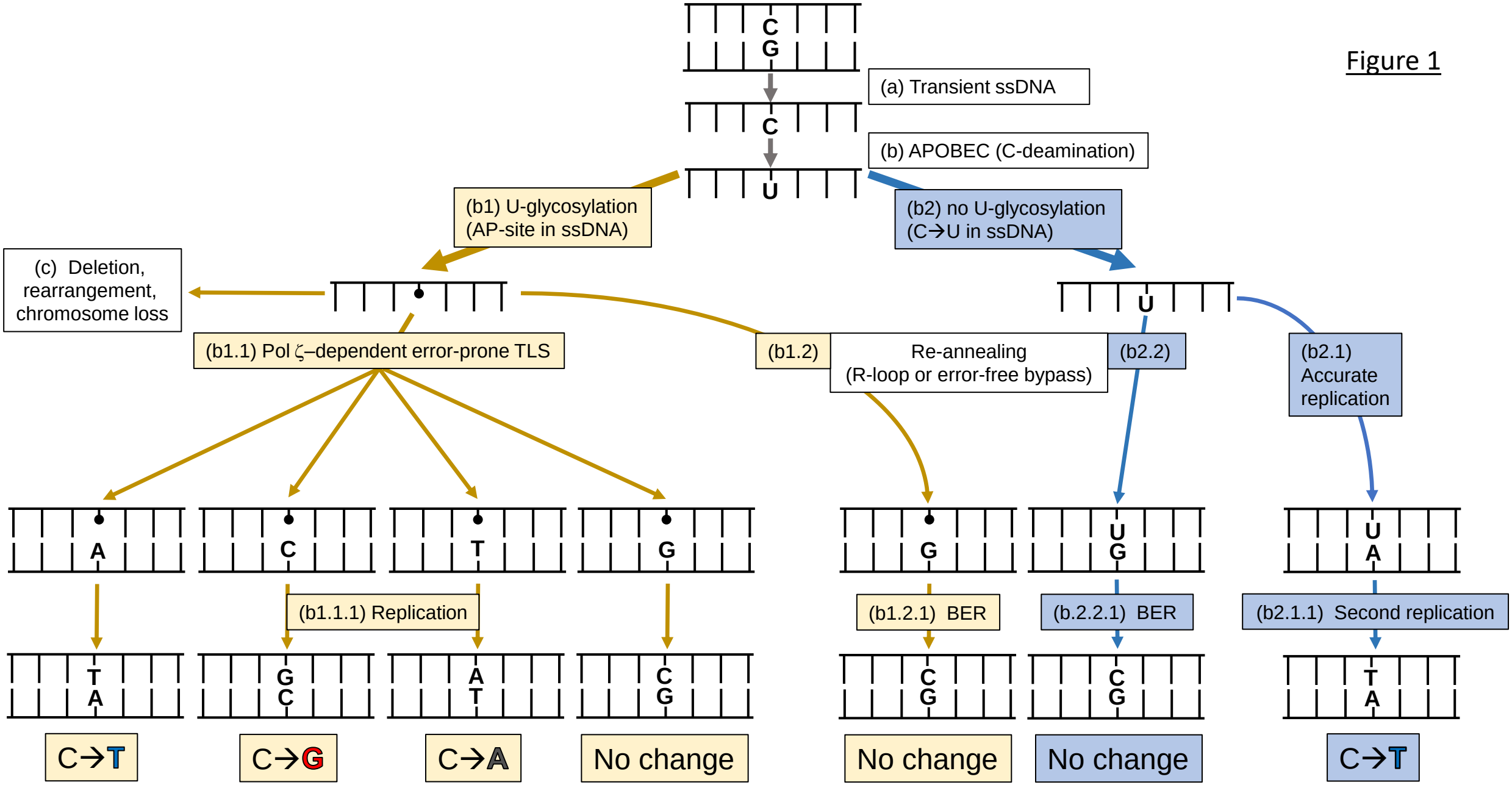


Figure 1
C
G
(a) Transient ssDNA
C
(b) APOBEC (C-deamination)
U
(b1) U-glycosylation
(AP-site in ssDNA)
(b2) no U-glycosylation
(C→U in ssDNA)
(c) Deletion, rearrangement, chromosome loss
U
(b1.1) Pol ζ–dependent error-prone TLS
(b1.2)
Re-annealing
(R-loop or error-free bypass)
(b2.2)
(b2.1)
Accurate
replication
A
C
T
G
G
U
G
U
A
(b1.1.1) Replication
(b1.2.1) BER
(b.2.2.1) BER
(b2.1.1) Second replication
T
A
G
C
A
T
C
G
C
G
C
G
T
A
C→T
C→G
C→A
No change
No change
No change
C→T

Figure 2

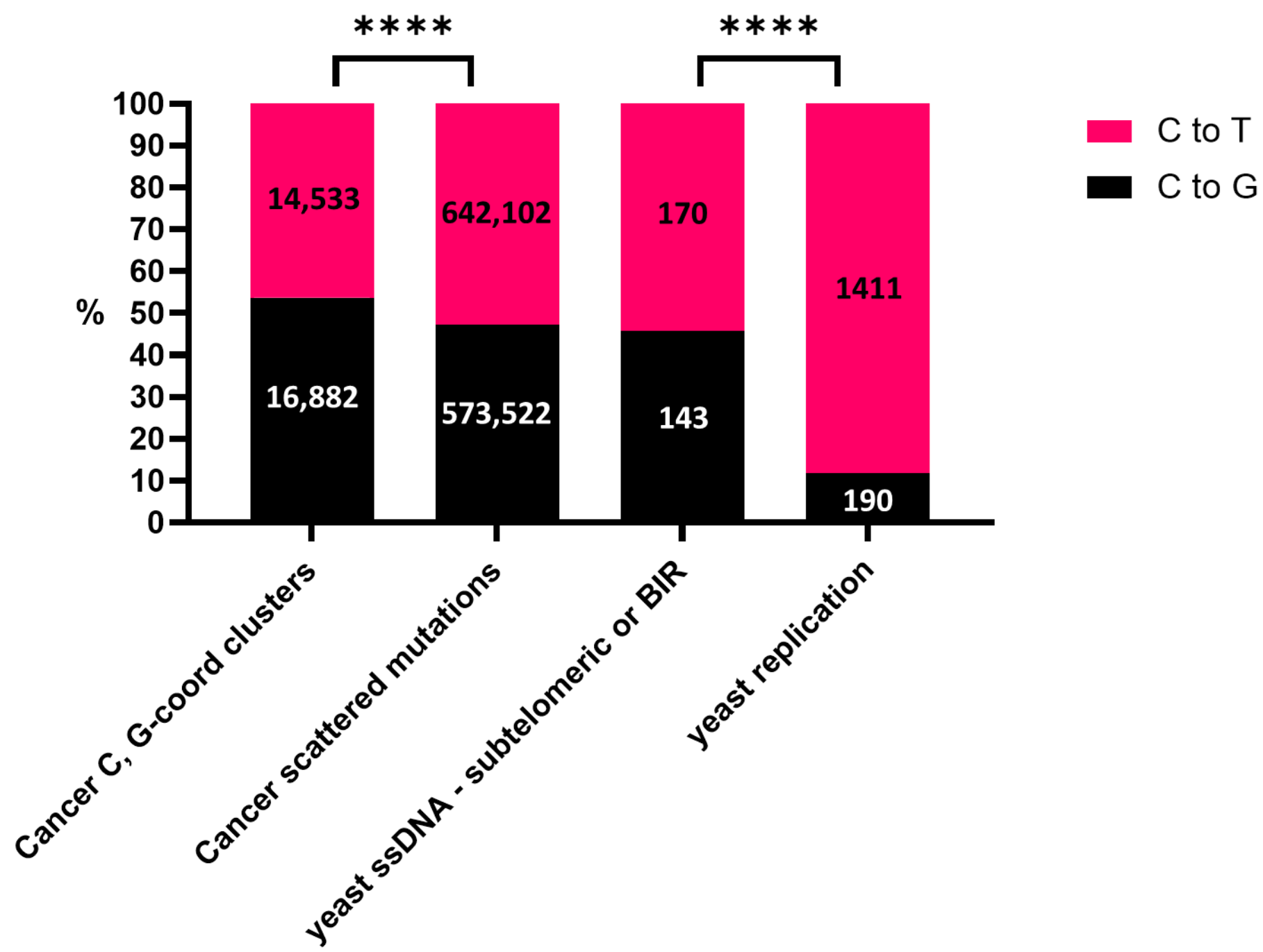
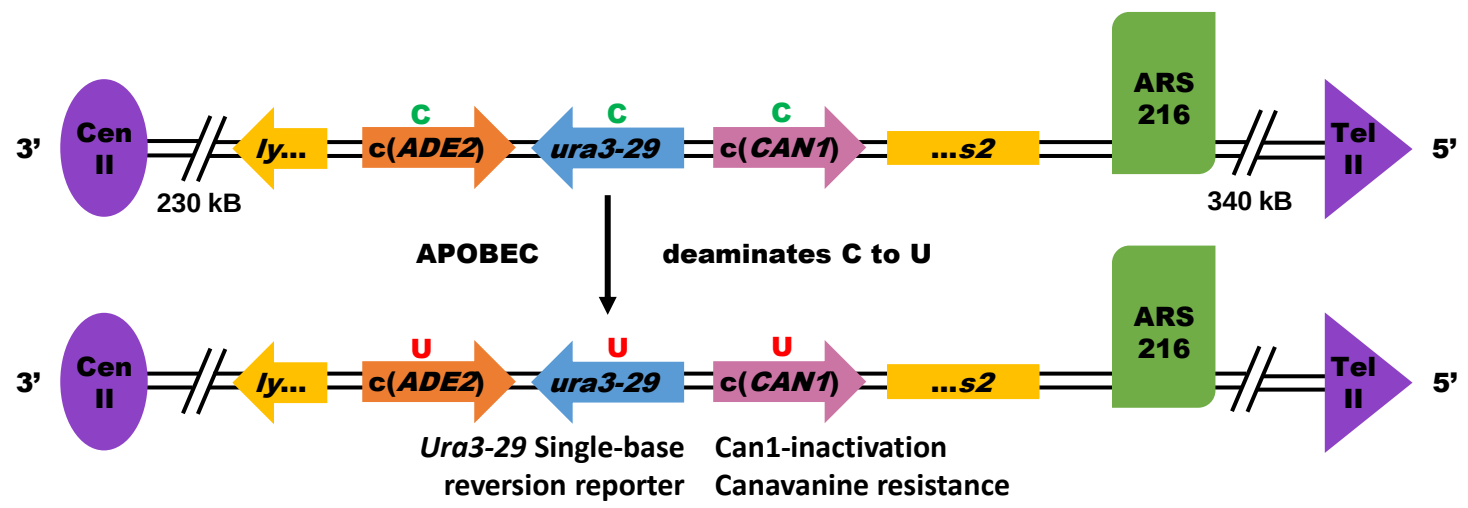
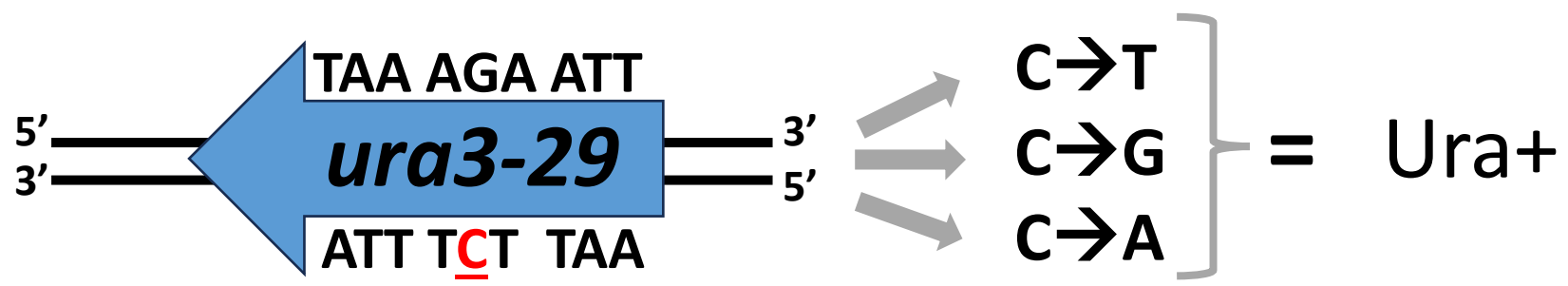

