Complex Mutation Profiles in Mismatch Repair and Ribonucleotide Reductase Mutations Reveal Novel Repair Substrate Specificity of MutS Homolog (MSH) Complexes

Natalie A. Lamb[1], Jonathan Bard[1,2], Raphael Loll-Krippleber[3], Grant W. Brown[3], Jennifer A. Surtees[1,4*]

[1] Department of Biochemistry, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo (SUNY), Buffalo, NY, 14203

[2] University at Buffalo Genomics and Bioinformatics Core, Buffalo, NY, 14203

[3] Department of Biochemistry and Donnelly Centre, University of Toronto, 160 College Street, Toronto, ON, Canada, M5S 3E1

[4] Genetics, Genomics and Bioinformatics Graduate Program, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo (SUNY), Buffalo, NY, 14203

Abstract

Determining mutation signatures is standard for understanding the etiology of human tumors and informing cancer treatment. Multiple determinants of DNA replication fidelity prevent mutagenesis that leads to carcinogenesis, including the regulation of free deoxyribonucleoside triphosphate (dNTP) pools by ribonucleotide reductase (RNR) and repair of replication errors by the mismatch repair (MMR) system. We identified genetic interactions between *rnr1* alleles that elevate dNTP levels and MMR. We then utilized a targeted deep-sequencing approach to determine mutational signatures associated with MMR pathway defects. By combining *rnr1* and *msh* mutations to increase dNTP levels and alter the mutational load, we uncovered previously unreported specificities of Msh2-Msh3 and Msh2-Msh6. Msh2-Msh3 is uniquely able to direct repair of G/C single base deletions in GC runs, while Msh2-Msh6 specifically directs repair of substitutions at G/C dinucleotides. We also identified broader sequence contexts that influence variant profiles in different genetic backgrounds. Finally, we observed that the mutation profiles in double mutants were not necessarily an additive relationship of mutation profiles in single mutants. Our results have implications for interpreting mutation signatures from human tumors, particularly when MMR is defective.

Introduction

Cancer is a genetic disease caused by mutation accumulation; DNA replication is an important source of mutation. Replicative polymerases, Polε and Polδ, minimize errors via a highly selective nucleotide binding domain to prevent deoxynucleotide triphosphate (dNTP) misincorporation and an exonuclease domain to proofread and remove errors (1-4). Together, these functions lead to error rates on the order of 10^{-7} (3,4). Appropriate levels and ratios of the four dNTPs are also essential for maintaining high fidelity polymerase function. This has been demonstrated in yeast using allosteric mutations in *RNR1* that alter dNTP pools in different ways. *RNR1* encodes the large subunit of the enzyme that catalyzes the rate-limiting step in dNTP synthesis, ribonucleotide reductase (RNR). Even a modest 2-fold increase above normal levels, as seen in *rnr1D57N*, increased nucleotide misincorporation by DNA polymerases and elevated mutation rates (5,6). More pronounced and skewed elevations in dNTP pools are generated in *rnr1Y285F* and *rnr1Y285A*, which increase dCTP and dTTP 3-fold and 20-fold, respectively (7). These increases in dNTP pool levels further compromise replication fidelity (7-10). Cancer cells have increased proliferative nature and thus elevated dNTP pools may be necessary to support accelerated DNA replication (11,12), which may, in turn, increase mutagenesis, promote molecular evolution and provide a selective advantage to the tumor.

Nucleotide misincorporation and DNA slippage events in DNA replication result in DNA structures that are substrates for the mismatch repair (MMR) system, which functions as a spell-check, recognizing and directing the repair of errors in replication (13), thereby increasing the fidelity of replication by an additional 10-1,000 fold (2,3). Once recognized by a MutS homolog (Msh) complex, errors are targeted for excision. In most eukaryotes, two heterodimeric complexes bind errors at the replication fork: Msh2-Msh6 and Msh2-Msh3, which recognize a broad spectrum of mismatches and insertion/deletion loops (IDLs) with different specificities. The current model of post-replicative repair

posits that Msh2-Msh3 recognizes, binds and directs repair of IDLs up to 17 nucleotides long (14,15), while Msh2-Msh6 targets mismatches and IDLs of 1-2 nucleotides. Msh2-Msh3 also has affinity for some mismatches, especially C-C, A-A and (possibly) G-G (16,17).

MMR defects promote tumorigenesis. MMR is deficient in ~25% of sporadic cancers caused by increased rate of mutagenesis (18). Deficiencies in MMR genes *MSH2, MSH6, PMS2* and *MLH1* cause hereditary Lynch syndrome, which leads to a strong predisposition to cancers of the gastrointestinal tract, endometrial cancer and lymphomas, and are defined by microsatellite instability (19,20). MMR mutations are also implicated in breast and ovarian cancer (21,22). While Msh2-Msh3 has not been directly linked to Lynch syndrome, mutations in *MSH3* lead to cancer predisposition within and outside the gastrointestinal tract (23-28) as well as chemoresistant tumors (29-31). Loss of *MSH3* function leads to elevated microsatellite alterations of selected tetranucleotide repeats, which is distinct from microsatellite instability and has been associated with a number of different cancers, including up to ~60% of colorectal cancers (32). Therefore, while *MSH3* also plays a role in tumorigenesis, its role is distinct from that of Msh2-Msh6.

In mammalian cells, characterizing mutation signatures is a powerful tool to identify potential pathways that are implicated in tumorigenesis. However, this analysis rarely considers the likely presence of altered dNTP pools as a mitigating factor. Notably, combining defects in both MMR and dNTP pool synthesis leads to synergistic increases in mutation rates and uncharacterized mutagenic potential(6). The synergistic effect on mutation rate seen in *rnr1D57N msh2Δ* and *rnr1D57N msh6Δ* (6) and in *rnr1Y285A msh2Δ* strains, (8,10) indicates that mismatch repair normally corrects the errors generated by DNA polymerases when dNTP pools are elevated. We sought to determine how altered dNTP pools and loss of functional MMR interact genetically, at the cell and nucleotide levels, with an eye to developing a mechanistic understanding of mutation signatures observed in tumors. At the

same time, the altered mutation profiles generated by *rnr1* alleles (33) allowed us to evaluate the role of Msh2-Msh3 and Msh2-Msh6 in directing repair of typically rare replication errors.

In *Saccharomyces cerevisiae*, we performed targeted deep sequencing of the *CAN1* gene to build mutation profiles from first principles. We characterized single *msh∆* mutants and evaluated combinatorial effects on the mutation profiles when combined with *rnr1* alleles. Our key findings are: 1) a genetic interaction between *rnr1* and MMR deletions, 2) the identification of novel and specific DNA substrates for Msh2-Msh3- versus Msh2-Msh6-mediated MMR and 3) that mutation profiles of *rnr1 msh∆* double mutants were not necessarily additive of the single mutant profiles.

Results

Deletion of *MSH* genes alters mutation profiles in distinct ways

To define the mutation profiles associated with specific MMR deficiencies, we deleted *MSH2, MSH6* or *MSH3*, which encode the subunits of the MMR recognition MSH complexes (**Tables S1, S2**), and used a targeted deep sequencing approach to derive mutation profiles from first principles (33). We selected ~2,000 canavanine resistant (Can^R) colonies (selected samples) for each different genetic background, pooled them and extracted the pooled genomic DNA. These samples were subjected to paired-end deep sequencing of the *CAN1* gene. Selected samples exhibited an average variant frequency of 99.35%. We applied a permissive variant filter, based on the average frequency of individual variants in permissive samples (grown in the absence of selection) and the positions in which they occurred, to remove background noise resulting from PCR/sequencing errors (33) (**Fig. S1**).

Deletion of *MSH2* effectively eliminates all MMR activity, retaining all errors made by replicative polymerases (34). We also removed only Msh2-Msh6 (*msh6Δ*) or Msh2-Msh3 (*msh3Δ*) to learn more about substrate specificity and repair efficiency of each complex. In general, *msh2Δ, msh3Δ* and *msh6Δ* mutation profiles were consistent with previous studies performed in yeast (**Fig. 1A**) (6,10,15,16,35-37), but our work significantly increased the number of colonies sequenced and the sequencing coverage. Previous studies focused primarily on *msh2Δ;* studies of *msh3Δ* and *msh6Δ* analyzed fewer than 100 colonies via Sanger sequencing (6,15,16), preventing statistically significant conclusions about variant types. We found that deleting *MSH2* resulted in a relative decrease in single nucleotide variants (SNVs) and increased deletions and insertions compared to wildtype, similar to previous observations (35,37). A/T single base deletions dominated the mutation profile in *msh2Δ* (**Table S3**, **Fig. 1B, C, D)**. In contrast, the *msh6Δ* spectrum was dominated by SNVs, which represented >80% of the mutations (**Table S3**, **Fig. 1A, B)**, although the *types* of SNVs generated were similar for

msh2Δ and *msh6Δ* (**Table S3**, **Fig. 1B).** In *msh3Δ* cells there was a similar proportion of deletion events compared to *msh2Δ*, but the types of mutation varied. There was a marked increase in G/C -1 bp deletions and complex deletions and insertions (> 1 base pair) compared to wild-type, *msh2Δ* or *msh6Δ*, consistent with the preference of Msh2-3 for binding larger IDLs (38,39). Approximately 30% of the mutations that accumulated in *msh3Δ* were SNVs, but again the distribution was distinct, with increased CG>GC and TA>AT changes, despite the fact that these errors should be efficiently repaired by Msh2-Msh6 (40,41); TA>AT errors in *msh3Δ* were more frequent than observed in wildtype (**Table S3**, **Fig. 1A)**. There were fewer TA>CG changes in *msh3Δ* compared to *msh2Δ* and *msh6Δ* (**Table S3**, **Fig. 1A)**. These observations are consistent with Msh2-Msh3 also playing a role in correcting a specific subset of misincorporation events (16,36).

Rates of canavanine resistance increase in *rnr1 msh*Δ

In assessing the ability of Msh2-Msh3 and Msh2-Msh6 to direct repair of replication errors using *msh* alleles, we are limited to evaluating repair of those mutations that arise because of replication error. Due to the inherent proofreading activity of Polδ and Polε, this rate is low and certain mutations are very rare (42-44). However, *rnr1* alleles alter dNTP pools and compromise replication fidelity to generate distinct mutation profiles, including variants that are rare in wild-type backgrounds (8-10,33). To evaluate the specificity of MMR MSH complexes for a broader range of mutations and to assess the impact of combined genotypes on mutation profiles, we combined *msh* deletions with *rnr1* alleles. We previously characterized three *rnr1* alleles, *rnr1D57N*, *rnr1Y285F* and *rnr1Y285A*, using our targeted deep sequencing approach (33). The mutation profile, SNVs, in/dels, and sequence contexts for mutation in each was distinct from wild-type, and from each other (**Fig. 1**) (33). in addition to changes in the proportion and/or ratios of SNVs, especially increased CG>TA

mutations, all three *rnr1* alleles exhibited a shift in the relative distribution of single base deletions toward G/C deletions, which are typically rare events in wild-type backgrounds. The frequency of single base G/C deletions was particularly elevated in *rnr1Y285A*. Notably, dNTP pools are likely elevated in cancer cells (11,45), including those caused by defects in MMR. Therefore, the effect of the combination of *msh∆* and altered dNTPs on replication fidelity could result in unique mutation signatures observed in tumors.

To determine whether the *rnr1 msh* double mutants displayed the anticipated increases in mutagenesis, we measured the forward mutation rates at *CAN1*. The relative mutation rates for *msh* deletions were consistent with previous observations (6,8,35,46), with *msh2∆* exhibiting the highest mutation rate, *msh6∆* exhibiting a lower but still increased mutation rate and *msh3∆* exhibiting a mild mutator effect in this assay (**Table 1**). The *rnr1* mutation rates increased with the increasing dNTP levels displayed by each allele as expected (7), with *rnr1Y285A* showing a 20-fold increase. When *rnr1D57N* and *rnr1Y285F* were combined with *msh* deletions, the mutation rates exhibited the same hierarchy (*msh2∆* > *msh6∆* > *msh3∆*), but with higher rates than either of the relevant single mutants. In most cases the mutation rate of the double mutant approximated the product of the single mutant mutation rates. The *rnr1Y285F msh6Δ* strain had a particularly large increase, 85-fold above wild-type, compared to the expected 19-fold given the mutation rates of the respective single mutants. Notably, *rnr1Y285A msh3∆* exhibited a strong, synergistic mutator phenotype, indicating that the mutations generated in the presence of *rnr1Y285A* are substrates for Msh2-Msh3-mediated MMR. It is possible that the mutation rates for *rnr1Y285A msh2∆* and *rnr1Y285A msh6∆* are underestimates, due to the increased cell death observed with these strains.

Previous work characterized *rnr1Y285F* and *rnr1Y285A* (+/- *MSH2)* in backgrounds that carried a silenced *pGAL-RNR1* gene (i.e. grown in the absence of galactose) (7,10). We previously demonstrated

that the mutation profiles of these strains were comparable to *rnr1Y285F* and *rnr1Y285A* without the *pGAL-RNR1* construct (33). They are included here, with and without *MSH* genes to provide a comparison with that previous work. In general, *pGAL-RNR1* did not impact the observed mutation profile (**Fig. S2, box2; Fig. S3**) and were combined with the equivalent strains lacking *pGAL-RNR1* for some of our analyses.

Synthetic genetic interaction of *rnr1Y285A* with MMR factors

The combined effect of *rnr1* alleles and *msh* deletions on mutation rates suggested that together these genes contribute to increased replication fidelity, and that therefore they might show synergistic effects on cell fitness. To test this possibility, we performed synthetic genetic array (SGA) analysis using each of the three *rnr1* alleles (**Fig. 2** and **Tables S4-S10**). The *rnr1Y285A* query had the greatest number of genetic interactions, consistent with the Y285A mutation having the greatest effect on dNTP levels. The three *rnr1* alleles showed surprisingly little overlap in their genetic interactions (**Fig. 2A**), supporting the idea that different dNTP levels and pool balances stress cells in different ways. Both the *rnr1Y285A* screen and the *rnr1D57N* screen showed enrichment for the GO term 'DNA replication' (**Fig. 2B**) in addition to displaying unique enrichments for 'maintenance of DNA repeat elements' and 'DNA repair' (*rnr1Y285A*) and 'lagging strand replication' (*rnr1D57N*). Despite having interactions with several DNA replication genes, the *rnr1Y285F* screen did not show any statistically supported GO term enrichment. To further assess the functional properties of each *rnr1* allele genetic interactions, we applied spatial analysis of functional enrichment (SAFE) (47) to determine if any regions of the functional genetic interaction similarity yeast cell map (48) are over-represented for the negative genetic interaction gene sets (**Fig. 2C**). We found a statistically supported over-representation of the negative interacting genes in the DNA replication and repair neighborhood of the genetic

interaction cell map for all three *rnr1* alleles, indicating that dNTP pool alterations impinge most dramatically on the DNA replication and DNA repair capacity of the cell. As with GO term enrichment, differences among the three *rnr1* alleles were also apparent in the SAFE analysis, with *rnr1Y285A* and *rnr1D57N* interactors being over-represented in the chromatin organization neighborhood, and *rnr1Y285F* and *rnr1Y285A* interactors showing more over-representation in the mitosis neighborhood than did *rnr1D57N* interactors.

Most notably, we found that almost the entire MMR pathway (*MSH2*, *MSH3*, *MSH6*, *MLH1*, *PMS1* and *EXO1*) was identified specifically in the *rnr1Y285A* screen and the 'mismatch repair' GO term was very strongly and specifically enriched in the *rnr1Y285A* screen (**Fig. 2C**), indicating a strong requirement for the mismatch repair pathway when dNTP pools are highly unbalanced. We validated the specificity of the genetic interaction between MMR and *rnr1Y285A* by performing tetrad analysis of *msh6*Δ crosses with each of the three *rnr1* alleles (**Fig. 2D**). Fitness defects in double mutant colonies were only evident in the *rnr1Y285A* cross. In tetrad analysis, *msh6*Δ, and particularly *msh2*Δ showed stronger fitness defects when combined with *rnr1Y285A* than did *msh3*Δ (**Fig. 2E**), consistent with the reduced viability observed in mutation rate experiments (**Table 1**). Our genetic interaction data indicate that all three *rnr1* alleles interface with DNA replication and repair pathways, and that *rnr1Y285A* might be expected to have a particularly dramatic effect in MMR deficient cells.

