Human mismatch repair system corrects errors produced during lagging strand replication more effectively

Maria A. Andrianova^{1,2,3*}, Georgii A. Bazykin^{1,2,3,4}, Sergey I. Nikolaev⁵ and Vladimir B. Seplyarskiy^{1,3*}

¹Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoi Karetny pereulok 19, Moscow 127994, Russia

²Moscow State University, Leninskie gory 1, Moscow 119234, Russia

³Pirogov Russian National Research Medical University, Ostrovitianov str. 1, Moscow 117997, Russia

⁴Skolkovo Institute of Science and Technology, Skolkovo 143026, Russia

⁵Department of Genetic Medicine and Development, University of Geneva Medical School, 1 Rue Michel-Servet, 1211 Geneva, Switzerland; Institute of Genetics and Genomics in Geneva, 1211 Geneva, Switzerland

*Correspondence to: MAA (baranova.mariia@gmail.com), VBS (alicodendrochit@gmail.com)

Abstract

Mismatch repair (MMR) is one of the main systems maintaining fidelity of replication. Different effectiveness in correction of errors produced during replication of the leading and the lagging DNA strands was reported in yeast, but this effect is poorly studied in humans. Here, we use MMR-deficient (MSI) and MMR-proficient (MSS) cancer samples to investigate properties of the human MMR. MSI, but not MSS, cancers demonstrate unequal mutation rates between the leading and the lagging strands. The direction of strand asymmetry in MSI cancers matches that observed in cancers with mutated exonuclease domain of polymerase δ , indicating that polymerase δ contributes more mutations than its leading-strand counterpart, polymerase ϵ . As polymerase δ primarily synthesizes DNA during the lagging strand replication, this implies that mutations produced in wild type cells during lagging strand replication are repaired by the MMR ~ 3 times more effectively, compared to those produced on the leading strand.

Results

Replication is a very accurate process. Its fidelity is achieved through three main components: base selectivity of polymerases, proofreading activity of their exonuclease domains, and repair of mismatches that escaped proofreading by mismatch repair (MMR) system¹. Studies in yeast indicate that the effectiveness of each of these steps depends on the mismatch type, and that MMR compensates for the infidelity of polymerases^{2,3}. According to the current model of the replication fork in a eukaryotic cell, polymerase ϵ preferentially copies the leading strand template, while polymerases α and δ copy the lagging strand template⁴⁻⁸. In yeasts, different replicative polymerases possess different biases in the types of mutations they introduce, leading to differences in mismatch types between the leading and the lagging strands². These biases can be further modulated by differences in MMR effectiveness between the two strands. For example, in yeast, the 8-oxo-guanine-adenine mismatches generated during the lagging strand

replication were shown to be repaired more effectively, compared to those generated during the leading strand replication⁹. More generally, in yeast, the biases between leading and lagging strands introduced by the MMR during repair tend to be opposite to the biases introduced during replication, thus compensating for them^{10,11}.

We first asked which types of substitutions are corrected by the MMR more effectively in humans. To address this, we compared the mutational spectra in MSI and MSS cancers of two different cancer types – colon adenocarcinoma (COAD) and uterine corpus endometrial carcinoma (UCEC) (Fig. 1a, b). In line with previous studies^{12–14}, we found that MSI cancers are enriched in C to T (C→T) mutations in the GpCpN context, A→G mutations in all contexts, and C→A mutations in the CpCpN context (Fig. 1a, c). Therefore, these mutation types are effectively corrected by the MMR.

The mutational spectra of the two cancer types are somewhat distinct. To address these differences, we calculated the weights of mutational signatures¹³ in the spectrum of each sample. In all MSI samples, we observed signatures of defective DNA mismatch repair (Signatures 6, 15, 20, 26). In many MSI samples of UCEC, we also observed clock-like signature 5 and other mutational signatures not related to MMR deficiency (Supplementary Table 1). Therefore, the mutational patterns in UCEC MSI cancers are slightly confounded by MMR-independent mutational processes. Among the MSS samples, a strong signature of the TpCpN→G and TpCpN→T mutations (Signatures 2 and 13) is observed in UCEC but not in COAD samples; otherwise, the MSS spectra of the two cancer types look very similar. Signatures 2 and 13 are mutational signatures of the APOBEC activity¹³. To eliminate these differences in MSS spectra, we excluded samples with these signatures and high APOBEC enrichment values (Supplementary Table 2) from analysis (Fig. 1a, c).