****
****
100
90
80
70
60
%
50
40
30
20
10
0
14,533
642,102
170
1411
16,882
573,522
143
190
C to T
C to G
Cancer C, G-coord clusters
Cancer scattered mutations
yeast ssDNA - subtelomeric or BIR
yeast replication

Figure 3

A


3'
Cen II
230 kB
ly...
c(ADE2)
ura3-29
c(CAN1)
...s2
ARS 216
340 kB
Tel II
5'
C
C
C
APOBEC
deaminates C to U
U
U
U
Ura3-29 Single-base reversion reporter
Can1-inactivation Canavanine resistance


5'
3'
TAA AGA ATT
ura3-29
ATT TCT TAA
3'
5'
C→T
C→G
C→A
= Ura+

B

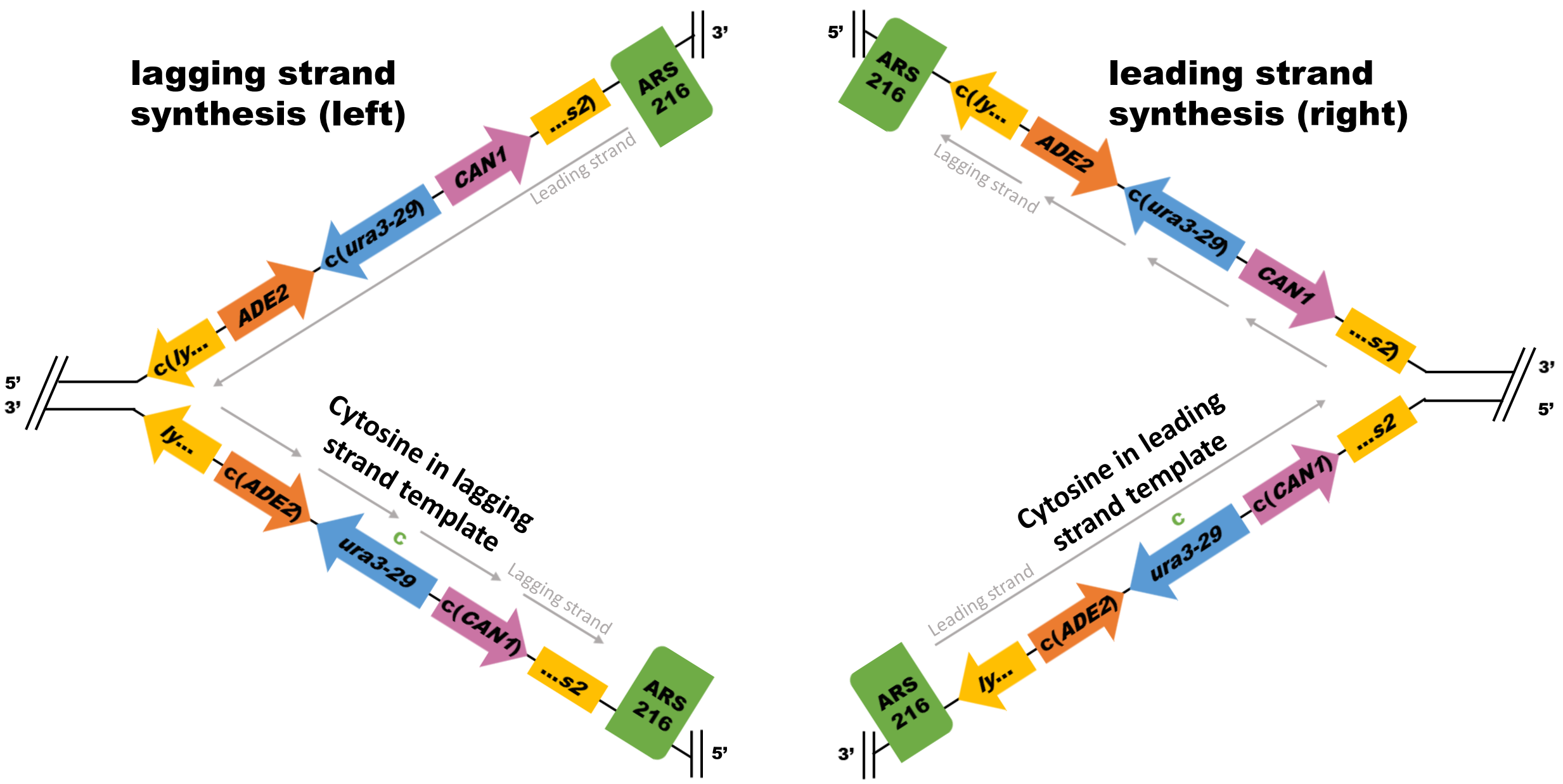

lagging strand synthesis (left)
leading strand synthesis (right)
Cytosine in lagging strand template
Cytosine in leading strand template
Leading strand
Lagging strand
ARS 216
...s2
CAN1
c(ura3-29)
ADE2
c(ly...
ly...
c(ADE2)
ura3-29
c(CAN1)
5'
3'