Mutation profiles are systematically genotype-specific

To refine the mutation profiles of the *msh∆* and *rnr1 msh∆* mutants, we characterized mutation events in two ways (**Table S3**), as previously described (33). First, we determined the number of a specific variant type, i.e., the number of C>A changes, at different positions along *can1* ("unique counts") (Counts in **Table S3**). Second, we calculated the frequency at which each of these unique

variants occurred, i.e., the combined frequency of all C>A changes at any position along *can1* ("sum of frequencies") (Freq. in **Table S3**). These analyses allowed us to determine whether different *types* of mutations occurred at specific locations in a genotype-dependent manner, independent of frequency, and whether variant frequencies were altered in a significant way by genotype (Counts/Freq. in **Table S3**). A decreased number for "unique counts" combined with unchanged or increased "sum of frequencies" would indicate that variant type is more localized, possibly indicating a mutational hotspot. For instance, *msh6Δ* exhibited the highest proportion of unique events contributing to the mutation spectrum (Counts/Freq. = 1.62; **Table S3**). In contrast, *rnr1Y285F msh3Δ* and *rnr1Y285A msh3Δ* exhibited the lowest proportion of unique variants; the mutation spectra were instead dominated by G/C single base deletions, which occur at high frequencies (**Table S3**).

We used Spearman rank correlation and cluster analysis to determine the quantitative relationship between mutation profile and genotype (see Materials and Methods). All unique variants for all genotypes were assessed in parallel, based on both the presence and frequency of unique variants, as described above. In general, biological replicates of the same genotype clustered because their mutational profiles were highly correlated (**Fig. S2, Table S11**). Therefore, we combined variants from all biological replicates within a genotype for the remainder of our analyses. Hierarchical cluster analysis using Spearman rank correlations based on the profile of unique variants between genotypes was consistent with genotype-specific mutation profiles (**Fig. 3A, Fig. S2)**, as was principal component analysis (PCA) based on unique variants (**Fig. 3B**). Combined, these results indicated that it is possible to distinguish among genotypes based on unique variant profiles observed from *can1* deep sequencing. Below, we parse these trends to develop genotype-specific mutation profiles.

***rnr1* alleles combined with MMR deletion result in unique mutation profiles**

When *msh2Δ* was combined with each *rnr1* allele, distinct mutation spectra (i.e. the proportion of each type of variant) were observed, revealing mechanistic insights into MMR and replication fidelity (**Table S**, **Fig. 1A**). Th3e *rnr1D57N*, *rnr1Y285F* and *rnr1Y285A* profiles exhibited an increase in G/C-1 deletions relative to A/T-1 deletions compared to wildtype. The opposite trend was observed in *msh2Δ* and in the *rnr1 mshΔ* backgrounds. This indicated that, while elevated/skewed dNTP increases the proportion of G/C-1 variants, A/T-1 deletions are preferentially repaired in the presence of MMR. There was also a significant increase in A/T+1 insertions in *rnr1D57N msh2Δ* relative to either single mutant. Finally, while the *rnr1Y285F* and *rnr1Y285A* SNV profiles were biased toward CG>TA and CG>AT SNVs, *rnr1Y285F msh2Δ* and *rnr1Y285A msh2Δ* were almost completely dominated by these variants, with proportions that differed from either single mutant (**Table S3**, **Fig. 1)**. These SNV profiles indicated that: 1) DNA polymerases primarily generated these errors when dCTP and dTTP were modestly skewed and elevated(*rnr1Y285F*) and 2) these elevated frequencies began to saturate MMR activity and/or these errors are inefficiently repaired by MMR.

Deleting *MSH6* in combination with *rnr1* alleles, so that only Msh2-Msh3-mediated MMR was present, also resulted in unique shifts in mutagenesis across *can1* (**Table S3 Fig. 1A)**. The effect was most dramatic in *rnr1Y285F msh6Δ* with a profile that was almost completely dominated by CG>TA and CG>AT SNV errors (**Table S3**, **Fig. 1B, C)**, indicating that Msh2-Msh3 is not efficient in correcting these mismatches. The proportion of CG>AT transversions was even higher in *rnr1Y285A msh6Δ*, although most of the variant classes were still observed. In both *rnr1Y285F msh6Δ* and *rnr1Y285A msh6Δ* there was almost a complete loss of the G/C -1 bp deletions that were observed at an increased frequency in *rnr1Y285F* and *rnr1Y285A* (**Table S3**, **Fig. 1C,D**), consistent with efficient repair of G/C -1 bp slippage events by Msh2-Msh3-mediated MMR. Strikingly, the opposite was observed in *msh3Δ* genotypes. In both *rnr1Y285F msh3Δ* and *rnr1Y285A msh3Δ*, there was a dramatic increase in G/C -1 bp deletions

compared to either single mutant, almost to the exclusion of other variants, indicating that Msh2-Msh6 was unable to correct this type of error.

Genotype-specific susceptibility to mutation within *CAN1*

In addition to unique variant spectra, the variant positions within *CAN1* were genotype-specific. (**Table S12-S17**). All variants were more likely to occur at or adjacent to repetitive sequences, although the specific repetitive sequences varied by genotype (49). We identified and defined susceptible positions as those in which the same variant was observed at the same position in >50% of biological replicates. Increased susceptibility sites also had a variant frequency above wild-type. Highly susceptible positions were defined as susceptible positions with a variant frequency >2-fold above wildtype. There were more susceptible positions in *msh2Δ* (49 positions) and *msh6Δ* (96 positions) than in *msh3Δ* (18 positions) and mean frequencies at these positions were higher in *msh2Δ* and *msh6Δ* than in *msh3Δ*, consistent with more stochastic, diffuse mutations in the absence of *MSH3*. The highly susceptible *CAN1* positions that we observed in *msh2Δ* overlapped previously observed "hotspots" (9,10); we noted additional susceptible positions, particularly in repetitive sequences (see below). We also noted novel increased susceptibility sites in *msh6Δ* and *msh3Δ*. The *msh6Δ* exhibited 13 highly susceptible positions, all SNVs, which were not observed in wildtype and occurred almost exclusively in repetitive sequence contexts, including at dinucleotides and in runs 3bp or greater in length. In *msh2Δ*, 11 highly susceptible positions were observed, mostly deletions within repetitive runs, with few positions overlapping with *msh6Δ*. We previously identified novel susceptible positions in *rnr1D57N*, *rnr1Y285F* and *rnr1Y285A* (33),many of which were associated with repetitive DNA sequences, particularly insertions and deletions.

When *msh* deletions were combined with *rnr1* alleles, the double mutants exhibited distinct profiles of susceptible positions in *CAN1*, typically differing from the individual single mutants in the number of replicates affected and/or the frequency of mutation. The susceptible positions in *rnr1Y285A msh2Δ* identified in our study largely overlapped with previously identified "hotspots" (10,50). Our study also revealed new susceptible positions. One noteworthy position is 32,940, which occurs in A-rich sequence (CCAA**G**AAAA) and is susceptible to G>T mutation in *msh2Δ rnr1Y285A* and *rnr1Y285A-pGAL*. G>A SNVs also occurred in a variety of genotypes at this position, indicating decreased replication fidelity in this context. The variant frequency at this position increased synergistically in *rnr1Y285A-pGAL msh2Δ* and *rnr1Y285A msh2Δ* double mutants, occurring at a frequency at least 10-fold greater than most single mutants. The majority of susceptible positions in *rnr1Y285A msh2Δ* hotspots are also susceptible in *rnr1Y285A* but not *msh2Δ*, indicating that highly skewed/elevated dCTP and dTTP levels promoted specific errors to a level where MMR was invoked, and approached saturation of the MMR system. By contrast, *rnr1Y285A msh6Δ* and *rnr1Y285A msh3Δ* tended to exhibit susceptible positions that were distinct from either single mutant.

Insertions and deletions occur in distinct sequence contexts depending on genotype

In general, susceptible *CAN1* positions tended to be in or near repetitive sequences and it is well established that insertion and deletion events occur more frequently in repetitive runs (4,51); our study found similar properties. Within *CAN1*, there are only short homopolymer runs and A/T repeats are more abundant than G/C repeats: 266 AA/TT dinucleotides versus 160 GG/CC dinucleotides; 98 mononucleotide runs of 2 or more A/T bases versus only 32 G/C runs. These ratios are consistent with our observation that more A/T insertion and deletion events occurred in A/T repeats than in G/C repeats in wildtype (**Fig. 1C & 1D**). 96.1% of all A/T -1 bp deletions occurred in a repetitive run of 3 bp

or greater. 89.6% of G/C deletions occurred in a repetitive run of 2 or more nucleotides. Notably, some G/C deletions occurred adjacent to repetitive runs in specific *rnr1* backgrounds. For example, a GΔ at position 31971 occurred in 4 out of 7 *rnr1D57N* biological replicates, between an A dinucleotide and C dinucleotide (ATAA**G**CCAA). CΔ deletions at positions 32398 (GAAA**C**GTAG) and 32630 (CCTT**C**GTTTT) were specific to r*nr1Y285A*. A GΔ at position 33018 was specific to *msh3Δ*; it is not flanked by a repetitive run but dinucleotides are nearby upstream and downstream (GGAT**G**TAAC).

Complex insertions and deletions (i.e. involving more than one nucleotide) were rare, but occurred at increased frequency in *msh3Δ*. The majority of these complex events, especially insertions, were observed in a single biological replicate. The complex insertion of CT in the repetitive run stretching from positions 33206- 33215 (CTTAAG**CT**CTCTC) is noteworthy. It was observed almost exclusively in *msh2Δ* genotypes, and more frequently when paired with *rnr1* alleles. Increased CT insertion in *rnr1 msh2*Δ genotypes indicates that positions 33206-33215 were particularly susceptible to mutation when dNTPs were elevated, even by a small amount as is the case in *rnr1D57N*. However, the CT insertion was very efficiently repaired by MMR, via either Msh2-Msh3 or Msh2-Msh6 directed repair as it was not observed in either *msh3Δ* or *msh6Δ*.

Unique mutation signatures revealed by analysis of SNV trinucleotide context

To gain mechanistic information about sequence context we analyzed SNVs and the nucleotides on the 5' and 3' sides of the variant. We determined the average number of times an SNV was observed in a particular triplet context per genotype, normalized by the number of times the triplet context occurs in *CAN1* (**Fig. 4)**. Notably, in all genotypes C→T changes (red bars, **Fig. 4**), particularly in G**C**C and G**C**G sequence contexts, dominated. The proportion of G**C**C and G**C**G changes increased with altered dNTPs, most dramatically in *rnr1Y285F* and *rnr1Y285A* samples (note change in y-axis), which

were highly correlated (r_s = 0.7284; **Table S18**). The G**C**C and G**C**G sequence contexts are mutated in *rnr1Y285A*, as detected by whole genome sequencing (9) possibly as a result of increased extension after misincorporation of dTTP, which is in excess, superseding proofreading (52).

The wildtype SNV trinucleotide pattern was highly correlated with that of *msh3Δ* (r_s = 0.7351), *rnr1D57N msh3Δ* (r_s = 0.6752) and *rnr1D57N* (r_s = 0.6682). We noted that no C>T SNVs occurred in G**C**T context in these four genotypes, while they were observed in all *rnr1Y285A* and *rnr1Y285F* double mutant backgrounds, still at relatively low frequencies. This example shows an error that was specific to skewed increases in dCTP and dTTP and was efficiently repaired by either Msh2-Msh3 or Msh2-Msh6.

We also observed a decrease in the number of unique SNVs that occurred in trinucleotide context when *rnr1Y285F* or *rnr1Y285A* alleles were paired with *msh2Δ, msh3Δ,* or *msh6Δ*. Calculation of Spearman correlation coefficients of each SNV in unique trinucleotide context revealed that these genotypes are also highly correlated to one another, with *rnr1Y285F/A msh6Δ* or *msh2Δ* double mutants showing the highest correlation values (**Table S18**). For example, the spectrum of SNVs in trinucleotide context of *rnr1Y285A msh2Δ* was most highly correlated to *rnr1Y285F-pGAL RNR1 msh2Δ* (0.8616)*, rnr1Y285F msh6Δ* (0.8595), *rnr1Y285F msh2Δ* (0.8458)*, rnr1Y285F pGAL-RNR1 msh6Δ* (0.8197), *rnr1D57N msh6Δ (0.7686), rnr1Y285A msh6Δ* (0.7662) and *rnr1Y285A pGAL msh2Δ* (0.7640)*,* descending in value. We observed a complete loss of C>G SNVs in all trinucleotide contexts in *rnr1Y285A msh3Δ*, *rnr1Y285A msh6Δ*, and *rnr1Y285F msh6Δ* backgrounds, as noted above (**Fig. 1, Table S3)**. The C>G variant rarely occurred in *msh2Δ* backgrounds, indicating that the replicative polymerases rarely generate these errors. It is noteworthy that double mutant profiles often showed distinct changes in the SNV signatures, compared to either single mutant. For example, *rnr1Y285A* sustained several C>G mutations in multiple trinucleotide contexts, which were also observed in

rnr1Y285A msh2Δ, but these were completely absent in *rnr1Y285A msh3Δ* and *rnr1Y285A msh6Δ*. Thus, there were combinatorial effects on the mutation spectra of specific MMR deletions in the presence of *rnr1* alleles. Different *rnr1* alleles paired with different MMR deletions result in distinct and unique mutational fingerprints.

Novel error substrate specificities revealed for Msh2-Msh3 and Msh2-Msh6 complexes

The depth of sequence within *can1* in single and double mutants allowed us to evaluate combinatorial effects of *rnr1* and *msh* alleles on mutation spectra. Using all single and double mutant variant data, we performed hierarchical cluster analysis of all unique errors to identify larger sequence contexts driving distinct mutational signatures (**Figs. 5-7)**. Genotypes were clustered based on the types of variants that were differentially enriched as a function of genotype, using Pearson correlation (Spearman correlation analysis yielded similar results) (**Fig. 5**). It is worth noting that overall, clusters were similar to those observed in **Fig. 3**, based on variant frequencies, which suggests that unique variants were the main drivers of genotype-specific mutation profiles. We identified differentially enriched variants and performed motif analysis (12 base window) to determine whether broader sequence context influenced the occurrence of these variants, indicating the same biological mechanism. We identified several motifs that were positively or negatively enriched in different genetic backgrounds. MMR status appeared to be the primary driver for enrichment, with G/C-1 variants positively enriched in *msh3Δ* genotypes and A/T-1 variants and several SNVs positively enriched in *msh2Δ* and *msh6Δ* genotypes (**Fig. 5**).

Some of the clusters with the most pronounced differential enrichment between genotypes were SNVs that occurred in CC dinucleotide context (**Fig. 5**). As observed previously (9,33), CC/GG dinucleotides bordered by G on either the 5' or the 3' side were frequently mutated in the presence of

rnr1 alleles, particularly *rnr1Y285F* and *rnr1D57N,* but not with *msh* alleles (**Fig. 6)**. MMR status, (especially *msh3Δ* or *msh6Δ*) in combination with these *rnr1* alleles appeared to reverse this bias with negative enrichment. The majority of SNVs that occurred in other CC dinucleotide contexts were positively enriched in *msh2Δ* and *msh6Δ* samples, but negatively enriched in *msh3Δ* (**Fig. 6 II, III, & IV)**, indicating that Msh2-Msh6 was uniquely required to recognize and direct repair of misincorporation events in the CC dinucleotide context.