Using the approach that determines the replication fork direction from data on replication timing^{15–17}, we compared the rates of complementary mutations between the leading and the lagging strands for germline mutations inferred from human-chimpanzee comparison, and for somatic mutations in MSS cancers and MSI cancers (Fig. 2). In line with previous results¹⁶, we found replication-driven asymmetry in MSI cancers. In general, we observed that the ratio of the rates of complementary mutations increased with the propensity of a DNA strand segment to be replicated as leading or lagging, indicating that mutation rates are asymmetric between the leading and the lagging strands. Separate analyses of UCEC and COAD cancer types provide concordant results (Supplementary Table 3), and small quantitative differences likely reflect admixture of MMR-independent mutational processes in these cancers (Supplementary Table 1). The asymmetry is very similar if we analyze only intergenic regions (Supplementary Fig. 1), suggesting that these results are not associated with transcription. We studied this asymmetry in the most extreme replication fork direction bin, because it corresponds to the genomic regions where we could predict fork polarity with the highest confidence¹⁷.

In human-chimpanzee divergence, complementary mutations are distributed nearly symmetrically between the two replicating strands, in line with previous results¹⁸. Lack of asymmetry in divergence data and in MSS cancers reflects a near lack of mutation biases between the two strands in MMR-proficient cells. By contrast, in MSI cancers where the MMR system is not proficient, we observe a strongly biased distribution of complementary mutations between the leading and the lagging DNA strands. A particularly strong asymmetry (1.5-1.8-fold) is observed for mutations effectively repaired by the MMR ($A \rightarrow G$, $CpCpN \rightarrow A$ and $GpCpN \rightarrow T$). The asymmetry is small or nearly absent for mutation types depleted in mutational spectra of MSI cancers ($A \rightarrow T$, $C \rightarrow G$, $DpCpN \rightarrow A$, $HpCpN \rightarrow T$ mutations), to the exception of the $A \rightarrow C$ mutation, where the asymmetry is 1.5-fold. In MMR-proficient cells, the observed biases have to be compensated by the MMR, which has to correct errors on one of the two

strands more effectively, thus equalizing the mutation rate between strands. Moreover, to compensate for the substitution biases, the MMR has to introduce its own repair biases.

To further validate this, we reproduced our results using exome sequences of MSI and MSS cancers from the TCGA (Supplementary Fig. 2). We also analyzed exomes and genomes separately by cancer type: COAD, stomach adenocarcinoma (STAD) and UCEC. The biases observed in this bigger set of cancer types were similar and concordant between cancer types, implying that they are determined not by the cancer type but by the (in)activity of the MMR system (Supplementary Table 3).

MMR corrects single-nucleotide insertions and deletions (indels) with the highest effectiveness among all single nucleotide errors. In MMR-deficient cells, indels in homopolymer tracts and dinucleotide tandem repeats are the most frequent type of mutations. As in yeast ¹⁹, in humans deletions are more common than insertions in MMR deficient cells, and their rate increases with the homopolymer tract length. We studied the most frequent deletions type, deletions of A or T in corresponding homopolymer tracts. Similarly to single-nucleotide substitutions, the efficiency of MMR in correction of deletions in homopolymer tracts may differ between the two DNA strands. We observe a small asymmetry for deletions (Supplementary Fig. 3), supporting the conjecture that the MMR activity differs between the two strands. This asymmetry has the same direction in the two cancer types, but is slightly higher in the COAD than in the UCEC.