Figure 4

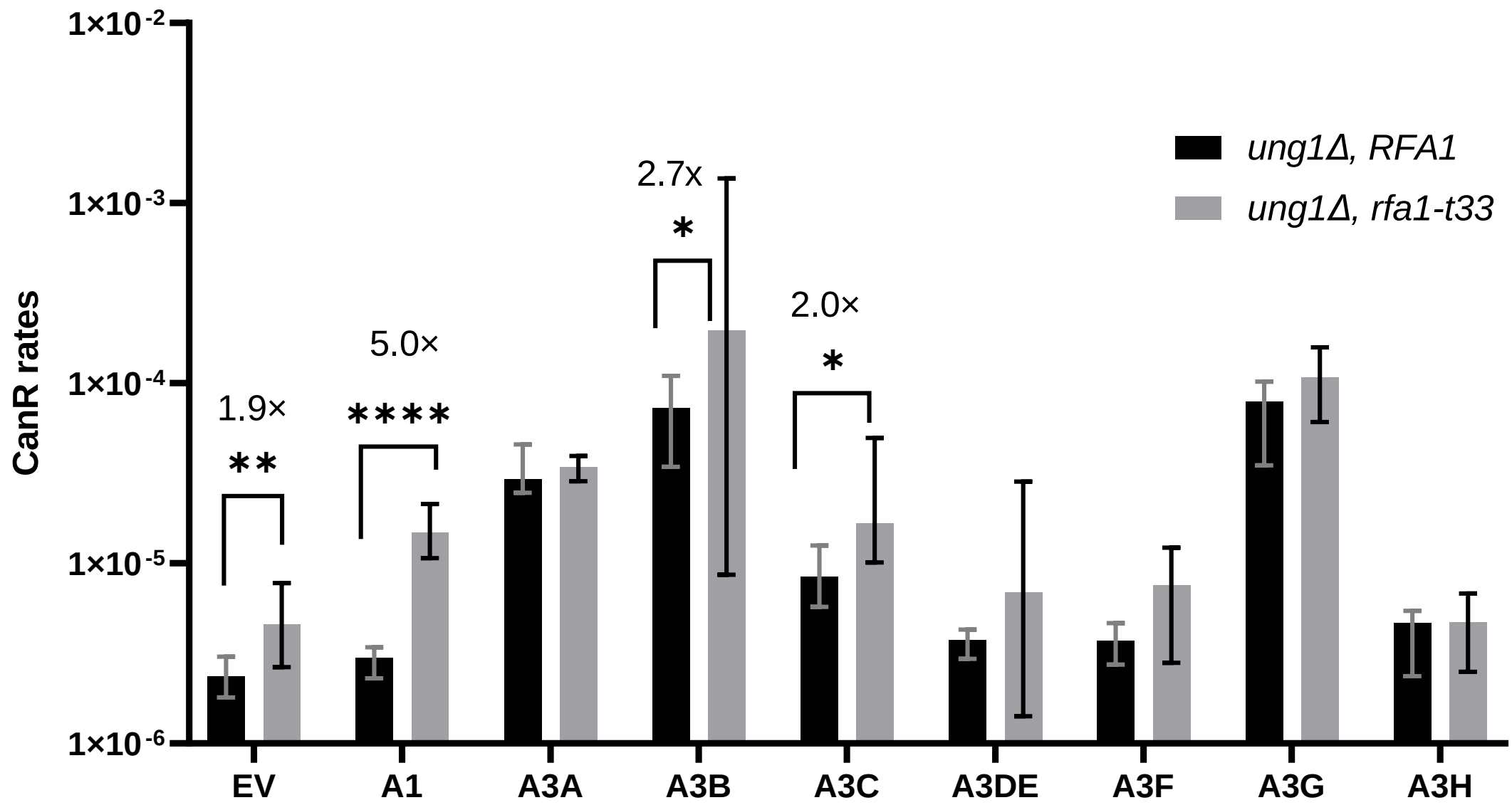
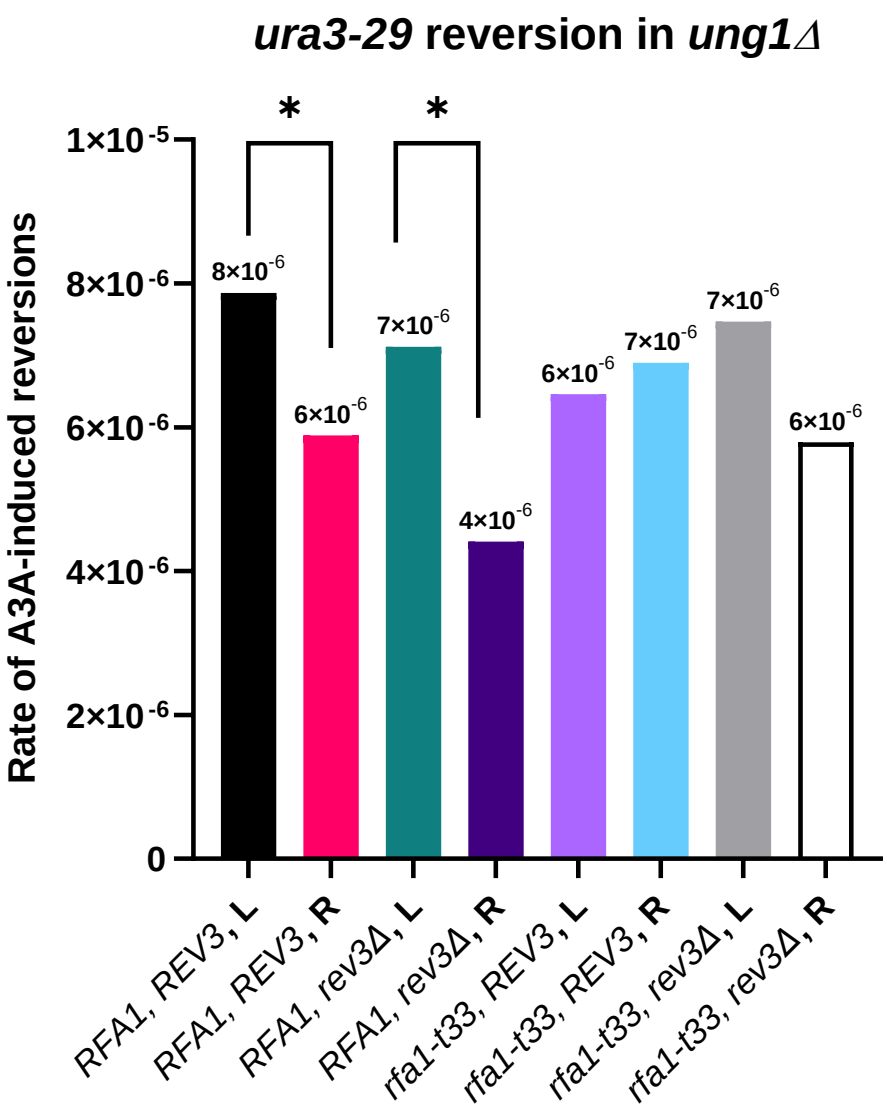

1×10-2
1×10-3
1×10-4
1×10-5
1×10-6
CanR rates
ung1Δ, RFA1
ung1Δ, rfa1-t33
1.9×
**
5.0×
****
2.7x
*
2.0×
*
EV
A1
A3A
A3B
A3C
A3DE
A3F
A3G
A3H

Figure 5

A


ura3-29 reversion in ung1Δ
Rate of A3A-induced reversions
1×10-5
8×10-6
6×10-6
4×10-6
2×10-6
0
*
*
8×10-6
6×10-6
7×10-6
4×10-6
6×10-6
7×10-6
7×10-6
6×10-6
RFA1, REV3, L
RFA1, REV3, R
RFA1, rev3Δ, L
RFA1, rev3Δ, R
rfa1-t33, REV3, L
rfa1-t33, REV3, R
rfa1-t33, rev3Δ, L
rfa1-t33, rev3Δ, R