Motif enrichment analysis also revealed synergistic increases in G/C-1 variants within specific contexts when *rnr1Y285F/*A and msh3D alleles were combined (**Fig. 7**). We previously found that *rnr1Y285F* and *rnr1Y285A* single mutants showed an increase in G/C single base deletions (33). Here we found that *msh3Δ* does as well (**Fig. 1; Table S3**). Individual positions that sustained G/C errors along *CAN1* were often *msh3Δ*-specific (**Table S3**) or *rnr1Y285F/A*-specific (33). Notably, the frequency of G/C-1 variants increased synergistically in the *rnr1Y285F msh3Δ* and *rnr1Y285A msh3Δ* double mutants (**Fig. 7).** In contrast, G/C-1 variant frequency was neutral or negatively enriched in all other genotypes, consistent with apparent specificity of Msh2-Msh3 for directing repair of G/C single base deletions (**Fig. 1, Table S3**). Loss of Msh2-Msh3 resulted in increased G/C single base deletions in homopolymer runs bordered by G/C rich sequence on the 5' side of the run (**Fig. 7A-C**). This error in this context occurred rarely in *rnr1Y285F/A* alone. There was a significant increase in G/C-1 mutations in *rnr1Y285F/A msh3Δ* mutants. G/C runs bordered by A/T nucleotides were more prone to mutagenesis when dCTP and dTTP were increased than in the absence of *MSH3* (**Fig. 7D-F)**. The frequency of these variants directly bordered by A/T increased synergistically when *MSH3* was deleted, but not when *MSH6* was deleted, indicating Msh2-Msh3 has specificity in directing repair of G/C -1 deletions in repetitive G/C context.

Discussion

Utilizing a *CAN1* selection-based deep sequencing approach (33), we characterized mutation spectra in *msh∆* and *rnr1 msh∆* double mutant genotypes. The sequencing depth afforded by our approach allowed us to expand our understanding of mismatch repair substrate recognition as well as the combined effects of MMR defects and altered dNTP pools on replication fidelity. By using *rnr1* backgrounds that alter the *type* and *frequency* of mutations sustained, we revealed previously unrecognized specificities for the MMR initiation complexes, Msh2-Msh3 and Msh2-Msh6. The combinatorial effects that we find highlight the importance of studying mutation signatures in different genetic contexts.

Different mechanisms of mutagenesis result from distinct elevations in dNTP levels

In *rnr1D57N msh2∆* the mutation rate increased 74-fold above wildtype and 3-fold above *msh2∆*, yet the mutation spectrum of *rnr1D57N msh2∆* is closely related to *msh2∆*, with the exception of an increase in A/T+1 insertions in the double mutant (**Fig. 1**). The same is true of *rnr1D57N msh3∆* and *rnr1D57N msh6∆;* their mutation spectra are most closely related to those of *msh3∆* and *msh6∆*, respectively, despite high increases in mutation rates (**Fig. 1, Tables 1 , S3**). Therefore, the elevated dNTP pools in *rnr1D57N*, which resulted in a mutation spectrum similar to wildtype (33), with the notable exception of G/C-1 deletions, did not substantially drive the type of mutation generated. The low frequency variants that accumulate in *rnr1D57N* were effectively repaired by MMR, in general, and even the absence of MMR did not result in an overt fitness defect in *rnr1D57N* (**Fig. 1, Table S7, S12**). We conclude that the balanced dNTP increases in *rnr1D57N* alter mutagenesis without a remarkable change in mutation spectrum.

The *rnr1Y285F* allele has a modest effect on mutation rate, yet yielded a distinct mutation spectrum especially when paired with MMR deletions. In fact, the *rnr1Y285F* spectrum closely resembles that of *rnr1Y285A*, despite *rnr1Y285A* having more significantly skewed dNTP pools, a higher mutation rate (**Table 1**) and a fitness defect (**Fig. 2**). We conclude that even modest skewed increases in dNTPs (*rnr1Y285F*) result in distinct error accumulation, likely due to both a decrease in selectivity of the replicative polymerases, and to an increase in efficient mismatch extension at the expense of proofreading (8,9). **Fig. 8A** illustrates an example of two different positions in the *rnr1Y285A* background that are predicted to be mutated at increased frequencies via this mechanism.

The increase in G/C deletions in both *rnr1Y285F* and *rnr1Y285A* genotypes can be explained by limiting levels of dGTP (**Fig. 8B)**. dGTP levels are limiting in both yeast and mammalian cells and notably led to distinct patterns of mutagenesis when the concentration decreases further, relative to the increase in the other 3 nucleotides, in *rnr1Y285F* and *rnr1Y285A* and even in *rnr1D57N*. Notably, *rnr1* alleles that cause increases in dGTP are also found to be extremely mutagenic (53), highlighting the importance of maintaining the proper level and relative abundance of dGTP.

New insights into MMR specificity

By altering dNTP pools, we altered the frequency and types of replication errors generated by DNA poylmerases, revealing new substrate specificities for Msh2-Msh3 and Msh2-Msh6. While G/C SNVs are common in wild-type backgrounds, G/C-1 deletions are relatively rare, making it difficult to determine the relative efficiencies of MSH complexes in repairing this type of error. However, elevating the dNTP pools to any extent increased the proportion of G/C-1 deletions, allowing us to assess MMR efficacy in their repair. In particular, the meta-analysis of our dataset showed that the majority of single base deletions in *rnr1Y285F*, *rnr1Y285A*, and *msh3Δ* occurred in G/C rich contexts

(**Fig. 7, 8B**). The double mutants *rnr1Y285F msh3Δ* and *rnr1Y285A msh3Δ* exhibited mutation profiles that were completely dominated by G/C-1 deletions. Therefore Msh2-Msh6 was not able to compensate for the loss of Msh2-Msh3 for repair of G/C-1 deletions within G/C-rich sequence contexts, despite the fact that both complexes have been implicated in directing repair of single base deletions (54). This suggests an important role for Msh2-Msh3 in replication fidelity within G/C-rich genomic regions.

It has been previously shown that there are few strand specific effects from altering dNTP levels(10), but numerous studies have indicated that MMR is more efficient on the lagging strand (55,56). It is also known that polymerase δ mutants deficient in proofreading (*pol3-L612M*) accumulate significantly more G/C deletions (4). This leads us to hypothesize that Msh2-Msh3 may have greater specificity for lagging strand DNA repair. Our targeted sequencing approach could be applied on strains with *CAN1* in the reverse orientation to further explore leading versus lagging strand DNA repair specificity of Msh2-Msh3 and Msh2-Msh6.

Similarly, SNVs at GG/CC dinucleotides were more prevalent in *rnr1D57N* and *rnr1Y285F* backgrounds than in wild-type (**Fig. 6, 7A**). We demonstrated that mutations in these G/C-rich patches were enriched in *msh2Δ* and *msh6Δ* backgrounds, but depleted in *msh3Δ*, indicating that Msh2-Msh6 directs more efficient repair of mutations in GG/CC dinucleotides than Msh2-Msh3. GG dinucleotides are mutated as a signature mutation in colorectal cancer (57); cancers which are defined by defects in MMR. Mutations at CC dinucleotides are also found in human lung cancer (58,59) and could be generated due to defects in MMR and elevations in dNTP levels in combination. Notably, RNR subunit overexpression and deletion of *MSH6* in combination caused a synergistic increase in lung carcinogenesis in a mouse model(6).

Notably, pairing elevations in dNTP levels with MMR deletions led to increased mutation rates and distinct mutation spectra. However, when measuring the rate of canavanine resistance in *rnr1Y285A msh2Δ* and *rnr1Y285A msh6Δ* backgrounds, we observed substantially lower colony numbers under permissive conditions, indicating reduced fitness even under conditions when all cells should be able to grow. Reduced fitness of *rnr1Y285A msh2Δ* and *rnr1Y285A msh6Δ* was also seen in SGA analysis and in tetrad dissections, consistent with the phenomenon of error extinction where the threshold of mutation rate that allows wild type cellular fitness is surpassed (60-62). We expect that the fitness defects we observe in the absence of selection would also reduce the number of cells that are able to grow under the selective conditions of mutation rate experiments, resulting in mutation rates that are underestimates (**Table 1**).

Application to mutation signatures in human cancer

The *msh2Δ, msh3Δ* and *msh6Δ* mutation spectra all had features of MMR deficient human tumor samples (63). Notably, the *msh6Δ* spectrum in our study closely resembles that of *Msh6*$^{-/-}$ in a HAP1 cell line (64). The percentage of substitutions and IDLs in the *msh6*Δ mutation spectrum is consistent with what is seen in a *Msh6*$^{-/-}$ cell line. While T>C variants did not dominate the yeast *msh6Δ* mutation signature as it did in the *Msh6*$^{-/-}$ cell line, the overall frequency and proportion of T>C changes did increase significantly. These data indicate that mutation signatures developed through defined mutations and deletions in *S. cerevisiae* will be broadly applicable to mammalian systems. Mutation signatures from *C. elegans* also resemble human cancer signatures, albeit with some minor discrepancies (54,65). We note that *C. elegans* lacks a *MSH3* homolog, which is present in both yeast and humans (66).

Mutation signatures observed in human cancers are routinely used to predict mechanisms of tumorigenesis. We compared our SNV trinucleotide context profiles with the COSMIC single base substitution (SBS) dataset (67), by hierarchical cluster analysis (**Fig. S4, Table S19**). Overall, the C>A and C>T changes appeared to drive the clustering. Our samples formed a distinct cluster, which included COSMIC signature SBS32, a mutation profile associated with azathioprine treatment (**Fig. S4, box I)**. The COSMIC cluster most correlated with our samples (**Fig. S4, box II, Table S20**) included SBS36 (associated with defective base excision repair), SBS38 (unknown association but found only in UV light-associated melanomas), SBS18 (possibly resulting from reactive oxygen species damage), SBS24 (associated with aflatoxin exposure), SBS29 (found in tobacco-chewing related cancers) and SBS52, SBS45 and SBS49 (noted as possible sequencing artefacts).

SBS6 and SBS15 are both most highly correlated with *rnr1Y285F* and *rnr1D57N* paired with *msh2Δ* or *msh6Δ* (**Table S20**). SBS6 most commonly occurs in colorectal and uterine cancers and is associated with defective MMR. It is possible that elevations in dCTP and dTTP contribute to this mutation signature in human cancers leading to synergistic increases in these types of mutations in the absence of certain MMR genes. SBS15 is characterized by C>T SNVs within NpCpG sequences in lung and stomach cancers with unknown origin. The origin of these cancers could be in part, due to defects in MMR coupled with skewed increases in dCTP and dTTP, consistent with the increased frequency of these SNVs at dinucleotides observed in *rnr1Y285F/A msh6Δ* backgrounds. SBS20 also correlated well with *rnr1Y285F* and *rnr1D57N* combined with *msh2Δ* or *msh6Δ*. This signature is associated with combined MMR and *POLD1* mutations (**Table S20**). This indicates that reduced replication fidelity resulting from either increased dNTP pools or decreased proofreading combine with MMR defects to generate similar mutation signatures.

In addition, we also compared mutation signatures from our study to the COSMIC insertion and deletion (ID) signatures (67). A major caveat to comparing *CAN1* ID mutations to human mutation signatures is the lack of homopolymeric sequences within *CAN1*. Nonetheless, within the sequence contexts available, we noted increased G/C single base deletions in G/C context was present in COSMIC ID signatures ID3, ID7, ID9 and ID15. The proposed aetiology of ID Signature 3 is cigarette smoking, ID7 is MMR defects and ID9 and ID15 have unknown etiology. Elevated dNTP levels has not been part of a clinical diagnosis, but skewed increases in dNTPs likely contribute to these signatures of unknown aetiology.

Our results highlight the importance of considering altered dNTP pools and combinatorial effects of genetic backgrounds, when defining the source of tumor mutation signatures.

Materials and Methods

Strains and plasmids

All strains were derived from the W303 *RAD5+* background (**Table S1**). Construction of *rnr1D57N*, *rnr1Y285F* and *rnr1Y285A*, with and without *pGL-RNR1*, was described previously (33). Mismatch repair genes were deleted by amplifying *msh2Δ::kanMX, msh6Δ::kanMX,* and *msh3Δ::kanMX* chromosomal fragments from deletion collection strains utilizing primers A and D specific to each locus (**Table S2**). PCR products from these strains were used for transformation to replace the endogenous *MSH* gene with the *kanMX* cassette, which confers resistance to the drug, G418. Transformants were selected on YPD plates containing G418 and deletions were confirmed by PCR.

Measuring mutation rates at *CAN1*

Mutation rates were measured at the *CAN1* locus as previously described (6,33). Briefly, strains were grown on complete media (YPD) until colonies reach 2 mm in size. Colonies were then suspended in 100 µl of 1x TE (10 mM Tris-HCl, pH 7.5; 1 mM EDTA) and diluted 1:10,000. 20 µl of the undilute colony suspension was plated on SC-ARG + Canavanine and 100 µl of the 10^{-4} dilution was plated on synthetic complete plates lacking arginine. The plates were incubated at 30°C until colonies reached ~1 mm in size. Colonies were counted and mutation rates were calculated though FluCalc fluctuation analysis software (68).

Synthetic Genetic Array Analysis

The genetic screens were performed using SGA technology (69). Briefly, query strains carrying each of the *rnr1* alleles were crossed to an ordered array of all the viable yeast deletion strains and an array of temperature sensitive alleles of yeast essential genes. Diploid cells were transferred to a sporulation-

inducing medium, after which the germinated spores were selected for the simultaneous presence of the gene deletion and the *rnr1* allele. Colony size was quantified as a measure of fitness, and SGA scores and p-values were calculated as described in (69). SGA scores from deletion mutants and ts mutants were merged by scaling the ts screen scores according to the SGA scores of the deletion mutants that are present in the ts allele array. A z-score was calculated for all the genes in each screen, and a cutoff of z=-2 was applied to identify negative genetic interactions. The SGA data is presented in Tables S4-S10.

Gene Ontology enrichment analysis

GO term analysis was performed using the GO term finder tool (http://go.princeton.edu/) using a P-value cutoff of 0.01 and applying Bonferroni correction, querying biological process enrichment for each gene set. GO term enrichment results were further processed with REViGO (70) using the "Medium (0.7)" term similarity filter and simRel score as the semantic similarity measure. Terms with a frequency greater than 15% in the REViGO output were eliminated as too general.

Spatial Analysis of Functional Enrichment

Network annotations were made with the Python implementation of Spatial Analysis of Functional Enrichment (SAFE) ((47); https://github.com/baryshnikova-lab/safepy). The yeast genetic interaction similarity network and its functional domain annotations were obtained from (48).

Sample preparation and analysis pipeline

A detailed description of sample preparation and the analytical pipeline used for data analysis can be found in (33). Briefly, we pooled ~2000 colonies CanR colonies for each biological replicate and

extracted genomic DNA from the pool. Each genotype was represented by at least 4 biological replicates (see **Table S21**). The *CAN1* gene was amplified by PCR using KAPA HiFi (Roche) in 6 overlapping fragments that were purified using the Zymo ZR-96 DNA Clean-up Kit. Nextera barcode adaptors were added to the amplicons, followed by attachment of Illumina Nextera XT index primers set A (Illumina). Excess adapters were removed using Ampure XP beads (Beckman Coulter). Pooled samples were diluted to 4 nM, denatured using NaOH and loaded onto an Illumina MiSeq sequencing platform (PE300, V3) with 20% PhiX control in two separate runs to increase coverage and as a check for reproducibility. Paired end (2x300) deep sequencing of *CAN1* provided enough sequencing depth to determine mutation spectra for 150 unique samples, representing over 30 different genotypes, and including biological and technical replicates. *CAN1* was sequenced at an average depth of approximately 16,000 reads per base in *CAN1* per sample allowing for detailed characterization of mutation spectra.

Sequence reads were trimmed (CutAdapt version 1.14), specifying a quality score of Q30, and then processed using CLC Genomics Workbench Version 11. Paired-end reads were merged, primer locations were trimmed, and processed reads were aligned to the SacCer3 reference genome. CLC low frequency variant caller was used to call variants, with required significance of 0.01%. Variant files were exported from CLC as VCF files and downstream analysis was performed in RStudio (version 1.2.1335), paired with custom python scripting. All sequence variants are provided in **Table S22**.