The asymmetry in mutation rates between the leading and the lagging DNA strands is probably due to replication of these strands by different polymerases. As MMR is primarily a co-replicative process ^{20,21}, we hypothesized that the observed asymmetry in MMR-deficient cancers is due to differences in types and rates of errors produced by replicative polymerases on the leading and on the lagging strands. To investigate the propensity of polymerases to particular mismatch types in humans, we employed data from patients with mutations in the exonuclease

domain of one of the two major replicative polymerases, polymerase ϵ (pol ϵ -exo⁻) or polymerase δ (pol δ -exo⁻), because the fidelity of the damaged polymerase is decreased by a factor of 100-1000, and most mutations are produced by it ^{2,22}. In particular, such mutations are frequent in cancers with inherited biallelic mismatch repair deficiency (bMMRD). We observe that the relative frequencies of different nucleotide mismatch types are radically distorted in these cancers (Fig. 3a, b). Assuming that disruption of the exonuclease domain does not affect the direction of the strand asymmetry of mutations, we can interpret these differences as mutational signatures of the corresponding replicative polymerases. We find that each polymerase usually produces one of the two complementary mutations with a higher frequency (Fig. 3 c). In particular, the biases associated with the C→A and T→G mutations are opposite: polymerase ϵ preferentially produces mismatches resulting in these mutations on the leading strand, while polymerase δ produces them on the lagging strand (Fig. 3c), in line with the observations in yeast ²³⁻²⁵. Whole genomes and whole exomes data for pol ϵ -exo⁻ and pol δ -exo⁻ UCEC and COAD cancers and bMMRD glioblastoma cancers demonstrate similar asymmetry patterns (Fig. 3, Supplementary Fig. 4, Supplementary Fig. 5, and Supplementary Table 4), which shows that this asymmetry is specific to the mutational processes rather than the cancer type.

Notably, the direction of the mutational strand asymmetry in pol δ -exo⁻ cancers generally matches that in MSI cancers with wild type polymerases (Fig. 4). We designed a model using the asymmetry observed in pol ϵ -exo⁻ and pol δ -exo⁻ cancers to estimate the contributions of pol ϵ and pol δ to mutagenesis in MSI cancers (see Methods). This model assumes that (i) all mutations in pol ϵ -exo⁻ (pol δ -exo⁻) cancers arise from the action of pol ϵ (pol δ) on the leading (respectively, lagging) strand; (ii) each of these polymerases independently contributes some fraction of mutations; and (iii) the direction of the replication fork in the 20% of the genome where replication asymmetry could be determined with the highest confidence is known exactly. We applied this model to mutation types presented in MSI spectra at high frequency (A→G, C→T,

C→A). We found that the contribution of pol δ to the asymmetry observed in MSI cancers is 1.9 to 4.3 fold higher than the contribution of pol ϵ (Supplementary Table 5). Accounting for the frequencies of the corresponding mutations, the overall contribution of pol δ to mutagenesis was ~3-fold greater than that of pol ϵ (Fig. 4b). An independent dataset of 12 MSI COAD cancers²⁶ yielded a 2-fold greater contribution of pol δ (data not shown). This implies that the main driver of the asymmetry in MSI cancers are the mutations introduced by pol δ during the replication of the lagging strand.

Our analysis is subject to several caveats. Firstly, we assumed that the mutational spectra of polymerases with disrupted exonuclease domain match those of wild type polymerases. This is likely to be true: the magnitude of the strand bias is huge in yeast^{23–25} and seems to be large in humans⁸, and is unlikely to be radically affected by the presence of exonuclease activity. Secondly, our approach allows predicting only the preferential direction of the replication fork; therefore, the magnitude of asymmetry may be underestimated, especially if it is very strong¹⁷. These caveats, however, cannot affect our conclusions qualitatively.

As no strand asymmetry is observed in wild type cells, the strand biases introduced by polymerases have to be compensated by the MMR. Since the mutations are biased towards the lagging strand, this implies that MMR is proportionally more effective in repairing mutations on the lagging strand.

Mutational spectra differed between the MMR-efficient and MMR-deficient cancers with somatic mutations in pol ϵ -exo or pol δ -exo (Supplementary Fig. 6). However, the level of replication asymmetry was similar among all pol ϵ -exo⁻ or pol δ -exo⁻ cancers (Supplementary Table 4). Therefore, while the MMR may change the mutational spectra, it is insufficient to compensate for the radical asymmetries introduced by pol ϵ -exo⁻ and pol δ -exo⁻.