B

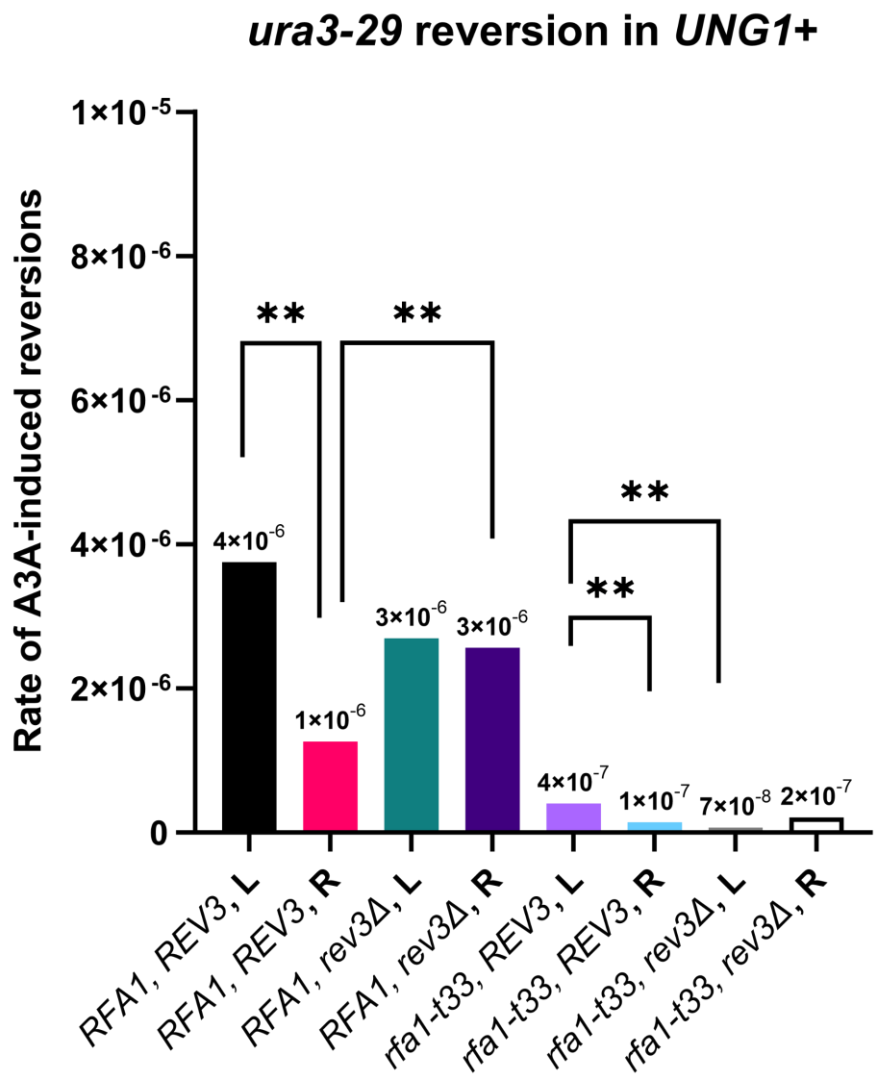
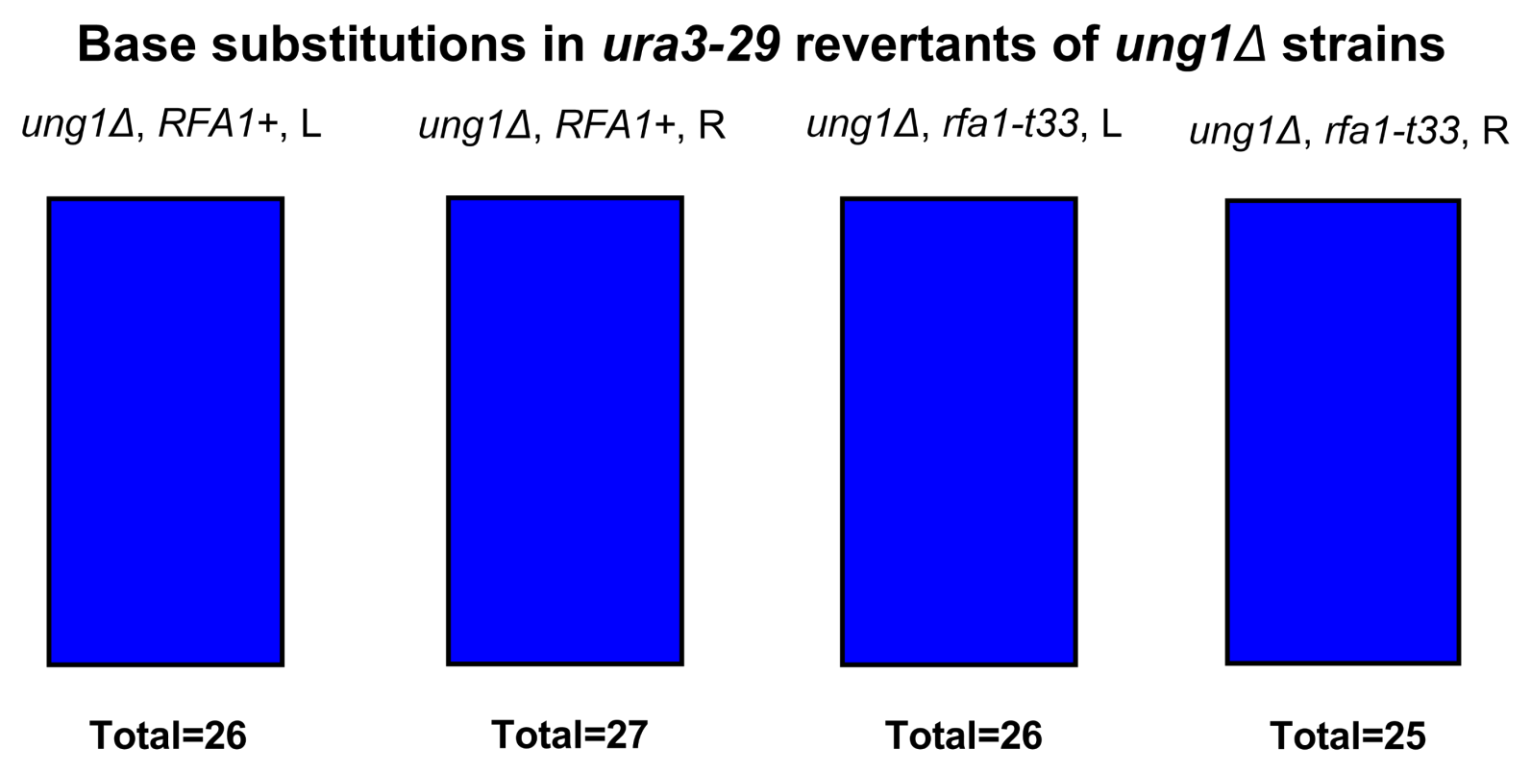

ura3-29 reversion in UNG1+
Rate of A3A-induced reversions
1×10-5
8×10-6
6×10-6
4×10-6
2×10-6
0
**
**
**
**
4×10-6
1×10-6
3×10-6
3×10-6
4×10-7
1×10-7
7×10-8
2×10-7
RFA1, REV3, L
RFA1, REV3, R
RFA1, rev3Δ, L
RFA1, rev3Δ, R
rfa1-t33, REV3, L
rfa1-t33, REV3, R
rfa1-t33, rev3Δ, L
rfa1-t33, rev3Δ, R