Spearman Rank and Hierarchical Cluster Analysis

We used Spearman rank correlations among all genotypes (**Fig. S2**) to assess the relationship between genotype and mutation profile. Biological replicates from strains with lower mutation rates (e.g. wild-type, *msh3Δ*, *rnr1D57N*) were less well correlated than those from strains with higher

mutation rates (e.g., *msh2Δ, msh6Δ, rnr1Y285A*), indicating that mutation events occurred more systematically (less stochastically) in these genetic backgrounds, consistent with specific mutation profiles. For example, the average correlation of all 7 wildtype biological replicates with all other samples was r_s = 0.192, while the average correlation within the 7 wildtype biological replicates was r_s = 0.191. In contrast, the *rnr1Y285A msh3Δ* biological replicates were much more highly correlated to one another (r_s = 0.527) than they were with all other samples (r_s = 0.264).

Hierarchical Cluster Analysis was performed on data that was grouped by biological replicates and the number of times the different classes of variants occurred within a genotype. Data was combined on the average of unique counts of variants from all the biological replicates within a genotype. The different classes of variants include 6 classes of single nucleotide variants (SNVs), single base A/T or G/C insertions and deletions, complex insertions and deletions, as well as mononucleotide variants (MNVs) and replacements (Replac.). MNVs and replacements are complex errors identified in CLC. MNVs are dinucleotide SNVs, where two neighboring nucleotides are both mutated, ex: CC> AT. Replacements are complex insertions or deletions, where the deleted or replaced base is a variant. Two examples include AAC > G and C > AT. Both MNVs and replacements are extremely low frequency events and rarely occur in our data set; neither had a significant impact on clustering.

The data was condensed based on genotype by combining biological replicates and adding the total number of times a variant was seen in the total number of replicates. This analysis does not take frequency into account and instead totals how many unique types of variants occur in a sample. If the same variant occurred in multiple biological replicates within a genotype it was counted as such. Spearman correlation coefficients were calculated across all samples. Spearman correlation was chosen because we did not consistently observe linear relationships between variant frequencies in

our data set, although Pearson correlation yielded nearly identical results. Correlation analysis was plotted in RStudio using the ggplots and ggplot2 libraries.

Principal Components Analysis (PCA)

PCA was performed on the number of times a unique variant was observed within a genotype. PCA was plotted using the "factoextra" package in RStudio, with a singular value decomposition (SVD) approach using the prcomp() function.

Determining SNV in trinucleotide context

A detailed description can be found in (33).

Cluster Analysis to look at correlation between genotypes

To identify variants specific to a particular genotype and to eliminate frequency bias from variants that occurred early in the growth of a particular sample, we condensed unique variants based on the number of biological replicates sequenced for that genotype. If a particular variant occurred in 4 out of 4 biological replicates it was represented as 1, if it occurred in 3 out of 6 replicates it was represented as .5. This gives an unbiased approach to score for the probability that a particular variant was present in a particular genotype, without considering the absolute frequency at which that variant occurred. These data were clustered on rows (or unique variants), after applying a row sum cutoff of greater than 2. This cutoff eliminates low frequency variants which are less likely to be driving the differences in mutation spectra that we observe. By clustering the data only on unique variants, it allows us to see groups of different *types* of variants in specific sequence contexts that are potentially diagnostic for a particular genotype.

CC Dinucleotide cluster analysis

A python script using a sliding window approach was used to identify all reference positions containing CC dinucleotides within *CAN1*. Our dataset was then subset to include only variants that occurred in these dinucleotide CC positions. Of the 138 CC dinucleotide contexts across *CAN1*, 110 (~80%) were mutated, compared to 857/1,711 base pairs or ~50% of the base pairs in *CAN1.* Unique CC run variants were clustered based on the number of times that variant occurred in each genotype, while accounting for (normalizing by) the number of biological replicates sequenced for each genotype, as described above. Heatmaps were plotted using the pheatmap library in RStudio and motif enrichment was performed using Berkeley web logos (71).

COSMIC SBS signature cluster analysis

We performed hierarchical clustering to compare the mutation spectra from our study with human mutation signatures, through an unbiased approach. The single base substitutions (SBS) COSMIC signatures from GRCh38 (v3.2- March 2021, https://cancer.sanger.ac.uk/signatures/downloads/) were combined with the normalized SNVs in trinucleotide context (**Fig. 4)** (67). Hierarchical cluster analysis was then performed on these combined data.

Acknowledgements

We gratefully acknowledge Dr. Anastasia Baryshnikova for her help with the SAFE analysis. N.A.L. is a University at Buffalo Presidential Scholar. This work was supported by the American Cancer Society (RSG-14-2350-01) to J.A.S. J.A.S. is an ACS Research Scholar. J.A.S. is also grateful for support from the

University at Buffalo's Genome, Environment and Microbiome Community of Excellence. This work was also supported by a grant from the Canadian Institutes of Health Research (FDN-159913 to GWB).

Figure Legends

Figure 1. Mutation Spectra are distinct across genotypes. (A) The relative distribution of SNVs, deletions, insertions, MNVs and replacements by genotype. **(B)** The SNV spectra normalized relative total variants. **(C)** The deletion spectra normalized relative to total variants. (D) The insertion spectra normalized relative to total variants.

Figure 2. Genome-scale screens for synthetic fitness defects with *rnr1* alleles.

(**A**) The overlap of the *rnr1* negative genetic interactions for the three SGA screens is plotted as a Venn diagram. The number of genes identified in each screen is indicated, as is the *rnr1* allele for each screen. (**B**) GO-term enrichments for the negative interacting genes from each *rnr1* screen are tabulated. The –fold enrichment for each term is indicated. Note that *rnr1Y285F* did not display any statistically supported enrichment. (**C**) Spatial analysis of functional enrichment. On the left, the yeast genetic interaction similarity network is annotated with GO biological process terms to identify major functional domains (48). 13 of the 17 domains are labeled and delineated by coloured outlines. On the right, the network is annotated with negative genetic interactions from each *rnr1* SGA screen. The overlays indicate the functional domains annotated on the left. Only nodes with statistically supported enrichments (SAFE score > 0.08, $p < 0.05$) are colored. (**D**) Tetrad analysis of *rnr1* x *msh6*Δ crosses. Ten tetrads were dissected for each cross, and colonies were imaged after 3 days. Each column of 4 colonies is the 4 spores from a single meiotic ascus. Genotypes are indicated by circles (*msh6*Δ) and

squares (*rnr1*). (**E**) Tetrad analysis of *rnr1Y285A* x *msh2*Δ and *rnr1Y285A* x *msh3*Δ crosses. Genotypes are indicated by circles (*msh2*Δ or *msh3*Δ) and squares (*rnr1Y285A*).

Figure 3. Distinct genotypes share unique features. (**A**) Hierarchical cluster analysis using Spearman Rank correlation on all the distinct genotypes in our study. Data was clustered based on the unique counts of the 14 different classes of variants that occurred in each genotype. The histogram shows the distribution of correlation coefficients across samples. (**B**) Principal components analysis performed on all unique variants from biological replicates within a genotype.

Figure 4. The average number of each SNV as it occurs in unique triplet nucleotide context differs by genotype. Bars are colored according to the six different types of SNVs. (**A**) The 16 different triplet contexts are lettered for display purposes. The variant change (C>A, turquoise bar) occurs at the middle nucleotide marked X in each triplet context. The same triplet context is repeated for each possible variant in the following panels. **(B)** Wildtype and single *rnr1* alleles. **(C)** Genotypes with *MSH2* deleted **(D)** Genotypes with *MSH3* deleted. **(E)** Genotypes with *MSH6* deleted.

Figure 5. Variants that occur in unique sequence contexts cluster together Hierarchical cluster analysis of all unique variants within our study by genotype (A) A heatmap displaying Pearson correlation enrichment value for a given variant between genotypes, with notable clusters boxed in black. (B) 12 base window motif enrichment on sequence contexts surrounding the notable clusters. (C) The type of variant observed in the center of sequence context from B. (D) Summary of genotypes which show negative or positive correlation in a given cluster.

Figure 6. SNVs occur in C-C dinucleotide sequence contexts with differential enrichment between genotypes. (A) Hierarchical cluster analysis of all SNVs that occur at CC dinucleotides. Clusters of interest are boxed in black, labeled by roman numerals. (**B**) Motif enrichment of a 12 base window surrounding the mutated nucleotide was performed using Berkley web logos. The mutated base is at the 7th nucleotide position in all logos, indicated by the black arrow. (**C**) The most predominant type(s) of SNV in the cluster are displayed. (**D**) A summary of genotypes that show negative or positive correlation in a given cluster.

Figure 7. G/C single base deletions that show synergistic increases in variant frequency in *rnr1Y285A msh3Δ* genotypes. The deleted base is starred. (**A**) Motif enrichment of a 12 base window surrounding two G deletions (starred nucleotide) that are specific to genotypes with *msh3Δ*. (**B**) The variant frequency from biological replicates in wildtype, *rnr1Y285F/A* and *msh3Δ* genotypes for the single base deletion that occurred at position 31979 is plotted. (**C)** The variant frequencies across replicates for the G deletion at 32272. In both cases (**B** and **C**) there are very few events in the single mutants, but significant frequencies in double mutant backgrounds. (**D**) Motif enrichment for two C deletions that are specific to genotypes with *rnr1Y285F* and *rnr1Y285A*, but increase synergistically in double mutants with *msh3Δ* in combination. (**E)** The variant frequencies across replicates for the G deletion at 32747. (**F**) The variant frequencies across replicates for the G deletion at 32658. In **E** and **F**, this event occurs in *rnr1Y285A*, but not *rnr1Y285F* or *msh3Δ*. It occurs at increased frequencies in double mutant backgrounds.

Figure 8. Mechanisms of mutagenesis for the incorporation of errors specific to either Msh2-Msh6 or Msh2-Msh3 repair. The mutated base is represented in red. (**A**)**.** Two examples of sequence context

surrounding CC dinucleotides from **Fig. 6B**, where mis-insertion is due to the nucleotide in excess in *rnr1Y285F/A* backgrounds. These errors are efficiently repaired by Msh2-Msh6. This specificity becomes apparent when *msh6Δ* is paired with *rnr1Y285F/A* alleles. (**B**) Two examples of sequence context from **Figs. 7A** and **7D**, where mis-alignment events occur due to the severely limiting amount of dGTP in *rnr1Y285F/A* genetic backgrounds. The run where the deletion occurred is underlined in black. These single base G/C deletions are efficiently repaired by Msh2-Msh3, but not Msh2-Msh6, a previously unidentified specificity of the repair complex.

References

1. Kunkel, T.A. (2004) DNA replication fidelity. *J Biol Chem*, **279**, 16895-16898.
2. McCulloch, S.D. and Kunkel, T.A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*, **18**, 148-161.
3. Ganai, R.A. and Johansson, E. (2016) DNA Replication-A Matter of Fidelity. *Mol Cell*, **62**, 745-755.
4. St Charles, J.A., Liberti, S.E., Williams, J.S., Lujan, S.A. and Kunkel, T.A. (2015) Quantifying the contributions of base selectivity, proofreading and mismatch repair to nuclear DNA replication in Saccharomyces cerevisiae. *DNA Repair (Amst)*, **31**, 41-51.
5. Chabes, A., Georgieva, B., Domkin, V., Zhao, X., Rothstein, R. and Thelander, L. (2003) Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. *Cell*, **112**, 391-401.
6. Xu, X., Page, J.L., Surtees, J.A., Liu, H., Lagedrost, S., Lu, Y., Bronson, R., Alani, E., Nikitin, A.Y. and Weiss, R.S. (2008) Broad overexpression of ribonucleotide reductase genes in mice specifically induces lung neoplasms. *Cancer Res*, **68**, 2652-2660.
7. Kumar, D., Viberg, J., Nilsson, A.K. and Chabes, A. (2010) Highly mutagenic and severely imbalanced dNTP pools can escape detection by the S-phase checkpoint. *Nucleic Acids Res*, **38**, 3975-3983.
8. Kumar, D., Abdulovic, A.L., Viberg, J., Nilsson, A.K., Kunkel, T.A. and Chabes, A. (2011) Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic Acids Res*, **39**, 1360-1371.
9. Watt, D.L., Buckland, R.J., Lujan, S.A., Kunkel, T.A. and Chabes, A. (2016) Genome-wide analysis of the specificity and mechanisms of replication infidelity driven by imbalanced dNTP pools. *Nucleic Acids Res*, **44**, 1669-1680.
10. Buckland, R.J., Watt, D.L., Chittoor, B., Nilsson, A.K., Kunkel, T.A. and Chabes, A. (2014) Increased and imbalanced dNTP pools symmetrically promote both leading and lagging strand replication infidelity. *PLoS Genet*, **10**, e1004846.
11. Mathews, C.K. (2015) Deoxyribonucleotide metabolism, mutagenesis and cancer. *Nat Rev Cancer*, **15**, 528-539.
12. Connor, A.A., Denroche, R.E., Jang, G.H., Timms, L., Kalimuthu, S.N., Selander, I., McPherson, T., Wilson, G.W., Chan-Seng-Yue, M.A., Borozan, I. *et al.* (2017) Association of Distinct Mutational Signatures With Correlates of Increased Immune Activity in Pancreatic Ductal Adenocarcinoma. *JAMA Oncol*, **3**, 774-783.
13. Kunkel, T.A. and Erie, D.A. (2015) Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet*, **49**, 291-313.
14. Jensen, L.E., Jauert, P.A. and Kirkpatrick, D.T. (2005) The large loop repair and mismatch repair pathways of Saccharomyces cerevisiae act on distinct substrates during meiosis. *Genetics*, **170**, 1033-1043.
15. Sia, E.A., Kokoska, R.J., Dominska, M., Greenwell, P. and Petes, T.D. (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol Cell Biol*, **17**, 2851-2858.
16. Harrington, J.M. and Kolodner, R.D. (2007) Saccharomyces cerevisiae Msh2-Msh3 acts in repair of base-base mispairs. *Mol Cell Biol*, **27**, 6546-6554.

17. Srivatsan, A., Bowen, N. and Kolodner, R.D. (2014) Mispair-specific recruitment of the Mlh1-Pms1 complex identifies repair substrates of the Saccharomyces cerevisiae Msh2-Msh3 complex. *J Biol Chem*, **289**, 9352-9364.
18. Mastrocola, A.S. and Heinen, C.D. (2010) Lynch syndrome-associated mutations in MSH2 alter DNA repair and checkpoint response functions in vivo. *Hum Mutat*, **31**, E1699-1708.
19. Pino, M.S., Mino-Kenudson, M., Wildemore, B.M., Ganguly, A., Batten, J., Sperduti, I., Iafrate, A.J. and Chung, D.C. (2009) Deficient DNA mismatch repair is common in Lynch syndrome-associated colorectal adenomas. *J Mol Diagn*, **11**, 238-247.
20. Heinen, C.D. (2010) Genotype to phenotype: analyzing the effects of inherited mutations in colorectal cancer families. *Mutat Res*, **693**, 32-45.
21. Davies, H., Morganella, S., Purdie, C.A., Jang, S.J., Borgen, E., Russnes, H., Glodzik, D., Zou, X., Viari, A., Richardson, A.L. *et al.* (2017) Whole-Genome Sequencing Reveals Breast Cancers with Mismatch Repair Deficiency. *Cancer Res*, **77**, 4755-4762.
22. Fusco, N., Lopez, G., Corti, C., Pesenti, C., Colapietro, P., Ercoli, G., Gaudioso, G., Faversani, A., Gambini, D., Michelotti, A. *et al.* (2018) Mismatch Repair Protein Loss as a Prognostic and Predictive Biomarker in Breast Cancers Regardless of Microsatellite Instability. *JNCI Cancer Spectr*, **2**, pky056.
23. Edelmann, W., Umar, A., Yang, K., Heyer, J., Kucherlapati, M., Lia, M., Kneitz, B., Avdievich, E., Fan, K., Wong, E. *et al.* (2000) The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. *Cancer Res*, **60**, 803-807.
24. van Oers, J.M., Edwards, Y., Chahwan, R., Zhang, W., Smith, C., Pechuan, X., Schaetzlein, S., Jin, B., Wang, Y., Bergman, A. *et al.* (2014) The MutSβ complex is a modulator of p53-driven tumorigenesis through its functions in both DNA double-strand break repair and mismatch repair. *Oncogene*, **33**, 3939-3946.
25. Santos, L.S., Silva, S.N., Gil, O.M., Ferreira, T.C., Limbert, E. and Rueff, J. (2018) Mismatch repair single nucleotide polymorphisms and thyroid cancer susceptibility. *Oncol Lett*, **15**, 6715-6726.
26. Morak, M., Käsbauer, S., Kerscher, M., Laner, A., Nissen, A.M., Benet-Pagès, A., Schackert, H.K., Keller, G., Massdorf, T. and Holinski-Feder, E. (2017) Loss of MSH2 and MSH6 due to heterozygous germline defects in MSH3 and MSH6. *Fam Cancer*, **16**, 491-500.
27. Adam, R., Spier, I., Zhao, B., Kloth, M., Marquez, J., Hinrichsen, I., Kirfel, J., Tafazzoli, A., Horpaopan, S., Uhlhaas, S. *et al.* (2016) Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *Am J Hum Genet*, **99**, 337-351.
28. Valle, L., de Voer, R.M., Goldberg, Y., Sjursen, W., Försti, A., Ruiz-Ponte, C., Caldés, T., Garré, P., Olsen, M.F., Nordling, M. *et al.* (2019) Update on genetic predisposition to colorectal cancer and polyposis. *Mol Aspects Med*, **69**, 10-26.
29. Takahashi, M., Koi, M., Balaguer, F., Boland, C.R. and Goel, A. (2011) MSH3 mediates sensitization of colorectal cancer cells to cisplatin, oxaliplatin, and a poly(ADP-ribose) polymerase inhibitor. *J Biol Chem*, **286**, 12157-12165.
30. Park, J.M., Huang, S., Tougeron, D. and Sinicrope, F.A. (2013) MSH3 mismatch repair protein regulates sensitivity to cytotoxic drugs and a histone deacetylase inhibitor in human colon carcinoma cells. *PLoS One*, **8**, e65369.
31. Nogueira, G.A.S., Costa, E.F.D., Lopes-Aguiar, L., Lima, T.R.P., Visacri, M.B., Pincinato, E.C., Lourenço, G.J., Calonga, L., Mariano, F.V., Altemani, A. *et al.* (2018) Polymorphisms in DNA