In summary, we show that the polymerase error rate is higher during lagging strand replication, that this asymmetry is primarily due to mutations produced by pol δ on the lagging strand, and that these mismatches are removed more effectively by the MMR on the lagging strand (Fig. 5). This is in agreement with the biochemical property of the MMR to preferentially eliminate mismatched nucleotides on the DNA strand containing the nick ²⁷. As the lagging strand is replicated in Okazaki fragments, the MMR has more time to discriminate between the template and nascent strands during its replication. More generally, the remarkable concordance between the biases in introduction and repair of mismatches helps reduce the genomic mutation rate.

Methods

Mutations Data. We used the following previously published data: (a) somatic mutations for whole-exome sequences of MSI (n = 159) and MSS (n = 782) cancers and for whole genome sequences of MSI (n=11) and MSS (n=27) cancers from the data portal of The Cancer Genome Atlas (TCGA) (<https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm>); (b) bam files with aligned reads for ultra-hypermuted cancers ²⁸; (c) human-chimpanzee-orangutan multiple alignment from the UCSC genome browser (<https://genome.ucsc.edu/>). Substitution rates were calculated as the number of substitutions of a particular type divided by the number of target sites. For asymmetry analyses of the TCGA data, cancers with mutations in replicative polymerases were excluded. For analysis of the interspecies data, we obtained mutations in the human line after its divergence from the chimpanzee by maximum parsimony, using orangutan as the outgroup.

Identification of somatic mutations in ultra-hypermuted cancers. Somatic mutations were identified using MuTect (v. 1.1.4) ²⁹ under the default parameters. Mutations were then filtered against common single-nucleotide polymorphisms (SNPs) found in dbSNP and against the Catalogue of somatic mutations in cancer (Cosmic database).

Leading vs. lagging strand asymmetry. The derivative of the replication timing at the position of the mutation was used as a proxy for the probability that the reference strand is replicated as leading or lagging in the current position, as described in ref. ¹⁷. The genome was categorized by these values into 10 equal bins, with the low value of the derivative corresponding to the propensity of the DNA segment to be replicated as lagging, and high value, as leading. For each bin, the numbers of substitutions and target sites were calculated. Each substitution was counted twice: as a substitution on the reference sequence with the corresponding derivative of the replication timing, and as a complementary substitution with the inverse derivative. Thus, each plot of substitutions asymmetry (Fig. 2 and 3c) is symmetric with respect to zero. Score confidence intervals were obtained for the relative risk in a 2x2 table.

Indels analyses. We used data on single nucleotide deletions in poly-A and poly-T tracts for MSI cancer genomes, for tracts with length six to eight identical nucleotides where enough deletions were found for analysis.

APOBEC enrichment. APOBEC enrichment was counted for each sample as ratio of C→K mutation rates in TpCpW and VpCpW contexts. Weights of mutational signature were calculated using R-package "deconstructSigs" ³⁰.

Model for mutational biases between strands. We calculated the ratios of the mutation rates using the following logic. From Fig. 4, for each type of mutation A→B, we obtained the ratio of its rate $r(A \rightarrow B)$ and the rate of its complement $r(A' \rightarrow B')$ on the leading strand. Then

$$\frac{x_{MMR}}{(1-x_{MMR})} = \frac{x_{\epsilon}\alpha + (1-x_{\delta})(1-\alpha)}{(1-x_{\epsilon})\alpha + x_{\delta}(1-\alpha)} \quad (1)$$

where α is the fraction of mutations A→B and A'→B' that are produced by polε on the leading strand; 1-α is the fraction of such mutations that are produced by polδ on the lagging strand; x_{ϵ} and 1- x_{ϵ} are the fractions of mutations A→B and A'→B' respectively produced by polε; x_{δ} and 1- x_{δ} are the fractions of mutations A→B and A'→B' respectively produced by polδ; and x_{MMR} and 1- x_{MMR} are the fractions of mutations A→B and A'→B' respectively on the leading strand in MSI cancers.