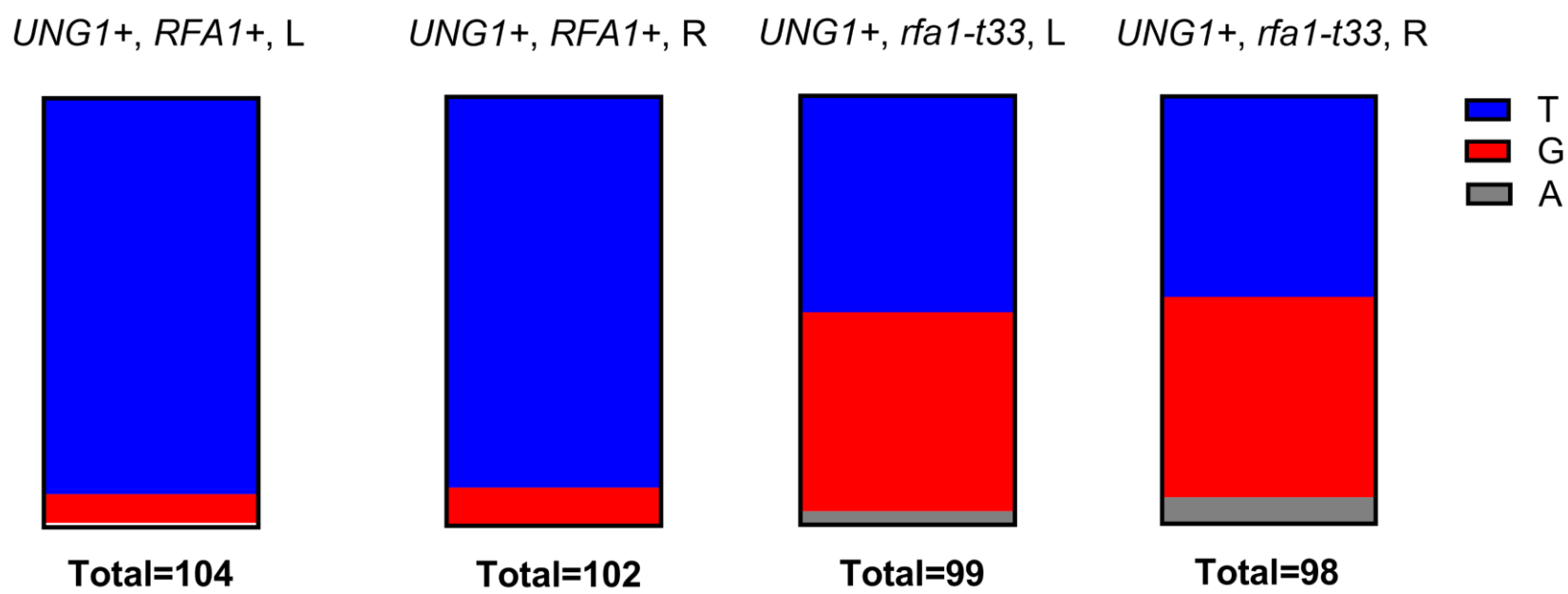
Figure 6

A



T	26	27	26	25
G	0	0	0	0
A	0	0	0	0
C	0	0	0	0

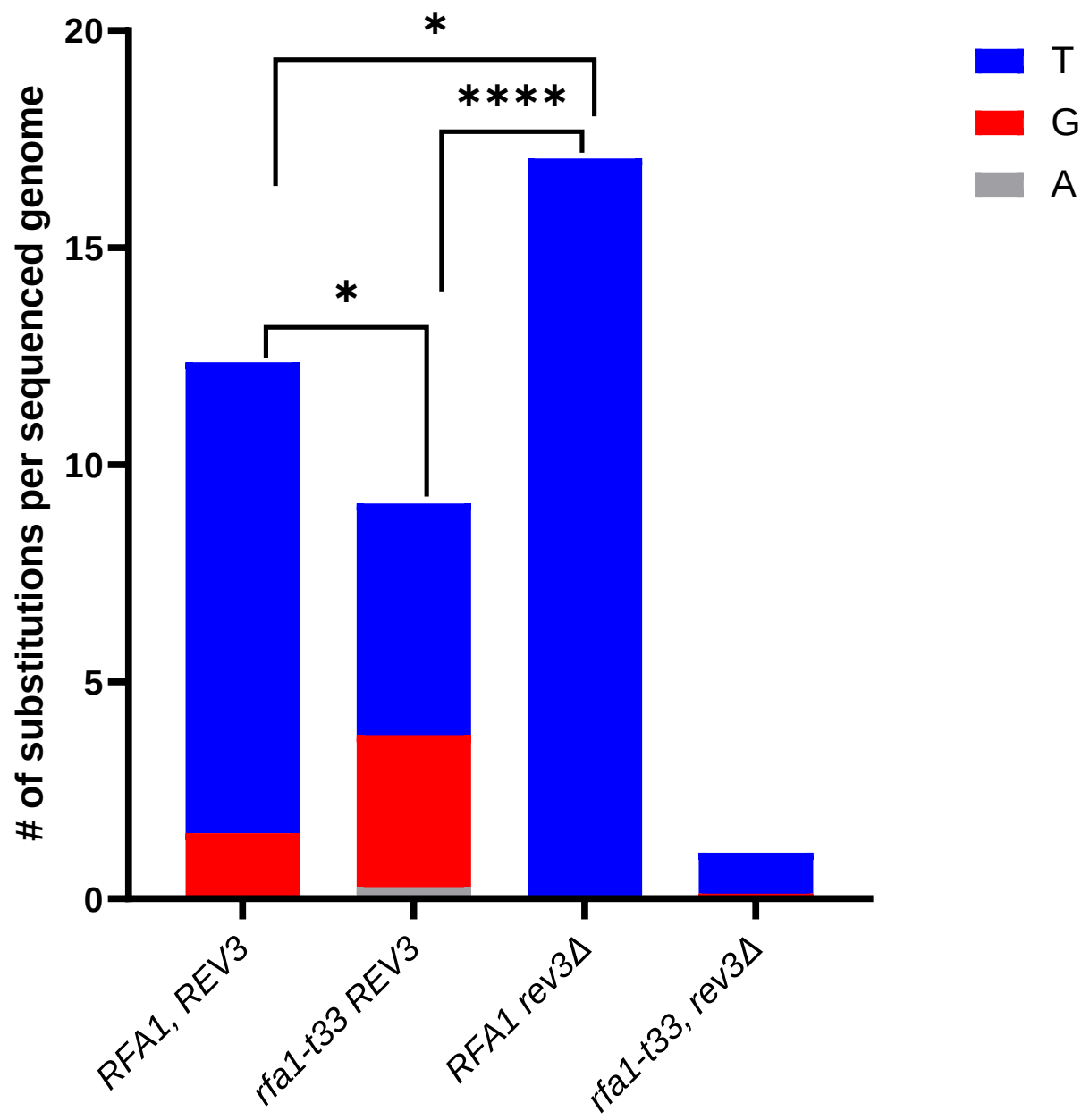
B



T	96	93	50	46
G	7	9	46	46
A	0	0	3	6
C	1	0	0	0

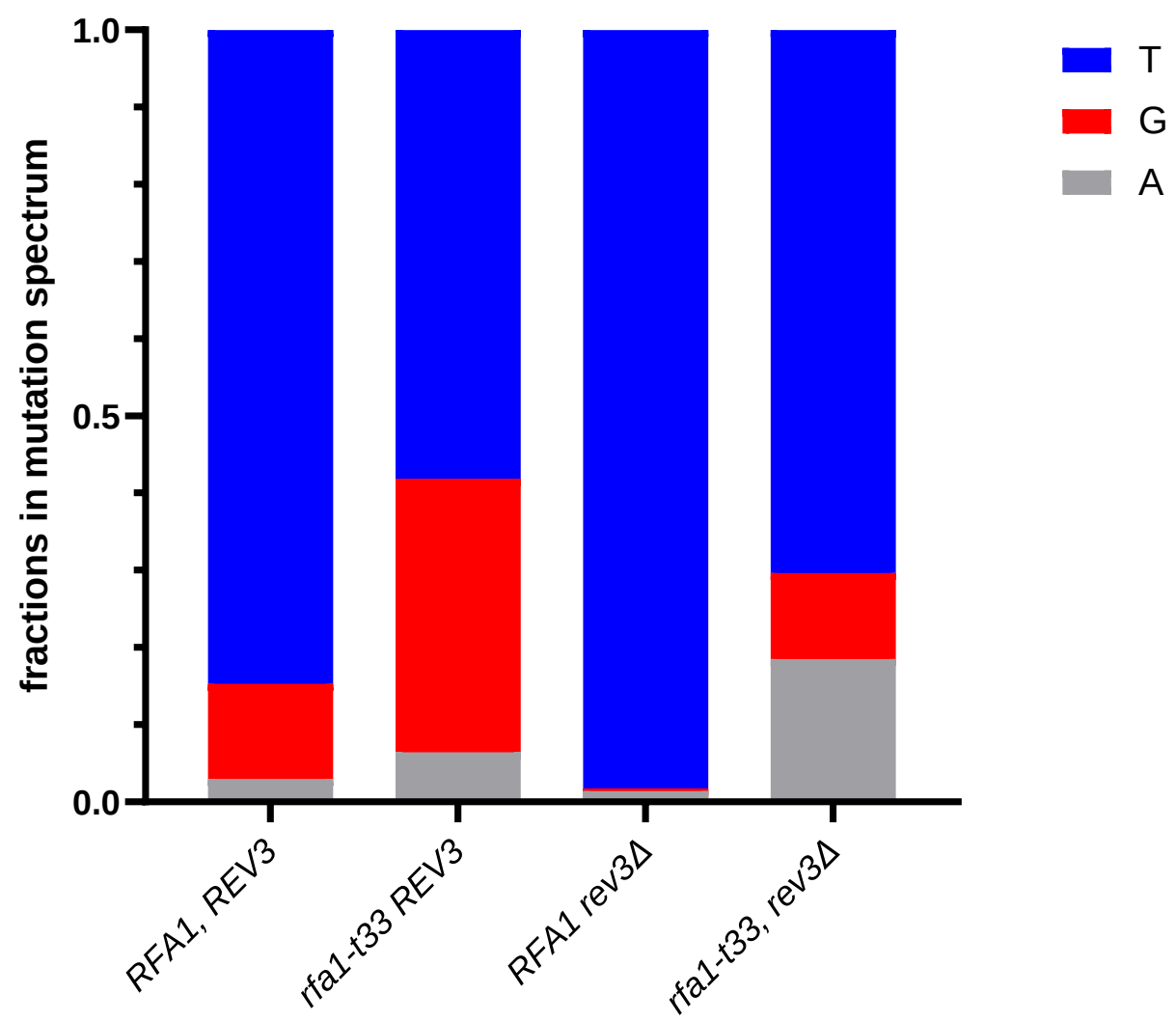
Figure 7

A



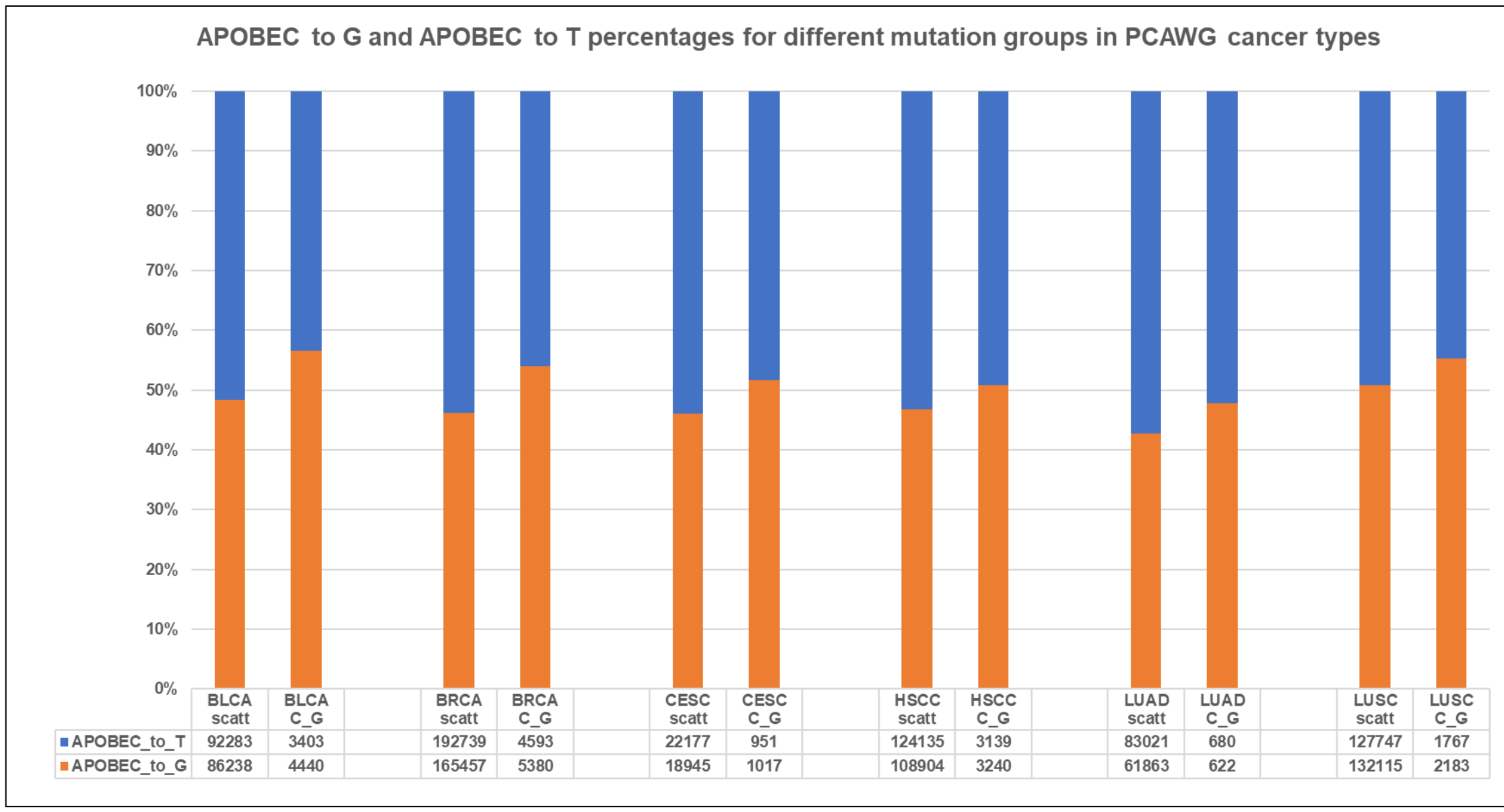
C to T	10.85	5.34	17.06	0.94
C to G	1.51	3.5	0	0.09
C to A	0	0.27	0	0.03

B



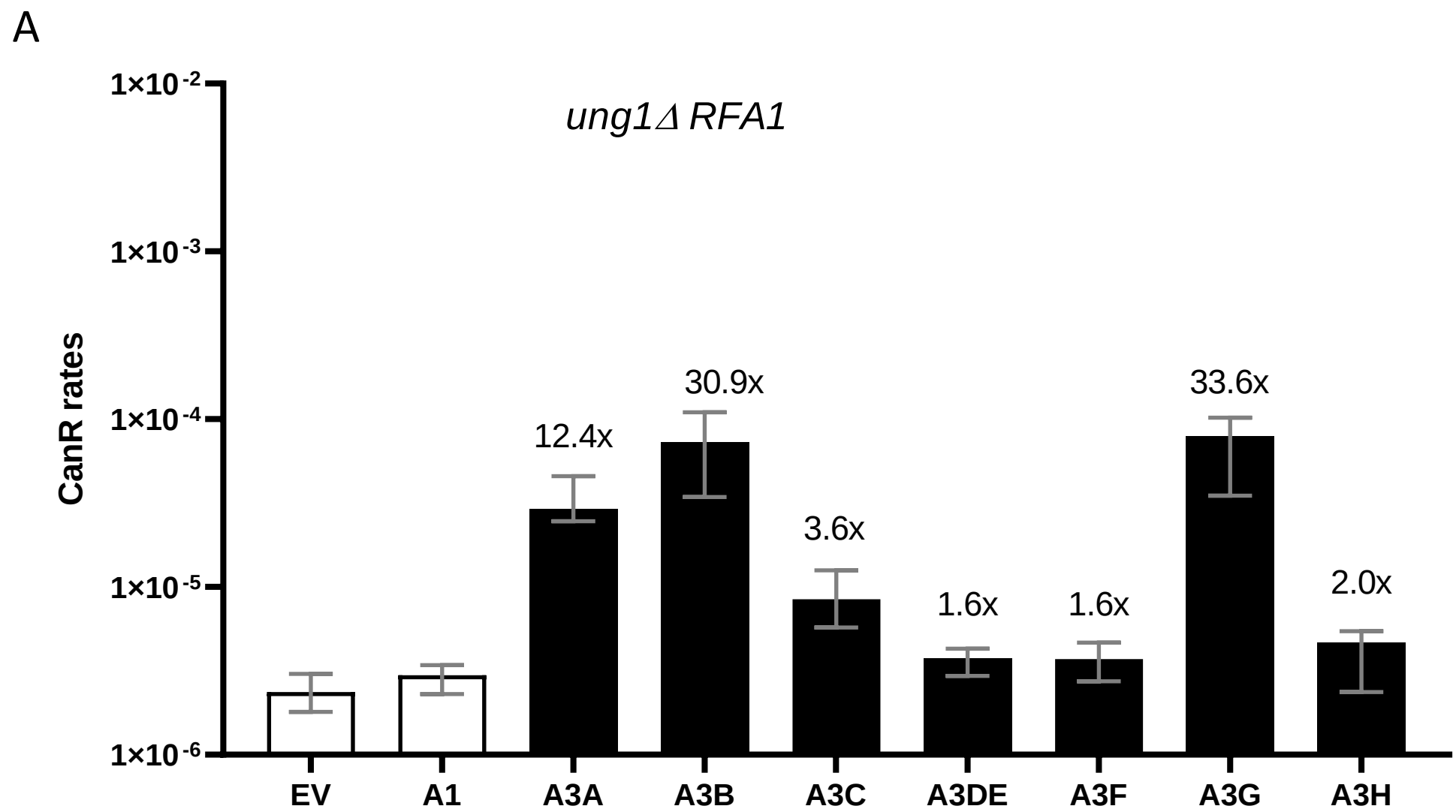
C to T	542	289	852	57
C to G	79	176	3	9
C to A	19	32	12	15

Figure S1



	BLCA scatt	BLCA C_G	BRCA scatt	BRCA C_G	CESC scatt	CESC C_G	HSCC scatt	HSCC C_G	LUAD scatt	LUAD C_G	LUSC scatt	LUSC C_G
APOBEC_to_T	92283	3403	192739	4593	22177	951	124135	3139	83021	680	127747	1767
APOBEC_to_G	86238	4440	165457	5380	18945	1017	108904	3240	61863	622	132115	2183

Figure S2


A
ung1Δ RFA1
CanR rates
1×10-2
1×10-3
1×10-4
1×10-5
1×10-6
12.4x
30.9x
3.6x
1.6x
1.6x
33.6x
2.0x
EV
A1
A3A
A3B
A3C
A3DE
A3F
A3G
A3H