mismatch repair pathway genes predict toxicity and response to cisplatin chemoradiation in head and neck squamous cell carcinoma patients. *Oncotarget*, **9**, 29538-29547.
32. Carethers, J.M., Koi, M. and Tseng-Rogenski, S.S. (2015) EMAST is a Form of Microsatellite Instability That is Initiated by Inflammation and Modulates Colorectal Cancer Progression. *Genes (Basel)*, **6**, 185-205.
33. Lamb, N.A., Bard, J.E., Buck, M.J. and Surtees, J.A. (2021) A selection-based next generation sequencing approach to develop robust, genotype-specific mutation profiles in Saccharomyces cerevisiae. *G3 (Bethesda)*.
34. Kunkel, T.A. and Erie, D.A. (2005) DNA mismatch repair. *Annu Rev Biochem*, **74**, 681-710.
35. Lang, G.I., Parsons, L. and Gammie, A.E. (2013) Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3 (Bethesda)*, **3**, 1453-1465.
36. Kumar, C., Piacente, S.C., Sibert, J., Bukata, A.R., O'Connor, J., Alani, E. and Surtees, J.A. (2011) Multiple factors insulate Msh2-Msh6 mismatch repair activity from defects in Msh2 domain I. *J Mol Biol*, **411**, 765-780.
37. Serero, A., Jubin, C., Loeillet, S., Legoix-Né, P. and Nicolas, A.G. (2014) Mutational landscape of yeast mutator strains. *Proc Natl Acad Sci U S A*, **111**, 1897-1902.
38. Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1996) Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. *Curr Biol*, **6**, 1185-1187.
39. Surtees, J.A. and Alani, E. (2006) Mismatch repair factor MSH2-MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination. *J Mol Biol*, **360**, 523-536.
40. Bowers, J., Sokolsky, T., Quach, T. and Alani, E. (1999) A mutation in the MSH6 subunit of the Saccharomyces cerevisiae MSH2-MSH6 complex disrupts mismatch recognition. *J Biol Chem*, **274**, 16115-16125.
41. Genschel, J., Littman, S.J., Drummond, J.T. and Modrich, P. (1998) Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. *J Biol Chem*, **273**, 19895-19901.
42. Kunkel, T.A. (2009) Evolving views of DNA replication (in)fidelity. *Cold Spring Harb Symp Quant Biol*, **74**, 91-101.
43. Arana, M.E. and Kunkel, T.A. (2010) Mutator phenotypes due to DNA replication infidelity. *Semin Cancer Biol*, **20**, 304-311.
44. Kunkel, T.A. (2011) Balancing eukaryotic replication asymmetry with replication fidelity. *Curr Opin Chem Biol*, **15**, 620-626.
45. Aye, Y., Li, M., Long, M.J. and Weiss, R.S. (2015) Ribonucleotide reductase and cancer: biological mechanisms and targeted therapies. *Oncogene*, **34**, 2011-2021.
46. Marsischky, G.T., Filosi, N., Kane, M.F. and Kolodner, R. (1996) Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Dev*, **10**, 407-420.
47. Baryshnikova, A. (2016) Systematic Functional Annotation and Visualization of Biological Networks. *Cell Systems*, **2**, 412-421.
48. Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D. *et al.* (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science*, **353**.

49. Washington, N.L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E.T., Barrett, K.M.S., Larsen, B.B., Anderson, C., White, S., Cassens, T. *et al.* (2021) Genomic epidemiology identifies emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *medRxiv*, 2021.2002.2006.21251159.
50. Kumar, D., Abdulovic, A.L., Viberg, J.r., Nilsson, A.K., Kunkel, T.A. and Chabes, A. (2011) Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic Acids Research*, **39**, 1360-1371.
51. Lujan, S.A., Clausen, A.R., Clark, A.B., MacAlpine, H.K., MacAlpine, D.M., Malc, E.P., Mieczkowski, P.A., Burkholder, A.B., Fargo, D.C., Gordenin, D.A. *et al.* (2014) Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res*, **24**, 1751-1764.
52. Kunkel, T.A. and Soni, A. (1988) Mutagenesis by transient misalignment. *J Biol Chem*, **263**, 14784-14789.
53. Schmidt, T.T., Sharma, S., Reyes, G.X., Gries, K., Gross, M., Zhao, B., Yuan, J.H., Wade, R., Chabes, A. and Hombauer, H. (2019) A genetic screen pinpoints ribonucleotide reductase residues that sustain dNTP homeostasis and specifies a highly mutagenic type of dNTP imbalance. *Nucleic Acids Res*, **47**, 237-252.
54. Meier, B., Volkova, N.V., Hong, Y., Schofield, P., Campbell, P.J., Gerstung, M. and Gartner, A. (2018) Mutational signatures of DNA mismatch repair deficiency in C. elegans and human cancers. *Genome Res*, **28**, 666-675.
55. Lujan, S.A., Williams, J.S., Pursell, Z.F., Abdulovic-Cui, A.A., Clark, A.B., Nick McElhinny, S.A. and Kunkel, T.A. (2012) Mismatch repair balances leading and lagging strand DNA replication fidelity. *PLoS Genet*, **8**, e1003016.
56. Pavlov, Y.I., Mian, I.M. and Kunkel, T.A. (2003) Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. *Curr Biol*, **13**, 744-748.
57. Rubin, A.F. and Green, P. (2009) Mutation patterns in cancer genomes. *Proc Natl Acad Sci U S A*, **106**, 21766-21770.
58. Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C. *et al.* (2007) Patterns of somatic mutation in human cancer genomes. *Nature*, **446**, 153-158.
59. Lee, W., Jiang, Z., Liu, J., Haverty, P.M., Guan, Y., Stinson, J., Yue, P., Zhang, Y., Pant, K.P., Bhatt, D. *et al.* (2010) The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature*, **465**, 473-477.
60. Herr, A.J., Kennedy, S.R., Knowels, G.M., Schultz, E.M. and Preston, B.D. (2014) DNA replication error-induced extinction of diploid yeast. *Genetics*, **196**, 677-691.
61. Schmidt, T.T., Reyes, G., Gries, K., Ceylan, C., Sharma, S., Meurer, M., Knop, M., Chabes, A. and Hombauer, H. (2017) Alterations in cellular metabolism triggered by URA7 or GLN3 inactivation cause imbalanced dNTP pools and increased mutagenesis. *Proc Natl Acad Sci U S A*, **114**, E4442-e4451.
62. Williams, L.N., Herr, A.J. and Preston, B.D. (2013) Emergence of DNA polymerase ε antimutators that escape error-induced extinction in yeast. *Genetics*, **193**, 751-770.
63. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415-421.

64. Zou, X., Owusu, M., Harris, R., Jackson, S.P., Loizou, J.I. and Nik-Zainal, S. (2018) Validating the concept of mutational signatures with isogenic cell models. *Nat Commun*, **9**, 1744.
65. Meier, B., Cooke, S.L., Weiss, J., Bailly, A.P., Alexandrov, L.B., Marshall, J., Raine, K., Maddison, M., Anderson, E., Stratton, M.R. *et al.* (2014) C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome Res*, **24**, 1624-1636.
66. Denver, D.R., Feinberg, S., Estes, S., Thomas, W.K. and Lynch, M. (2005) Mutation rates, spectra and hotspots in mismatch repair-deficient Caenorhabditis elegans. *Genetics*, **170**, 107-113.
67. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N. *et al.* (2020) The repertoire of mutational signatures in human cancer. *Nature*, **578**, 94-101.
68. Radchenko, E.A., McGinty, R.J., Aksenova, A.Y., Neil, A.J. and Mirkin, S.M. (2018) Quantitative Analysis of the Rates for Repeat-Mediated Genome Instability in a Yeast Experimental System. *Methods Mol Biol*, **1672**, 421-438.
69. Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.Y., Ou, J., San Luis, B.J., Bandyopadhyay, S. *et al.* (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Methods*, **7**, 1017-1024.
70. Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, **6**, e21800.
71. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188-1190.

Supplementary Figure Legends

Figure S1. Absolute variant frequency is consistent with selection at *CAN1.* (A) The absolute variant frequency was calculated by taking the sum of variant frequencies in each genotype and averaging for all biological replicates sequenced in a genotype. Error bars represent the standard deviation between biological replicates within a genotype. The number of total biological replicates sequenced varied by genotype; numbers are displayed in Table S3. Samples grown under permissive conditions, from four genotypes, are graphed in peach. All other samples grown under selection are graphed in turquoise. **(B)** The average absolute variant frequency before (red) and after (turquoise) the permissive variant filter was applied.

Figure S2. Hierarchical cluster analysis using Spearman rank correlation on all individual samples in our study. Boxed in black are notable clusters where biological replicates from the same genotypes cluster together. Throughout, biological replicates tended to cluster. Similarly, *rnr1Y285F pGAL-RNR1* and *rnr1Y285A pGAL-RNR1* strains clustered with *rnr1Y285F* and *rnr1Y285A* strain, respectively.

Figure S3. Mutation spectra comparison between rnr1 alleles and *rnr1-pGAL-RNR1* counterparts. **(A)** The SNV spectra normalized out of total SNVs and **(B)** total variants. **(C)** The deletion spectra normalized out of total deletions and **(D)** total variants. **(E)** The insertion spectra normalized out of total insertions and **(F)** total variants.

Figure S4. Hierarchical cluster analysis of mutation signatures from this study compared with COSMIC SBS signatures. The single base substitutions (SBS) COSMIC signatures from GRCh38 (v3.2-March 2021, https://cancer.sanger.ac.uk/signatures/downloads/) were combined with the normalized

SNVs in trinucleotide context (**Fig. 4).** Our samples formed a distinct cluster, which included COSMIC signature SBS32, a mutation profile associated with azathioprine treatment (**box I**). A second cluster (**box II**) correlated well with several of our samples (**Table S20**).

Table 1. Canavanine mutation rates

Genotype	Mutation Rate [95% Confidence Intervals]	Fold Change
Wildtype (n=115)	2.7×10^{-7} [$2.3 \times 10^{-7} – 3.2 \times 10^{-7}$]	1
msh2Δ (n=24)	6.7×10^{-6} [$5.8 \times 10^{-6} – 7.8 \times 10^{-6}$]	24.8
msh3Δ (n=32)	7.8×10^{-7} [$6.1 \times 10^{-7} – 9.7 \times 10^{-7}$]	2.9
msh6Δ (n=24)	1.8×10^{-6} [$1.4 \times 10^{-6} – 2.2 \times 10^{-6}$]	6.7
rnr1D57N (n=91)	7.3×10^{-7} [$6.3 \times 10^{-7} – 8.3 \times 10^{-7}$]	2.7
rnr1D57N msh2Δ (n=24)	2.0×10^{-5} [$1.8 \times 10^{-5} – 2.3 \times 10^{-5}$]	74.1
rnr1D57N msh3Δ (n=24)	2.2×10^{-6} [$1.7 \times 10^{-6} – 2.7 \times 10^{-6}$]	8.2
rnr1D57N msh6Δ (n=24)	5.6×10^{-6} [$4.7 \times 10^{-6} – 6.6 \times 10^{-6}$]	20.7
rnr1Y285F (n=42)	7.5×10^{-7} [$5.8 \times 10^{-7} – 9.5 \times 10^{-7}$]	2.8
rnr1Y285F msh2Δ (n=24)	2.9×10^{-5} [$2.6 \times 10^{-5} – 3.2 \times 10^{-5}$]	107.4
rnr1Y285F msh3Δ (n=24)	3.7×10^{-6} [$3.0 \times 10^{-6} – 4.5 \times 10^{-6}$]	13.7
rnr1Y285F msh6Δ (n=24)	2.3×10^{-5} [$2.0 \times 10^{-5} – 2.6 \times 10^{-5}$]	85.2
rnr1Y285F pGAL-RNR1 (n=24)	7.8×10^{-7} [$5.7 \times 10^{-7} – 1.0 \times 10^{-6}$]	2.9
rnr1Y285F pGAL-RNR1 msh2Δ (n=24)	1.5×10^{-5} [$1.3 \times 10^{-5} – 1.8 \times 10^{-5}$]	55.6
rnr1Y285F pGAL-RNR1 msh6Δ (n=12)	1.6×10^{-5} [$1.3 \times 10^{-5} – 1.9 \times 10^{-5}$]	59.3
rnr1Y285A (n=46)	5.5×10^{-6} [$4.7 \times 10^{-6} – 6.2 \times 10^{-6}$]	20.4
rnr1Y285A msh2Δ (n=33)	2.5×10^{-5} [$2.1 \times 10^{-5} – 3.0 \times 10^{-5}$]	92.6
rnr1Y285A msh3Δ (n=49)	6.2×10^{-5} [$5.7 \times 10^{-5} – 6.6 \times 10^{-5}$]	229.6
rnr1Y285A msh6Δ (n=33)	1.0×10^{-5} [$8.9 \times 10^{-6} – 1.2 \times 10^{-5}$]	37.0
rnr1Y285A pGAL-RNR1 (n=22)	1.5×10^{-5} [$1.2 \times 10^{-5} – 1.7 \times 10^{-5}$]	55.6
rnr1Y285A pGAL-RNR1 msh2Δ (n=24)	1.3×10^{-5} [$1.1 \times 10^{-5} – 1.5 \times 10^{-5}$]	48.2
rnr1Y285A pGAL-RNR1 msh3Δ (n=36)	7.3×10^{-5} [$6.8 \times 10^{-5} – 7.9 \times 10^{-5}$]	270.4
rnr1Y285A pGAL-RNR1 msh6Δ (n=30)	1.0×10^{-5} [$8.7 \times 10^{-6} – 1.2 \times 10^{-5}$]	37.0