For example, consider mutation C→A/G→T. From Fig. 4, $x_{MMR}/(1-x_{MMR})$ is 0.65. As the C→A/G→T ratio for the leading strand in polε-exo⁻ cancers is 1.96 (Fig. 4), the fraction of C→A mutations produced by polε is $x_{\epsilon} = 0.66$, and the fraction of G→T mutations produced by polε is 1- $x_{\epsilon} = 0.34$. Similarly, as the C→A/G→T ratio for the lagging strand in polδ-exo⁻ cancers is 2.07, the fraction of C→A mutations produced by polδ is $x_{\delta} = 0.67$, and the fraction of G→T mutations produced by polδ is 1- $x_{\delta} = 0.33$. From equation 1, $(1-\alpha)/\alpha = 4.0$. In other words, for this mutation type, polδ produces mismatches leading to this mutation type on the lagging strand ~4 times more often than polε produces them on the leading strand. As no strand bias is observed in MSS cancers, the repair bias by MMR has to be exactly inverse, repairing polδ mutations 4 times as effectively as polε mutations. The results for all 5 mutation types are calculated similarly (Table S3). The overall strand asymmetry of MMR was calculated as the mean asymmetry across the five mutation types, weighted by the proportions of each mutation in MSI cancers.

References and Notes:

1. Kunkel, T. A. Evolving views of DNA replication (in)fidelity. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 91–101 (2009).
2. St Charles, J. A., Liberti, S. E., Williams, J. S., Lujan, S. A. & Kunkel, T. A. Quantifying the contributions of base selectivity, proofreading and mismatch repair to nuclear DNA replication in *Saccharomyces cerevisiae*. *DNA Repair* **31**, 41–51 (2015).
3. Kunkel, T. A. Balancing eukaryotic replication asymmetry with replication fidelity. *Curr. Opin. Chem. Biol.* **15**, 620–626 (2011).
4. Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M. J. & Kunkel, T. A. Division of labor at the eukaryotic replication fork. *Mol. Cell* **30**, 137–144 (2008).
5. Larrea, A. A. *et al.* Genome-wide model for the normal eukaryotic DNA replication fork. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17674–17679 (2010).
6. Pursell, Z. F., Isoz, I., Lundström, E.-B., Johansson, E. & Kunkel, T. A. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* **317**, 127–130 (2007).
7. Georgescu, R. E. *et al.* Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* **21**, 664–670 (2014).
8. Shinbrot, E. *et al.* Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res.* **24**, 1740–1750 (2014).
9. Pavlov, Y. I., Mian, I. M. & Kunkel, T. A. Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. *Curr. Biol. CB* **13**, 744–748 (2003).
10. Lujan, S. A. *et al.* Mismatch repair balances leading and lagging strand DNA replication fidelity. *PLoS Genet.* **8**, e1003016 (2012).

11. Kunkel, T. A. & Erie, D. A. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu. Rev. Genet.* **49**, 291–313 (2015).
12. Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81–84 (2015).
13. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
14. Zhao, H. *et al.* Mismatch repair deficiency endows tumors with a unique mutation signature and sensitivity to DNA double-strand breaks. *eLife* **3**, e02725 (2014).
15. Baker, A. *et al.* Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLoS Comput. Biol.* **8**, e1002443 (2012).
16. Haradhvala, N. J. *et al.* Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell* **164**, 538–549 (2016).
17. Seplyarskiy, V. B. *et al.* APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Res.* **26**, 174–182 (2016).
18. Chen, C.-L. *et al.* Replication-associated mutational asymmetry in the human genome. *Mol. Biol. Evol.* **28**, 2327–2337 (2011).
19. Lujan, S. A., Clark, A. B. & Kunkel, T. A. Differences in genome-wide repeat sequence instability conferred by proofreading and mismatch repair defects. *Nucleic Acids Res.* **43**, 4067–4074 (2015).
20. Hombauer, H., Srivatsan, A., Putnam, C. D. & Kolodner, R. D. Mismatch repair, but not heteroduplex rejection, is temporally coupled to DNA replication. *Science* **334**, 1713–1716 (2011).
21. Liao, Y., Schroeder, J. W., Gao, B., Simmons, L. A. & Biteen, J. S. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6898–6906 (2015).