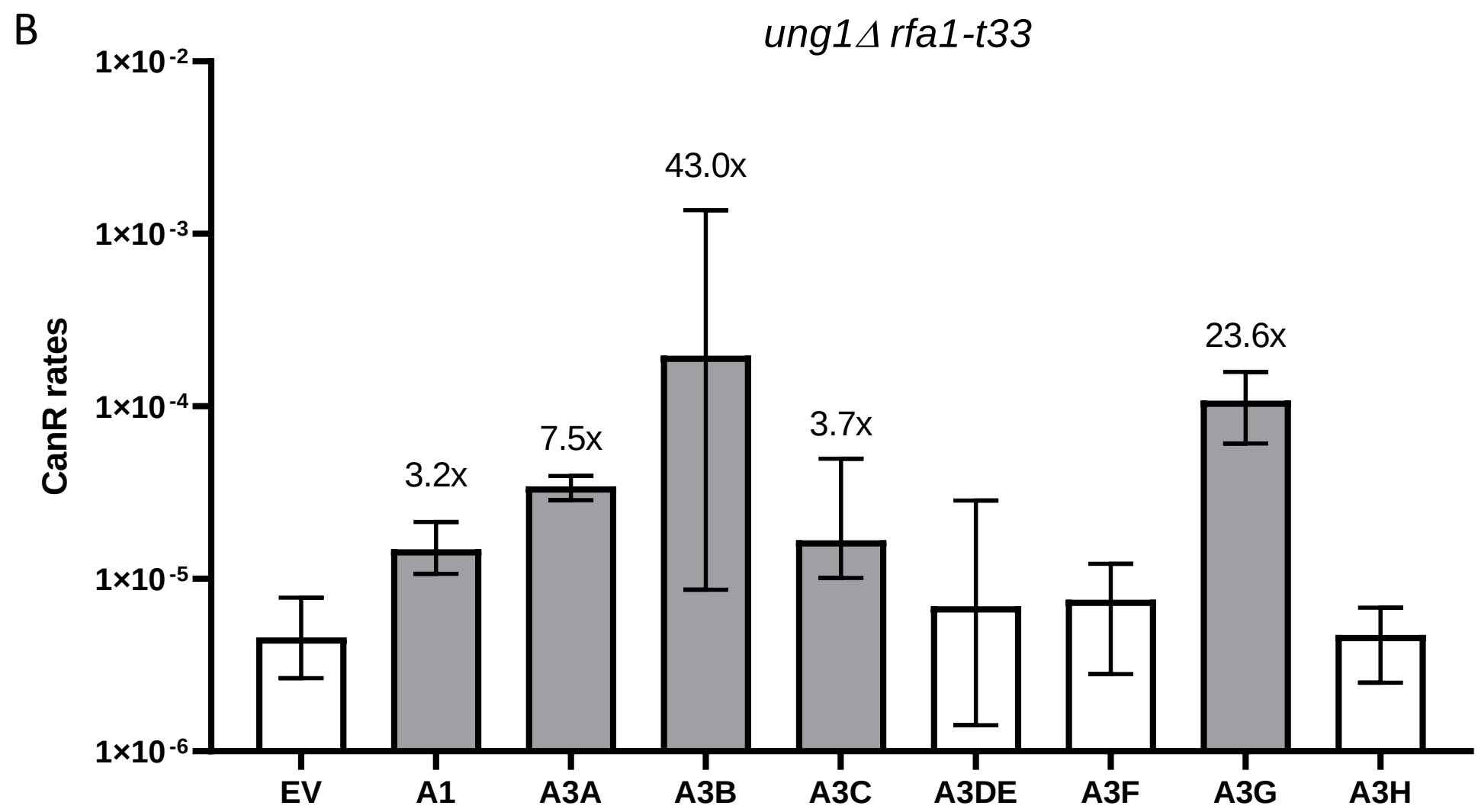
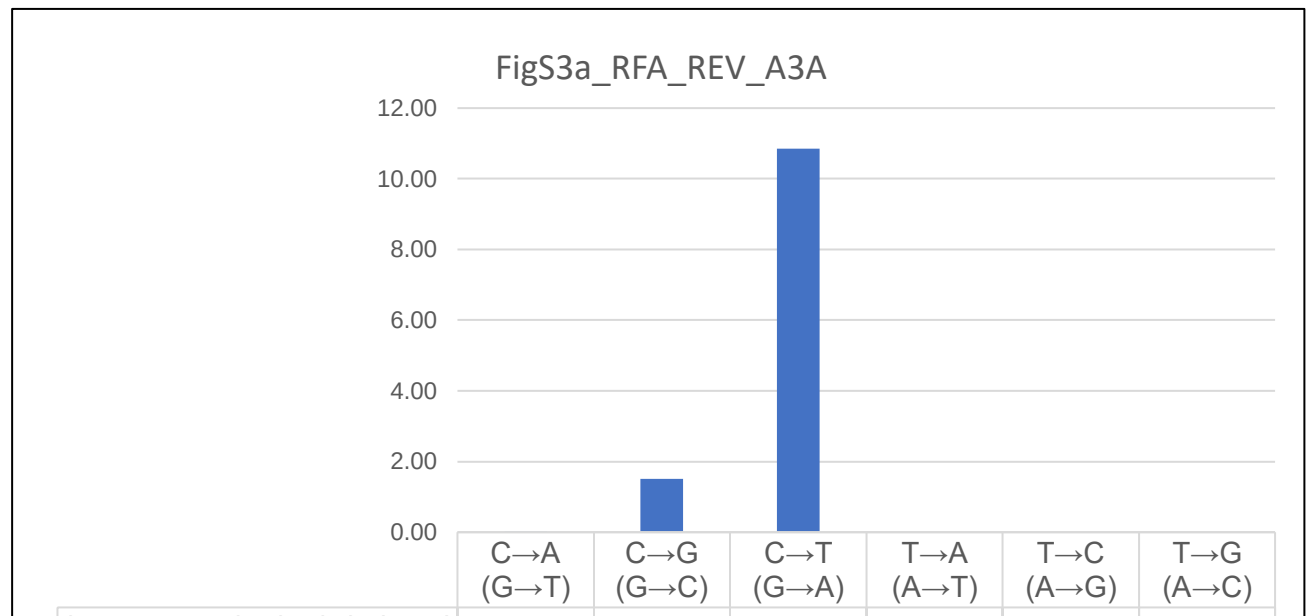
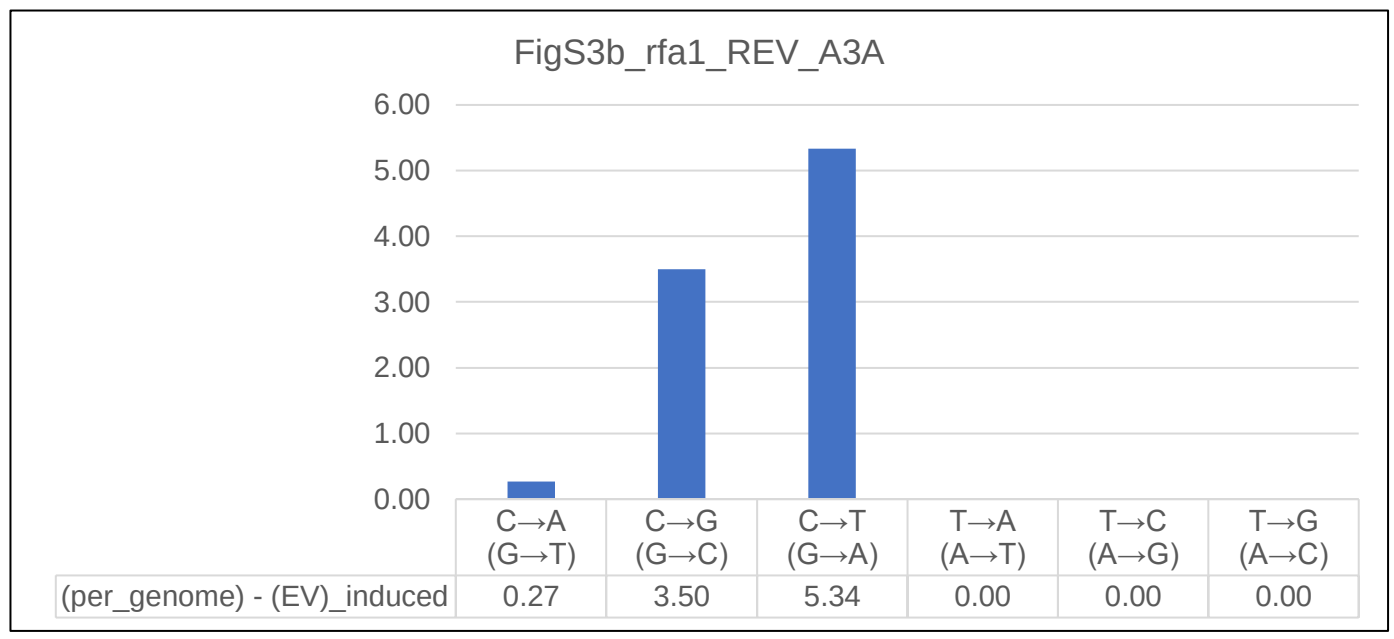