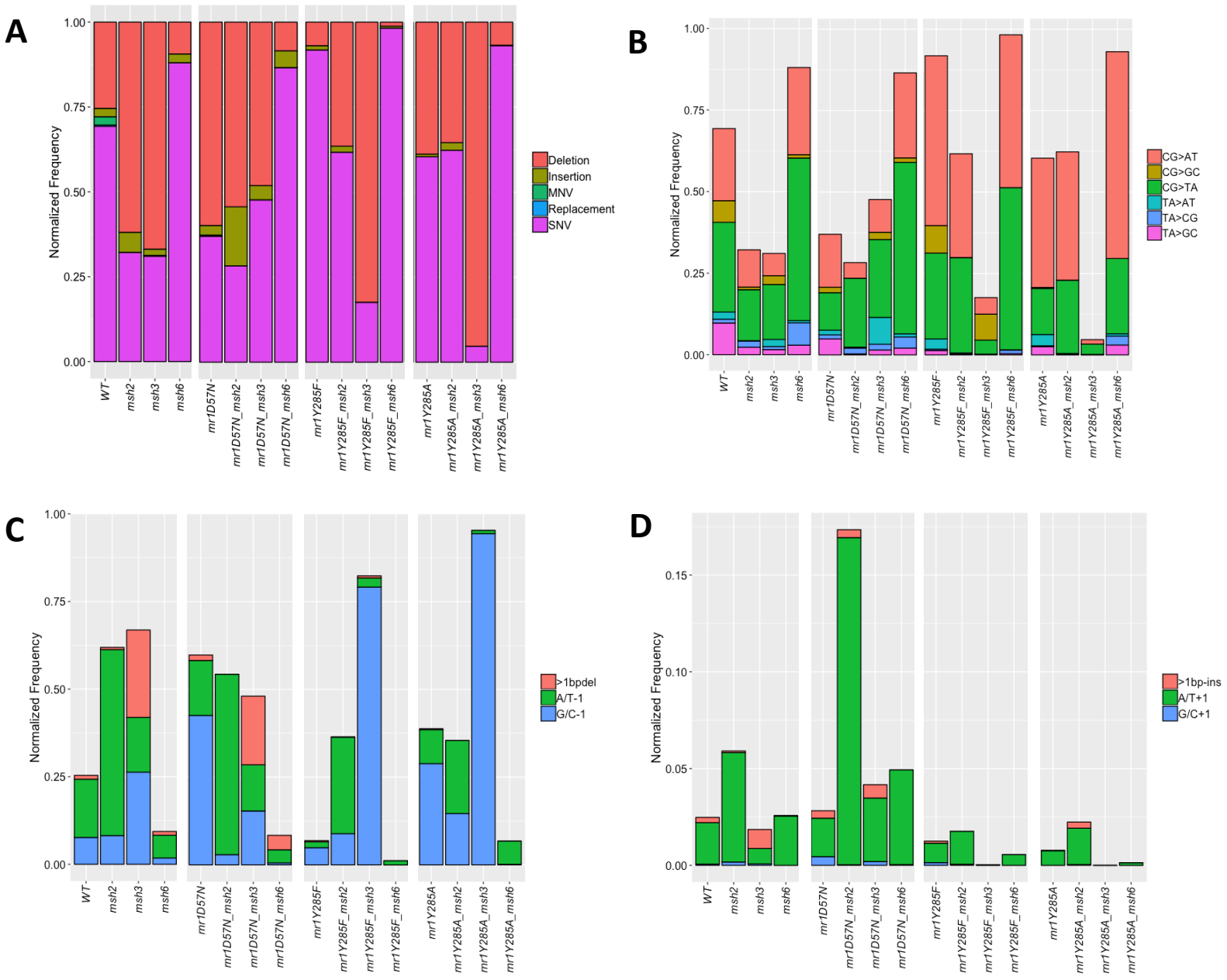
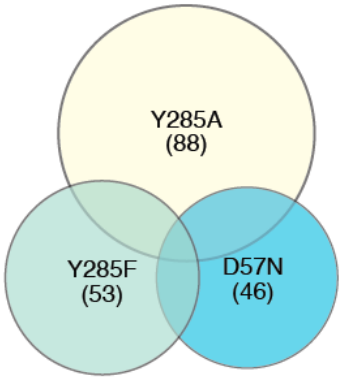

A
Normalized Frequency
1.00
0.75
0.50
0.25
0.00
Deletion
Insertion
MNV
Replacement
SNV
WT-
msh2-
msh3-
msh6-
mr1D57N-
mr1D57N_msh2-
mr1D57N_msh3-
mr1D57N_msh6-
mr1Y285F-
mr1Y285F_msh2-
mr1Y285F_msh3-
mr1Y285F_msh6-
mr1Y285A-
mr1Y285A_msh2-
mr1Y285A_msh3-
mr1Y285A_msh6-
B
CG>AT
CG>GC
CG>TA
TA>AT
TA>CG
TA>GC
C
>1bpdel
A/T-1
G/C-1
D
0.15
0.10
0.05
>1bp-ins
A/T+1
G/C+1

Figure 1.

A



B

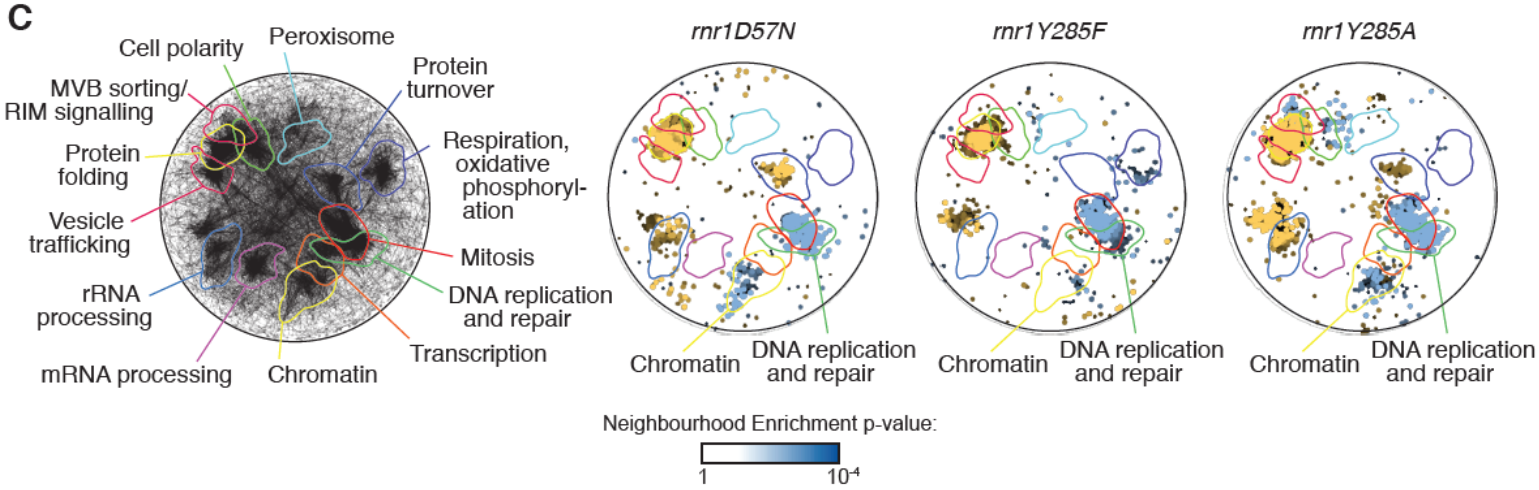
rnr1Y285A

GO term ID	Description	Enrichment
GO:0006298	mismatch repair	17
GO:0043570	maintenance of DNA repeat elements	14
GO:0006260	DNA replication	6
GO:0006281	DNA repair	4

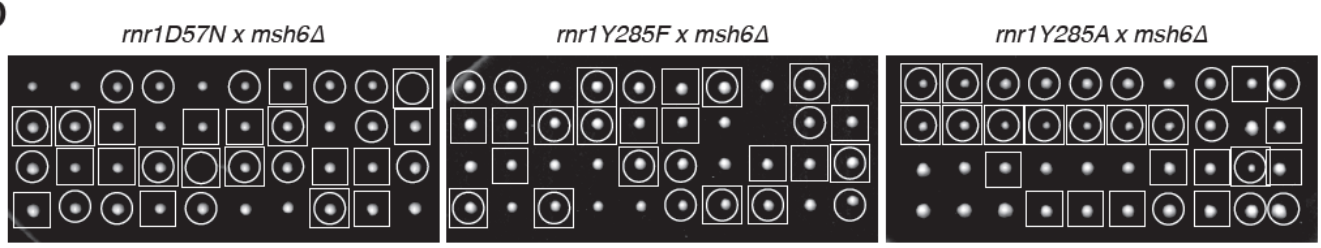
rnr1D57N

GO term ID	Description	Enrichment
GO:0006273	lagging strand replication	38
GO:0006260	DNA replication	7

C



D



E

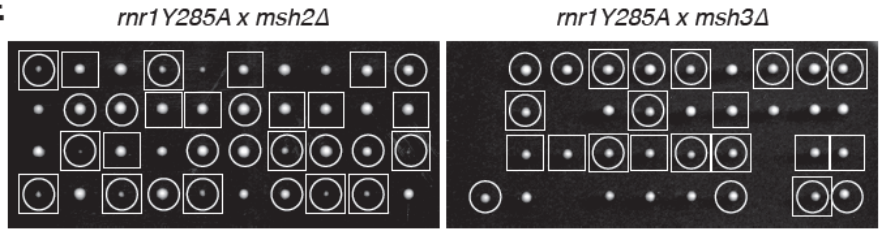


Figure 2.

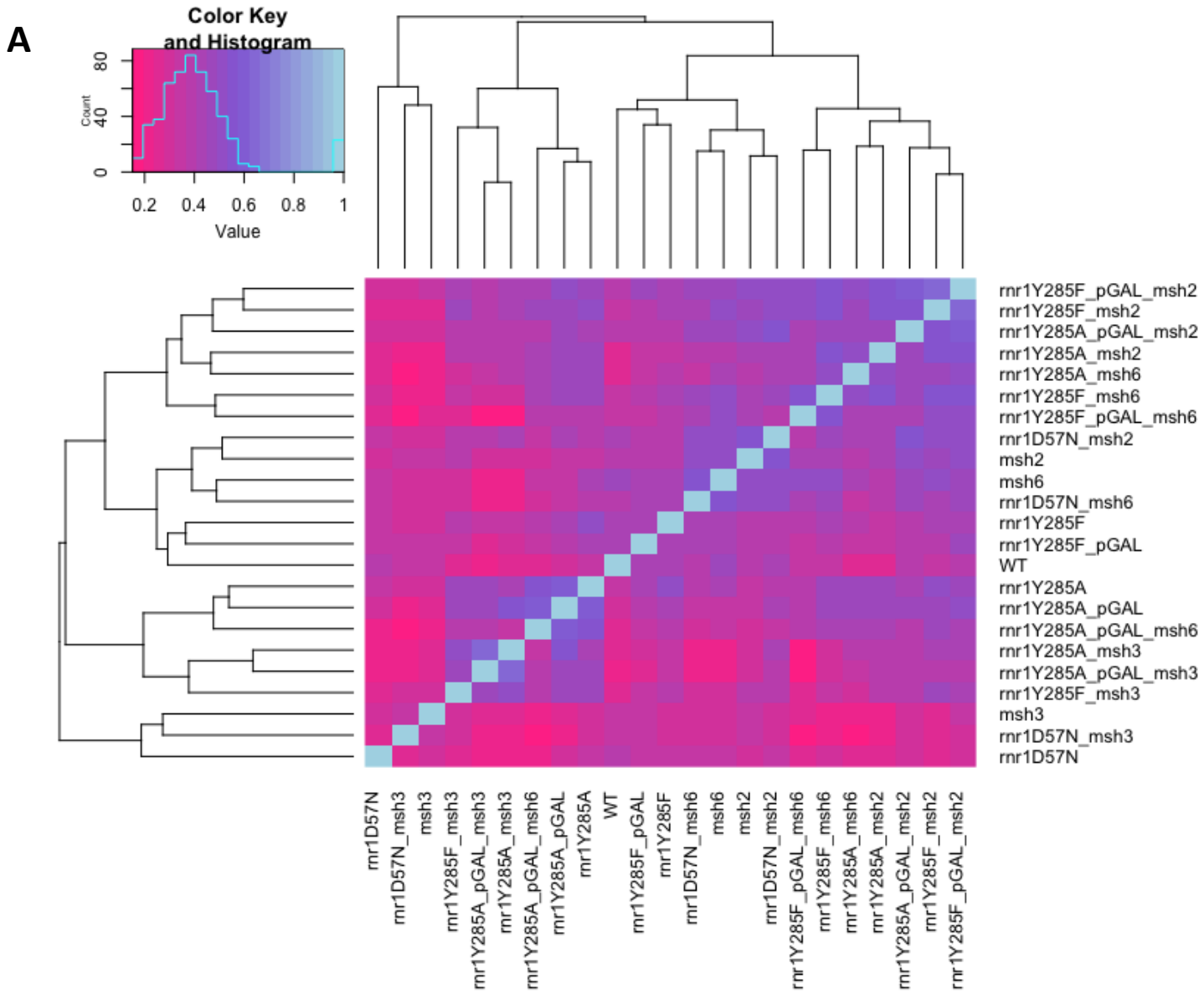

A
Color Key
and Histogram
Count
0
40
80
0.2
0.4
0.6
0.8
1
Value
rnr1Y285F_pGAL_msh2
rnr1Y285F_msh2
rnr1Y285A_pGAL_msh2
rnr1Y285A_msh2
rnr1Y285A_msh6
rnr1Y285F_msh6
rnr1Y285F_pGAL_msh6
rnr1D57N_msh2
msh2
msh6
rnr1D57N_msh6
rnr1Y285F
rnr1Y285F_pGAL
WT
rnr1Y285A
rnr1Y285A_pGAL
rnr1Y285A_pGAL_msh6
rnr1Y285A_msh3
rnr1Y285A_pGAL_msh3
rnr1Y285F_msh3
msh3
rnr1D57N_msh3
rnr1D57N
rnr1D57N
rnr1D57N_msh3
msh3
rnr1Y285F_msh3
rnr1Y285A_msh3
rnr1Y285A_pGAL_msh6
rnr1Y285A_pGAL
rnr1Y285A
WT
rnr1Y285F_pGAL
rnr1Y285F
rnr1D57N_msh6
msh6
msh2
rnr1D57N_msh2
rnr1Y285F_pGAL_msh6
rnr1Y285F_msh6
rnr1Y285A_msh6
rnr1Y285A_msh2
rnr1Y285A_pGAL_msh2
rnr1Y285F_msh2
rnr1Y285F_pGAL_msh2

Figure 3.