22. Reha-Krantz, L. J. DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochim. Biophys. Acta* **1804**, 1049–1063 (2010).
23. Fortune, J. M. *et al.* *Saccharomyces cerevisiae* DNA polymerase delta: high fidelity for base substitutions but lower fidelity for single- and multi-base deletions. *J. Biol. Chem.* **280**, 29980–29987 (2005).
24. Shcherbakova, P. V. *et al.* Unique error signature of the four-subunit yeast DNA polymerase epsilon. *J. Biol. Chem.* **278**, 43770–43780 (2003).
25. Lujan, S. A. *et al.* Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.* **24**, 1751–1764 (2014).
26. Katainen, R. *et al.* CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* **47**, 818–821 (2015).
27. Pluciennik, A. *et al.* PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16066–16071 (2010).
28. Shlien, A. *et al.* Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermuted cancers. *Nat. Genet.* **47**, 257–262 (2015).
29. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
30. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, (2016).

Acknowledgements We thank Shamil Sunyaev, Pasha Mazin and Sonya Garushyants for useful discussion. This work was performed at IITP RAS and supported by the Russian Science Foundation grant no. 14-50-00150.

Author Contributions V.B.S. designed the study, M.A.A. performed analyses, M.A.A., G.A.B. and V.S.B. wrote the paper, S.I.N. provided data for cancer samples and assisted in interpretation of results and writing the paper.

The authors declare no competing financial interests.

Figure 1. Mutational spectrum of MMR deficient cancers. Relative frequencies of the 96 mutation types in MSI and MSS cancers. Complementary mutation types were pooled. **a**, MSI (n = 3) and MSS (n = 8) colon adenocarcinoma samples; **b**, MSI (n = 7) and MSS (n = 19) uterine corpus endometrial carcinoma; **c**, MSI (n = 7) and MSS (n = 14) uterine corpus endometrial carcinoma excluding APOBEC enriched samples.

Figure 2. Strand asymmetry of mutations. Panels show ratios of the mutation rates for complementary mutations in 11 MSI (3 COAD and 7 UCEC) and 22 MSS (8 COAD and 14 UCEC) cancers, and in human-chimpanzee divergence. Horizontal axis, propensity of the replication fork to replicate the strand as lagging or leading. Vertical axis, the ratio of the frequencies of the two complementary mutation types on the strand in this category. Vertical bars represent 95% confidence intervals. D corresponds to nucleotides A, G or T. H corresponds to nucleotides A, C or T. Note the logarithmic vertical axis.

Figure 3. Mutation patterns in bMMRD cancers with subsequent mutations in *polε* or *polδ*. Data for 7 exomes of bMMRD cancer (5 *polε*-exo⁻ and 2 *polδ*-exo⁻). **a**, Relative frequencies of single-nucleotide substitutions irrespective of the strand; **b**, Context-dependent mismatch frequency for each mutation type averaged between

patients in each group. Vertical axis, 5' context of the substitution; horizontal axis, 3' context. For each single-nucleotide mutation type, the color of each of the 16 cells indicates how frequently this mutation occurs in this context; **c**, The ratio of the mutation rates for complementary mutations as a function of the propensity of the replication fork to replicate the reference strand as lagging or leading. Axes and notations same as in Fig. 2.

Figure 4. Strand asymmetry of mutations in MSI cancers and in bMMRD pol ϵ -exo⁻ and pol δ -exo⁻ cancers. a, The asymmetry for complementary mutations in the 20% of the genome where replication asymmetry could be determined with the highest confidence (corresponding to bins 1 and 10 in Fig. 2). Error bars represent 95% confidence intervals. **b,** Model-estimated ratio of MMR effectiveness on the lagging and the leading strands.

Figure 5. The schematic representation of MMR effectiveness during the leading and the lagging strand replication. While mismatches (red asterisks) are introduced more frequently during replication of the lagging strand by polδ, the lagging strand mismatches are corrected more effectively by the MMR.