B
ung1Δ rfa1-t33
CanR rates
1×10-2
1×10-3
1×10-4
1×10-5
1×10-6
3.2x
7.5x
43.0x
3.7x
23.6x
EV
A1
A3A
A3B
A3C
A3DE
A3F
A3G
A3H

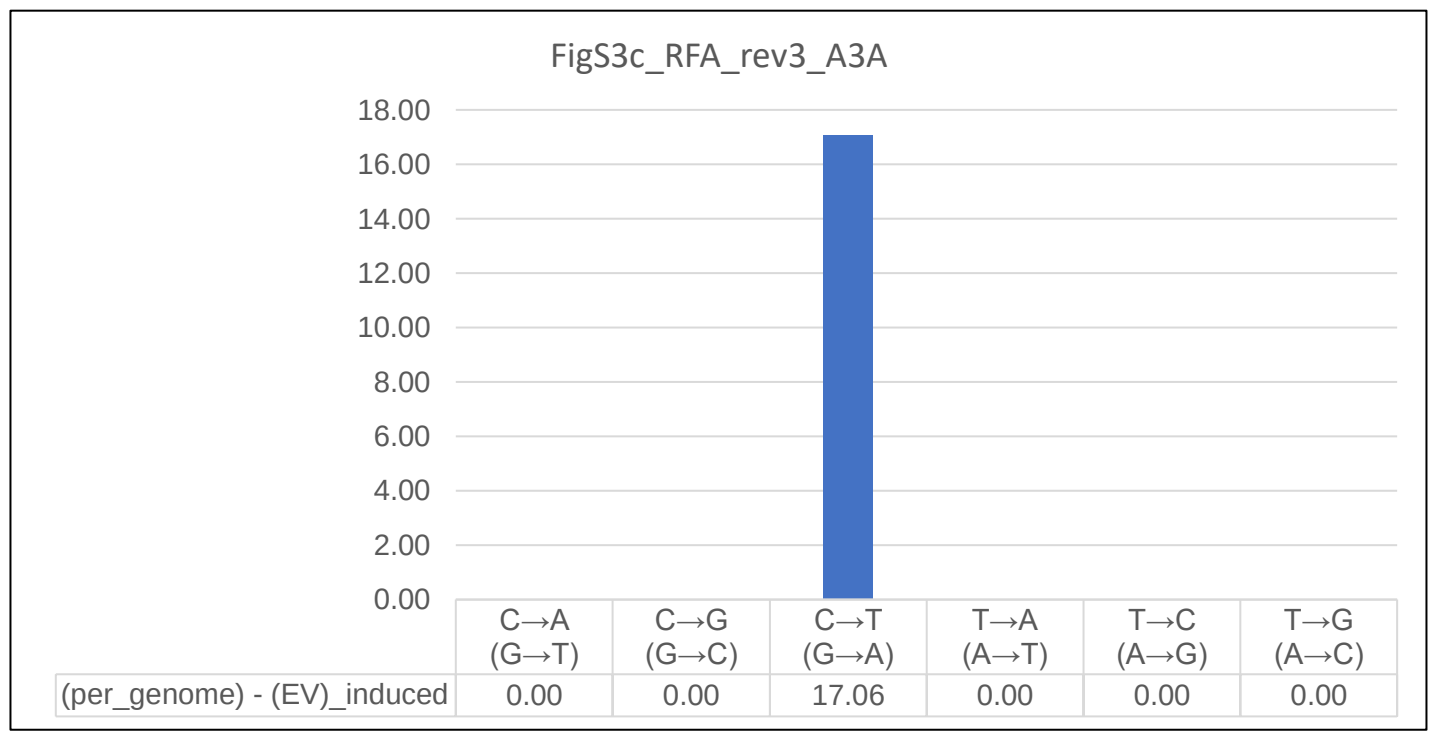
Figure S3



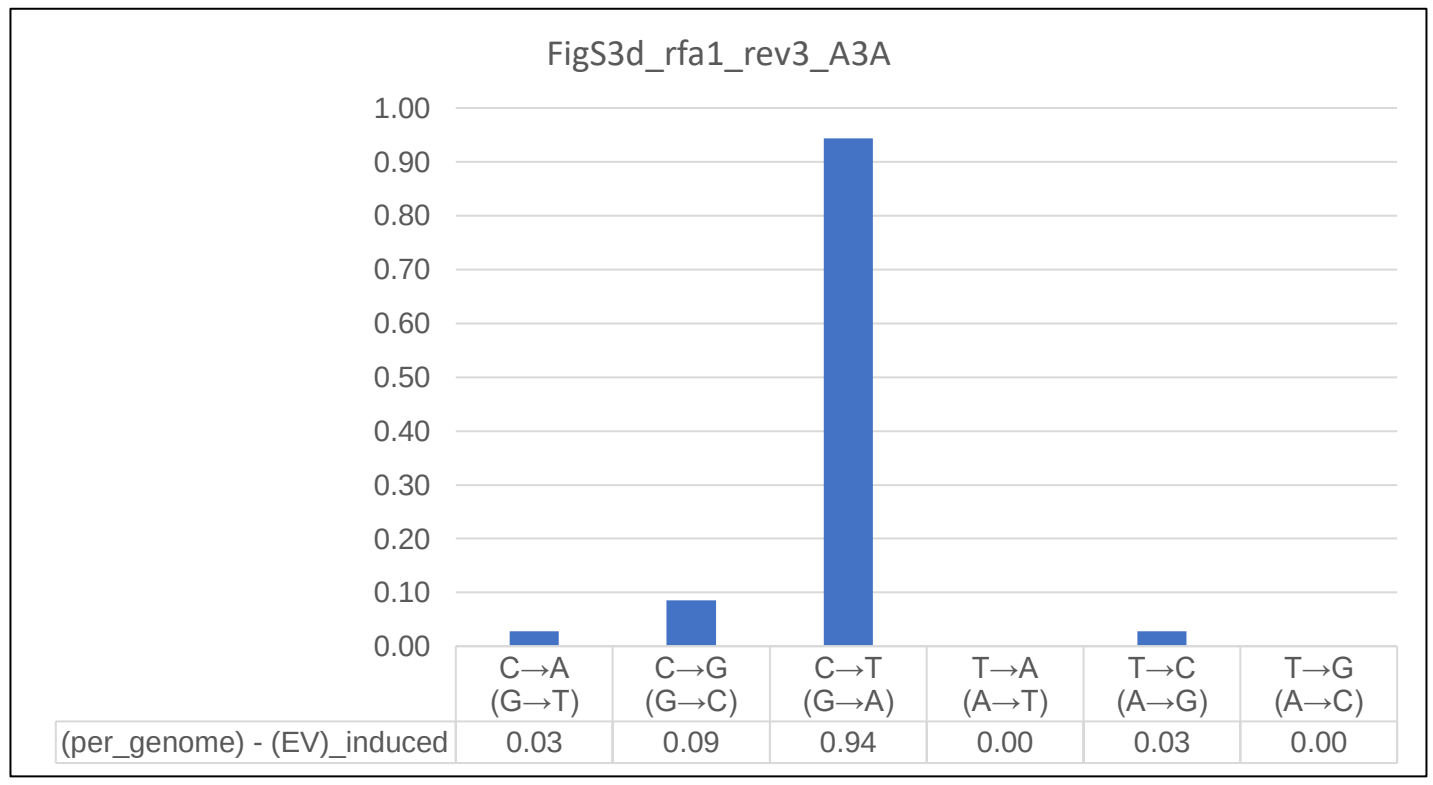
	C→A (G→T)	C→G (G→C)	C→T (G→A)	T→A (A→T)	T→C (A→G)	T→G (A→C)
(per_genome) - (EV)_induced	0.00	1.51	10.85	0.00	0.00	0.00



	C→A (G→T)	C→G (G→C)	C→T (G→A)	T→A (A→T)	T→C (A→G)	T→G (A→C)
(per_genome) - (EV)_induced	0.27	3.50	5.34	0.00	0.00	0.00



	C→A (G→T)	C→G (G→C)	C→T (G→A)	T→A (A→T)	T→C (A→G)	T→G (A→C)
(per_genome) - (EV)_induced	0.00	0.00	17.06	0.00	0.00	0.00



	C→A (G→T)	C→G (G→C)	C→T (G→A)	T→A (A→T)	T→C (A→G)	T→G (A→C)
(per_genome) - (EV)_induced	0.03	0.09	0.94	0.00	0.03	0.00