B

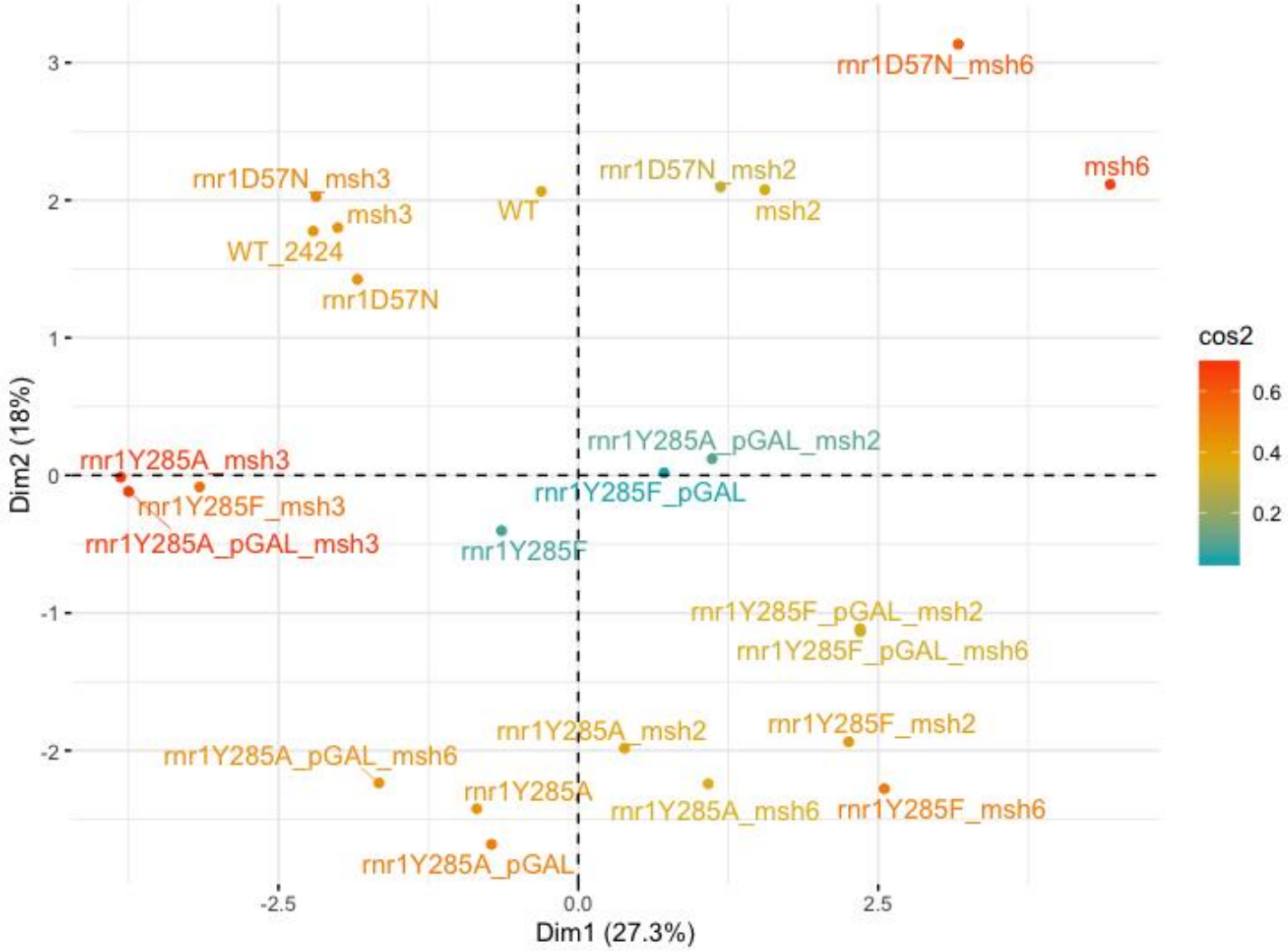


Figure 3. (continued)

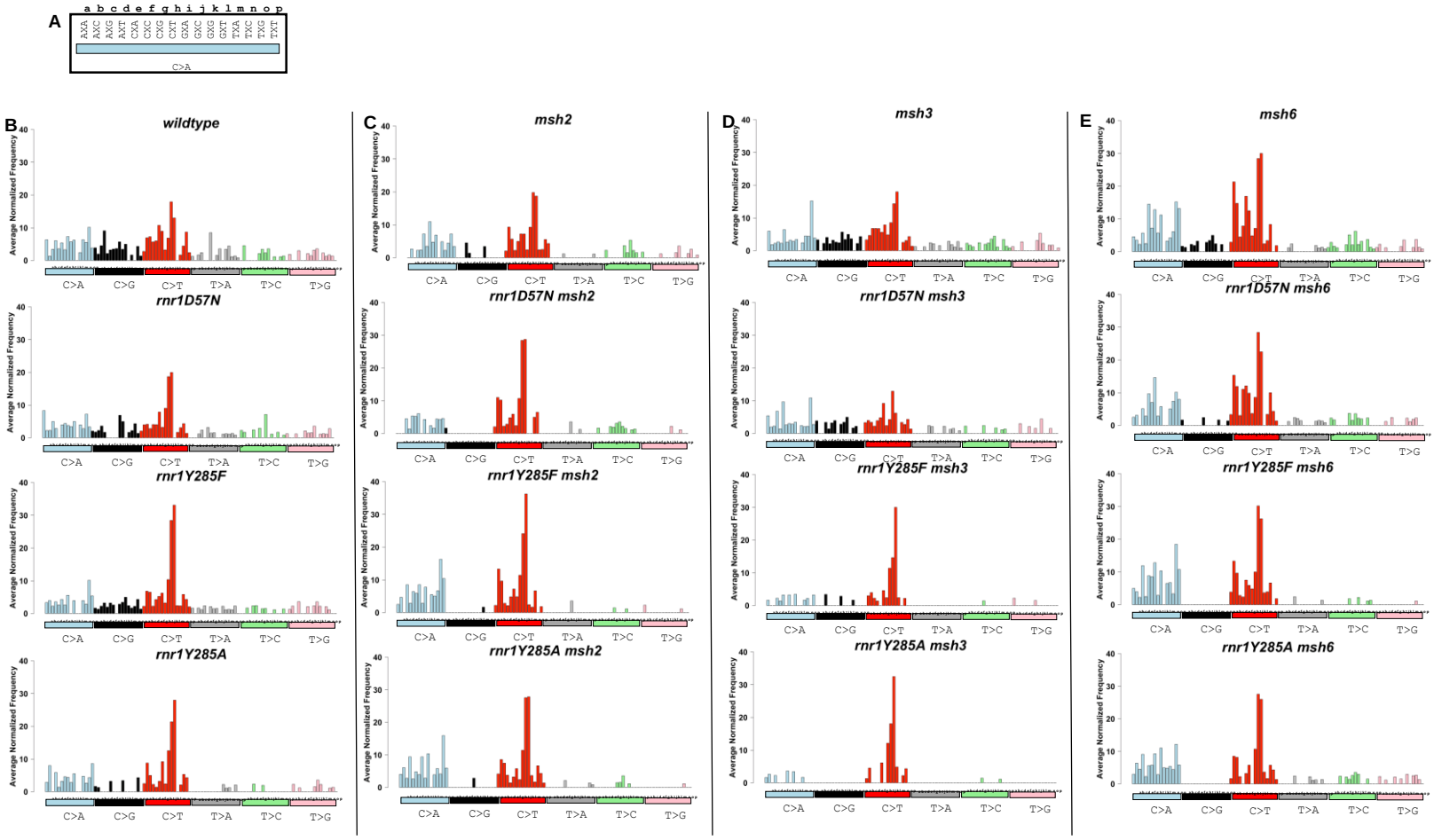

A
a b c d e f g h i j k l m n o p
AXA AXC AXG AXT CXA CXC CXG CXT GXA GXC GXG GXT TXA TXC TXG TXT
C>A
B
wildtype
rnr1D57N
rnr1Y285F
rnr1Y285A
C
msh2
rnr1D57N msh2
rnr1Y285F msh2
rnr1Y285A msh2
D
msh3
rnr1D57N msh3
rnr1Y285F msh3
rnr1Y285A msh3
E
msh6
rnr1D57N msh6
rnr1Y285F msh6
rnr1Y285A msh6
Average Normalized Frequency
C>A C>G C>T T>A T>C T>G

Figure 4.

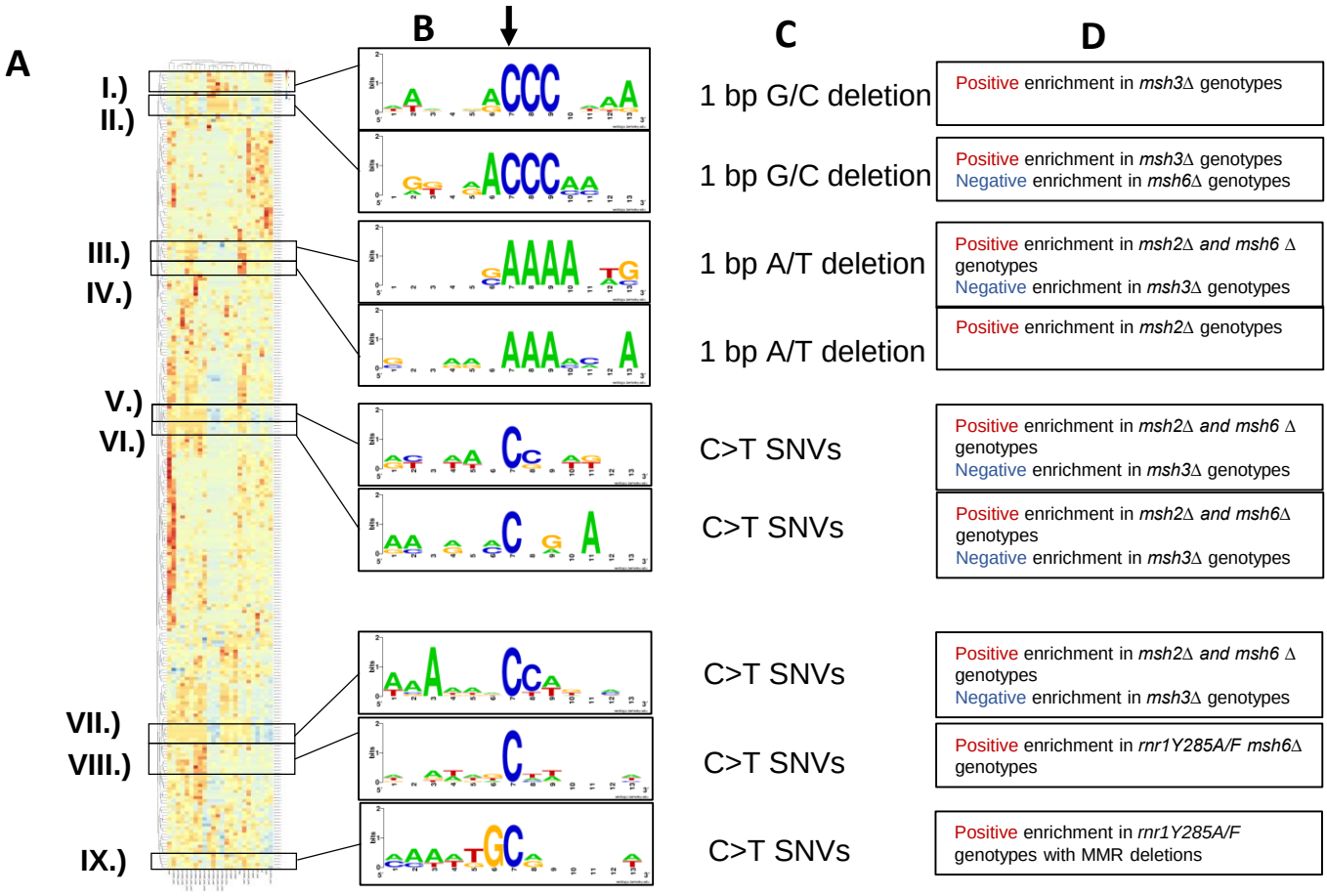

A
B
C
D
I.)
II.)
III.)
IV.)
V.)
VI.)
VII.)
VIII.)
IX.)
1 bp G/C deletion
Positive enrichment in msh3Δ genotypes
1 bp G/C deletion
Positive enrichment in msh3Δ genotypes
Negative enrichment in msh6Δ genotypes
1 bp A/T deletion
Positive enrichment in msh2Δ and msh6 Δ genotypes
Negative enrichment in msh3Δ genotypes
1 bp A/T deletion
Positive enrichment in msh2Δ genotypes
C>T SNVs
Positive enrichment in msh2Δ and msh6 Δ genotypes
Negative enrichment in msh3Δ genotypes
C>T SNVs
Positive enrichment in msh2Δ and msh6Δ genotypes
Negative enrichment in msh3Δ genotypes
C>T SNVs
Positive enrichment in msh2Δ and msh6 Δ genotypes
Negative enrichment in msh3Δ genotypes
C>T SNVs
Positive enrichment in rnr1Y285A/F msh6Δ genotypes
C>T SNVs
Positive enrichment in rnr1Y285A/F genotypes with MMR deletions

Figure 5.

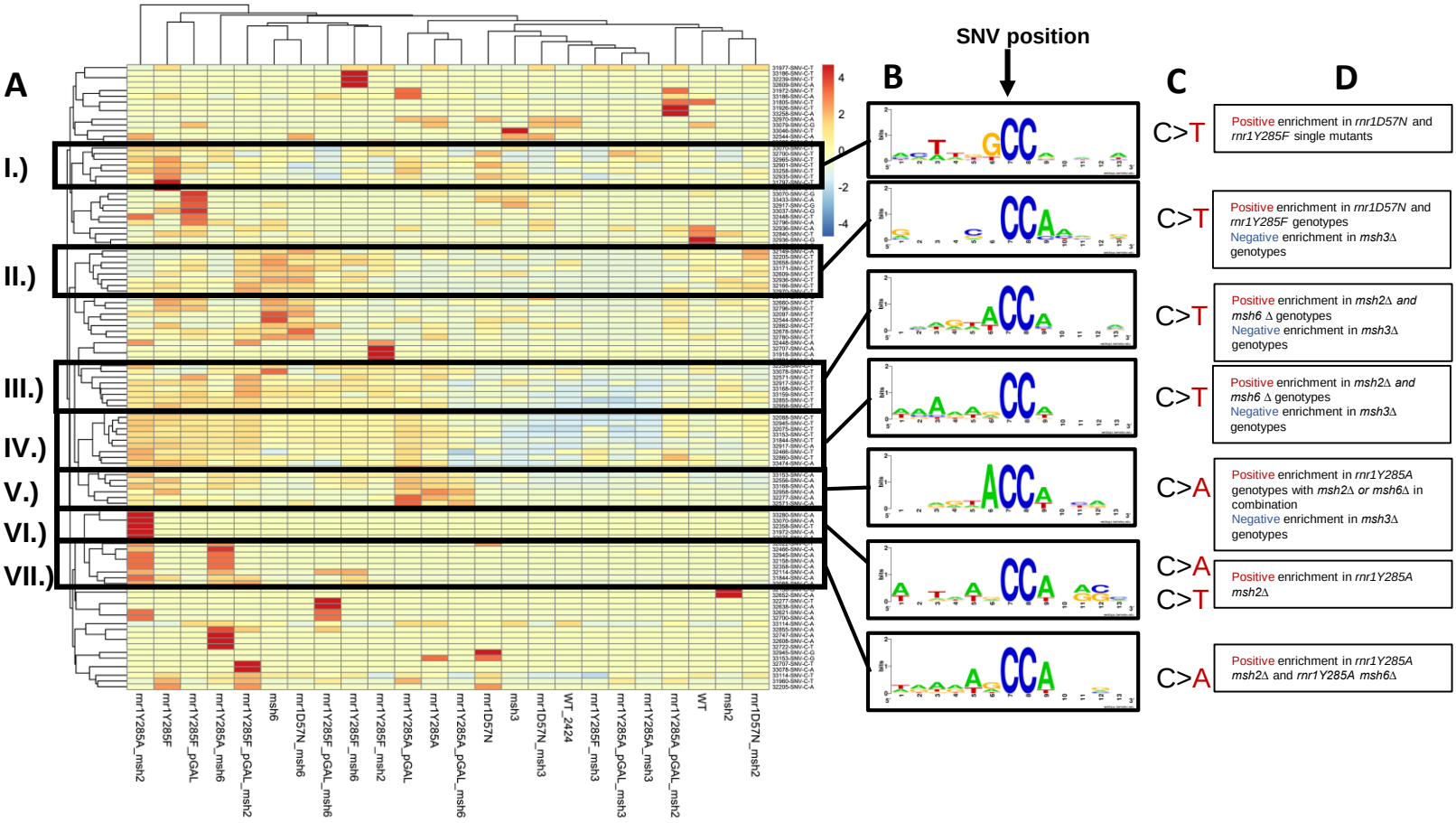

A
B
C
D
SNV position
I.)
II.)
III.)
IV.)
V.)
VI.)
VII.)
4
2
0
-2
-4
rnr1Y285A_msh2
rnr1Y285F
rnr1Y285F_pGAL
rnr1Y285A_msh6
rnr1Y285F_pGAL_msh2
msh6
rnr1D57N_msh6
rnr1Y285F_pGAL_msh6
rnr1Y285F_msh6
rnr1Y285F_msh2
rnr1Y285A_pGAL
rnr1Y285A
rnr1Y285A_pGAL_msh6
rnr1D57N
msh3
rnr1D57N_msh3
WT_2424
rnr1Y285F_msh3
rnr1Y285A_pGAL_msh3
rnr1Y285A_msh3
rnr1Y285A_pGAL_msh2
WT
msh2
rnr1D57N_msh2
C>T
Positive enrichment in rnr1D57N and rnr1Y285F single mutants
C>T
Positive enrichment in rnr1D57N and rnr1Y285F genotypes
Negative enrichment in msh3Δ genotypes
C>T
Positive enrichment in msh2Δ and msh6 Δ genotypes
Negative enrichment in msh3Δ genotypes
C>T
Positive enrichment in msh2Δ and msh6 Δ genotypes
Negative enrichment in msh3Δ genotypes
C>A
Positive enrichment in rnr1Y285A genotypes with msh2Δ or msh6Δ in combination
Negative enrichment in msh3Δ genotypes
C>A
C>T
Positive enrichment in rnr1Y285A msh2Δ
C>A
Positive enrichment in rnr1Y285A msh2Δ and rnr1Y285A msh6Δ

Figure 6.

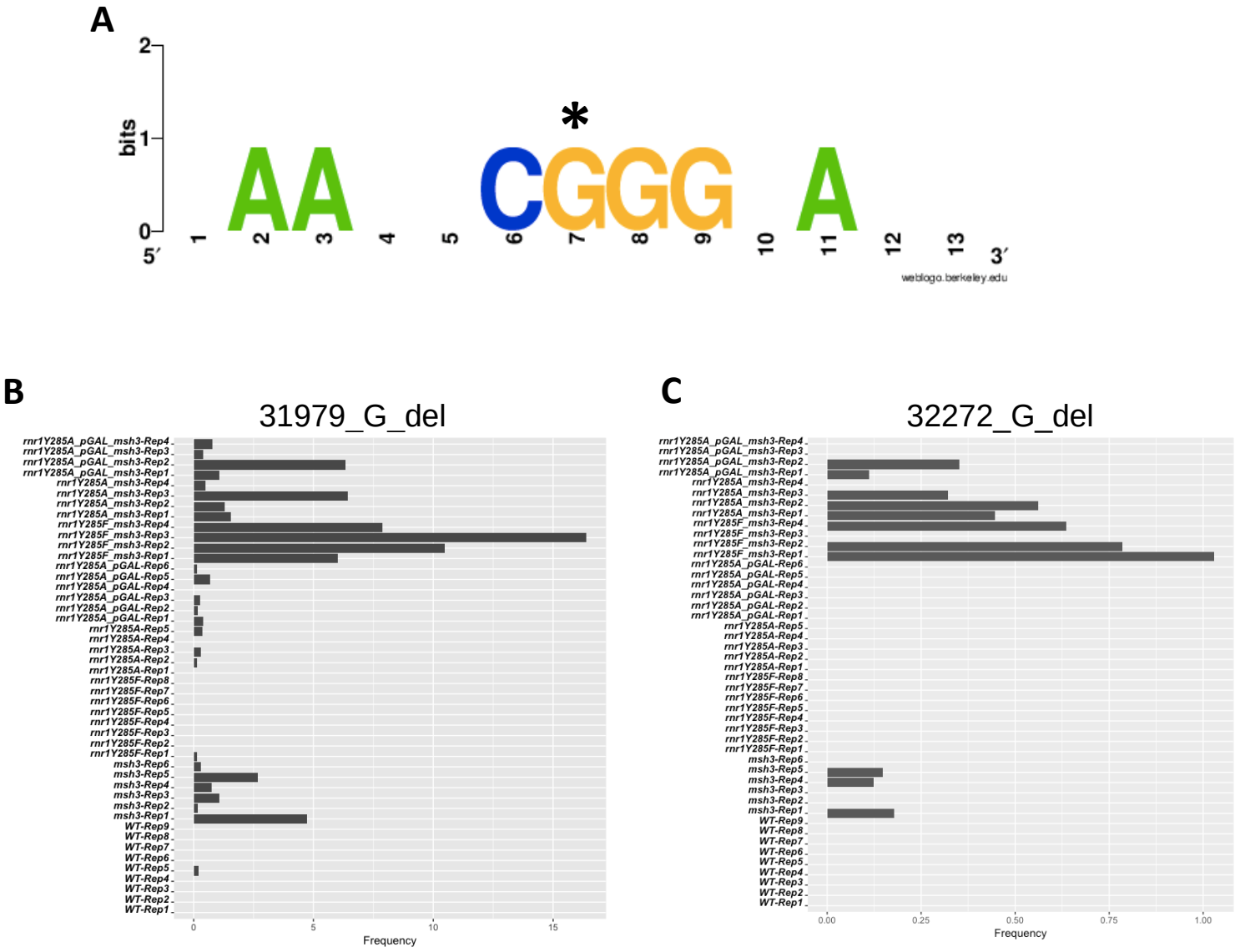

A
bits
2
1
0
5′
1
2
3
4
5
6
7
8
9
10
11
12
13
3′
*
AA
CGGG
A
weblogo.berkeley.edu
B
31979_G_del
rnr1Y285A_pGAL_msh3-Rep4
rnr1Y285A_pGAL_msh3-Rep3
rnr1Y285A_pGAL_msh3-Rep2
rnr1Y285A_pGAL_msh3-Rep1
rnr1Y285A_msh3-Rep4
rnr1Y285A_msh3-Rep3
rnr1Y285A_msh3-Rep2
rnr1Y285A_msh3-Rep1
rnr1Y285F_msh3-Rep4
rnr1Y285F_msh3-Rep3
rnr1Y285F_msh3-Rep2
rnr1Y285F_msh3-Rep1
rnr1Y285A_pGAL-Rep6
rnr1Y285A_pGAL-Rep5
rnr1Y285A_pGAL-Rep4
rnr1Y285A_pGAL-Rep3
rnr1Y285A_pGAL-Rep2
rnr1Y285A_pGAL-Rep1
rnr1Y285A-Rep5
rnr1Y285A-Rep4
rnr1Y285A-Rep3
rnr1Y285A-Rep2
rnr1Y285A-Rep1
rnr1Y285F-Rep8
rnr1Y285F-Rep7
rnr1Y285F-Rep6
rnr1Y285F-Rep5
rnr1Y285F-Rep4
rnr1Y285F-Rep3
rnr1Y285F-Rep2
rnr1Y285F-Rep1
msh3-Rep6
msh3-Rep5
msh3-Rep4
msh3-Rep3
msh3-Rep2
msh3-Rep1
WT-Rep9
WT-Rep8
WT-Rep7
WT-Rep6
WT-Rep5
WT-Rep4
WT-Rep3
WT-Rep2
WT-Rep1
0
5
10
15
Frequency
C
32272_G_del
rnr1Y285A_pGAL_msh3-Rep4
rnr1Y285A_pGAL_msh3-Rep3
rnr1Y285A_pGAL_msh3-Rep2
rnr1Y285A_pGAL_msh3-Rep1
rnr1Y285A_msh3-Rep4
rnr1Y285A_msh3-Rep3
rnr1Y285A_msh3-Rep2
rnr1Y285A_msh3-Rep1
rnr1Y285F_msh3-Rep4
rnr1Y285F_msh3-Rep3
rnr1Y285F_msh3-Rep2
rnr1Y285F_msh3-Rep1
rnr1Y285A_pGAL-Rep6
rnr1Y285A_pGAL-Rep5
rnr1Y285A_pGAL-Rep4
rnr1Y285A_pGAL-Rep3
rnr1Y285A_pGAL-Rep2
rnr1Y285A_pGAL-Rep1
rnr1Y285A-Rep5
rnr1Y285A-Rep4
rnr1Y285A-Rep3
rnr1Y285A-Rep2
rnr1Y285A-Rep1
rnr1Y285F-Rep8
rnr1Y285F-Rep7
rnr1Y285F-Rep6
rnr1Y285F-Rep5
rnr1Y285F-Rep4
rnr1Y285F-Rep3
rnr1Y285F-Rep2
rnr1Y285F-Rep1
msh3-Rep6
msh3-Rep5
msh3-Rep4
msh3-Rep3
msh3-Rep2
msh3-Rep1
WT-Rep9
WT-Rep8
WT-Rep7
WT-Rep6
WT-Rep5
WT-Rep4
WT-Rep3
WT-Rep2
WT-Rep1
0.00
0.25
0.50
0.75
1.00
Frequency

Figure 7.

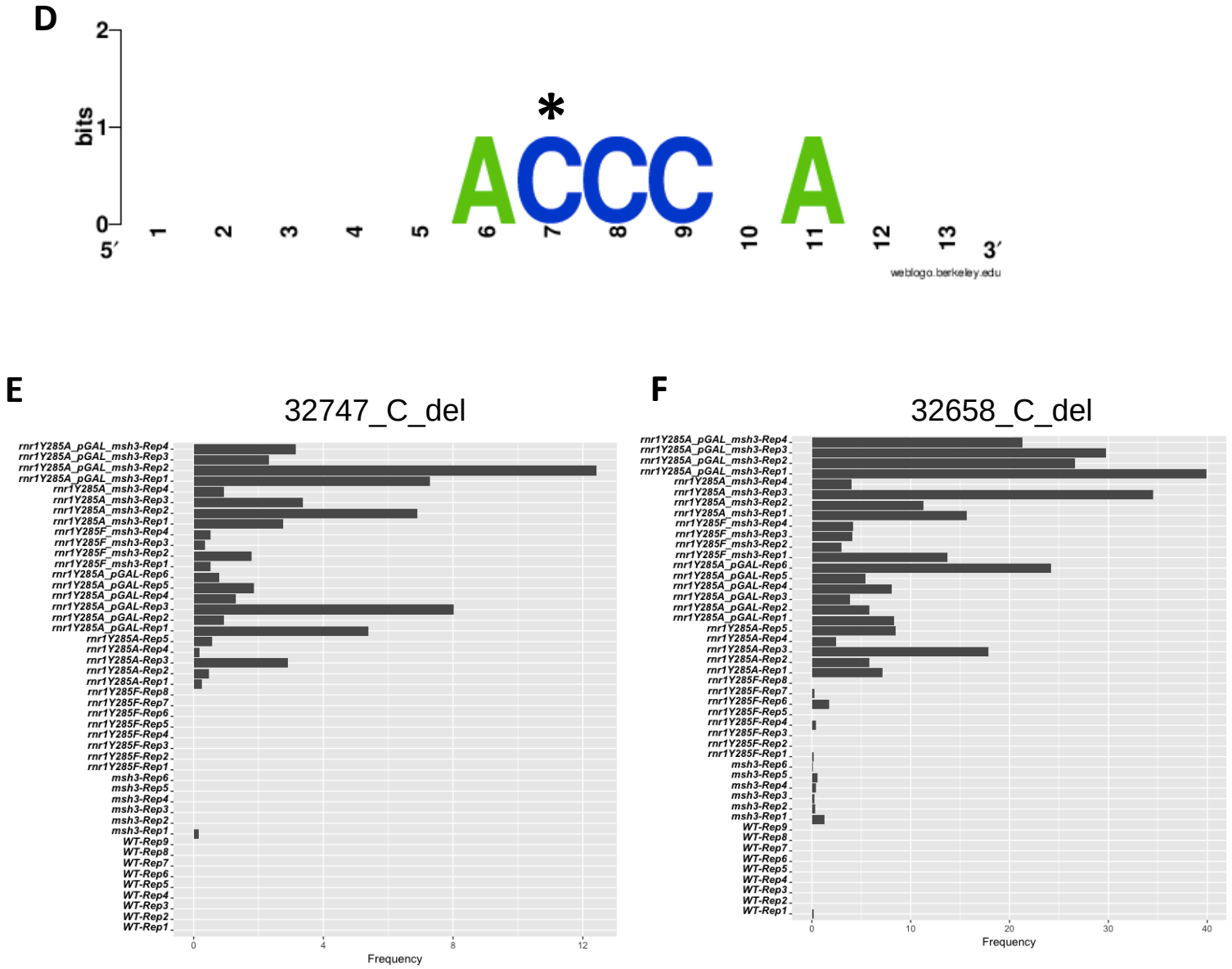


Figure 7. (continued)

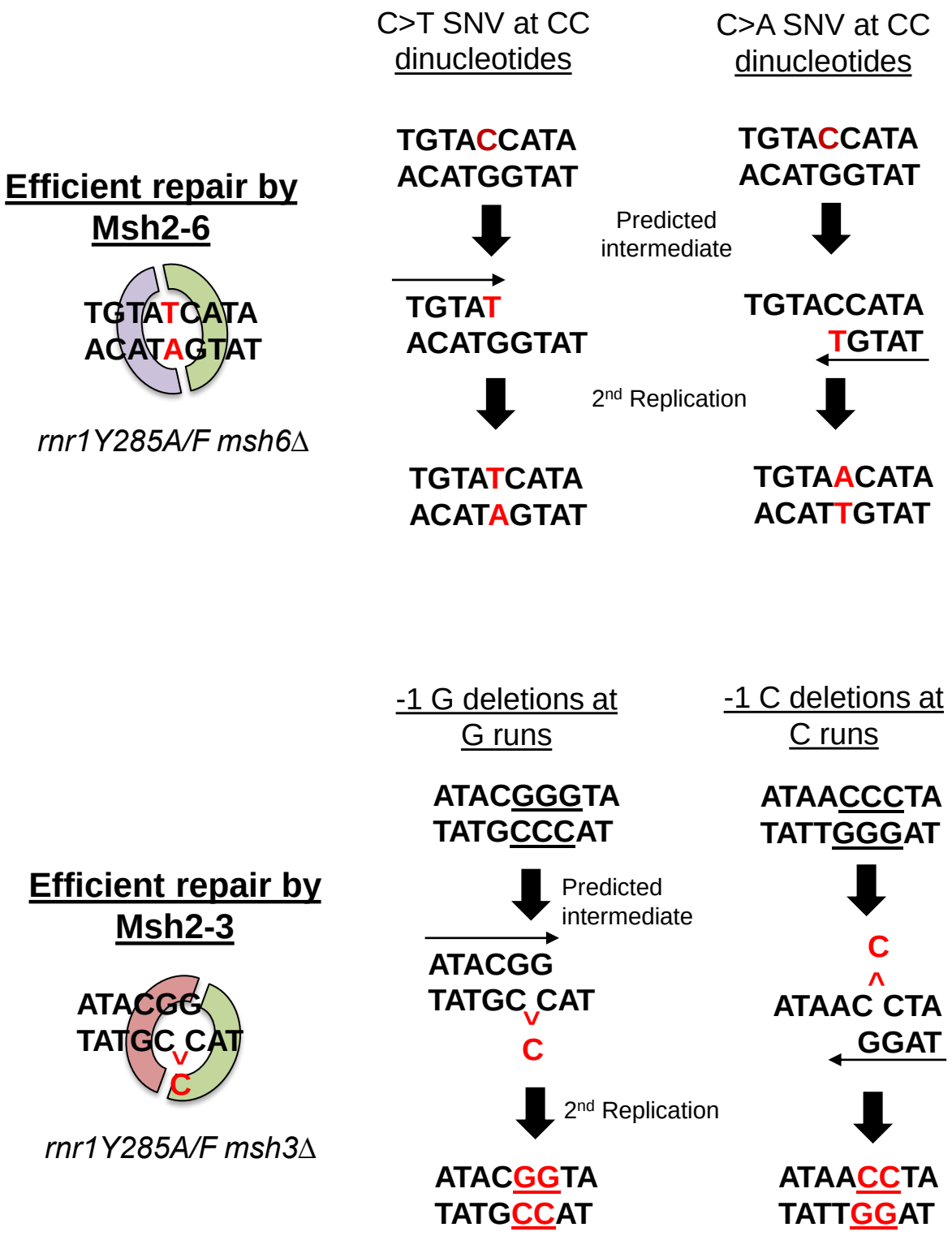

A
C>T SNV at CC dinucleotides
C>A SNV at CC dinucleotides
Efficient repair by Msh2-6
TGTATCATA
ACATAGTAT
rnr1Y285A/F msh6Δ
TGTACCATA
ACATGGTAT
TGTACCATA
ACATGGTAT
Predicted intermediate
TGTAT
ACATGGTAT
TGTACCATA
TGTAT
2nd Replication
TGTATCATA
ACATAGTAT
TGTAACATA
ACATTGTAT
B
-1 G deletions at G runs
-1 C deletions at C runs
Efficient repair by Msh2-3
ATACGG
TATGC CAT
C
rnr1Y285A/F msh3Δ
ATACGGGTA
TATGCCCAT
ATAACCCTA
TATTGGGAT
Predicted intermediate
ATACGG
TATGC CAT
C
C
ATAAC CTA
GGAT
2nd Replication
ATACGGTA
TATGCCAT
ATAACCTA
TATTGGAT

Figure 8.