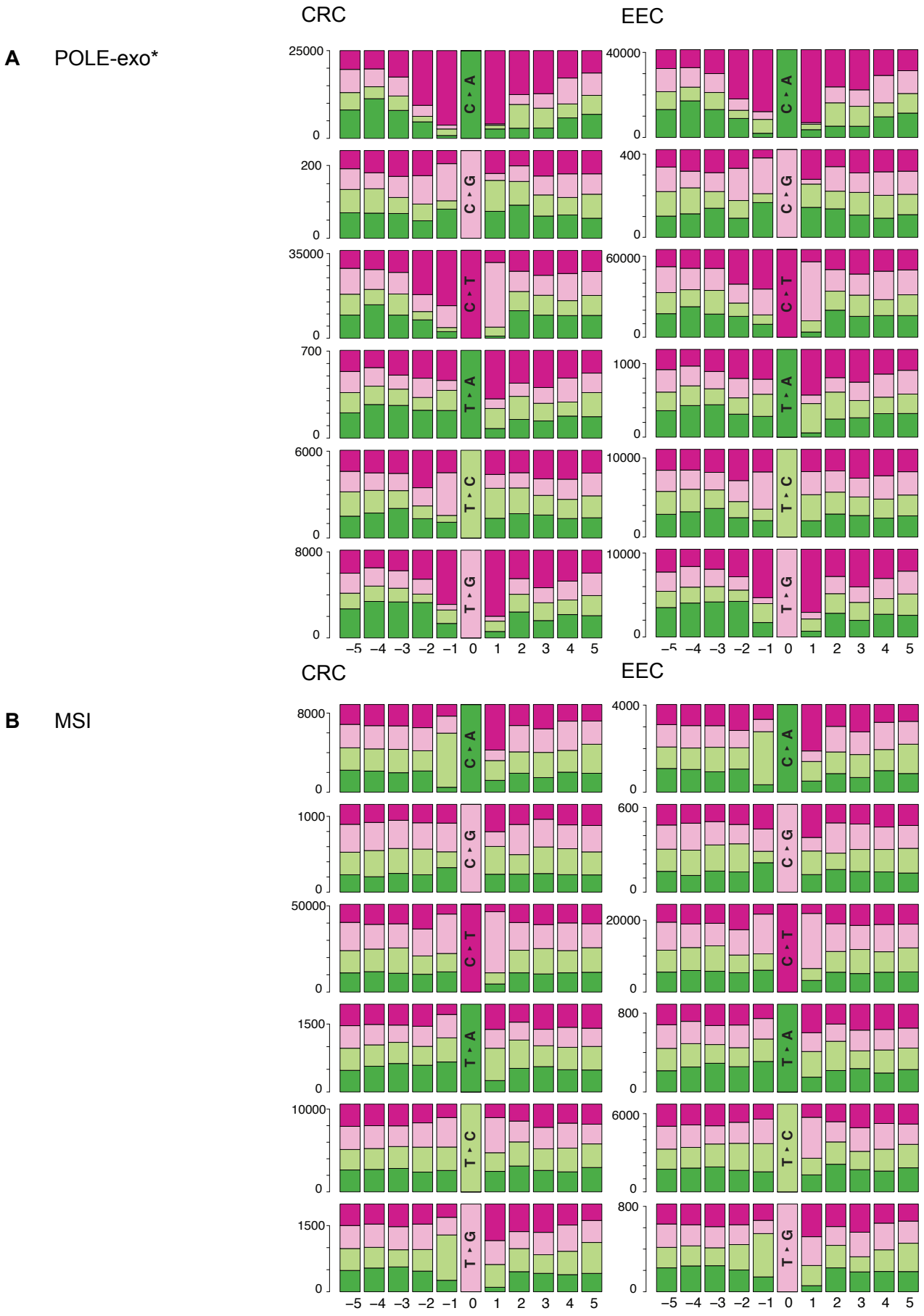
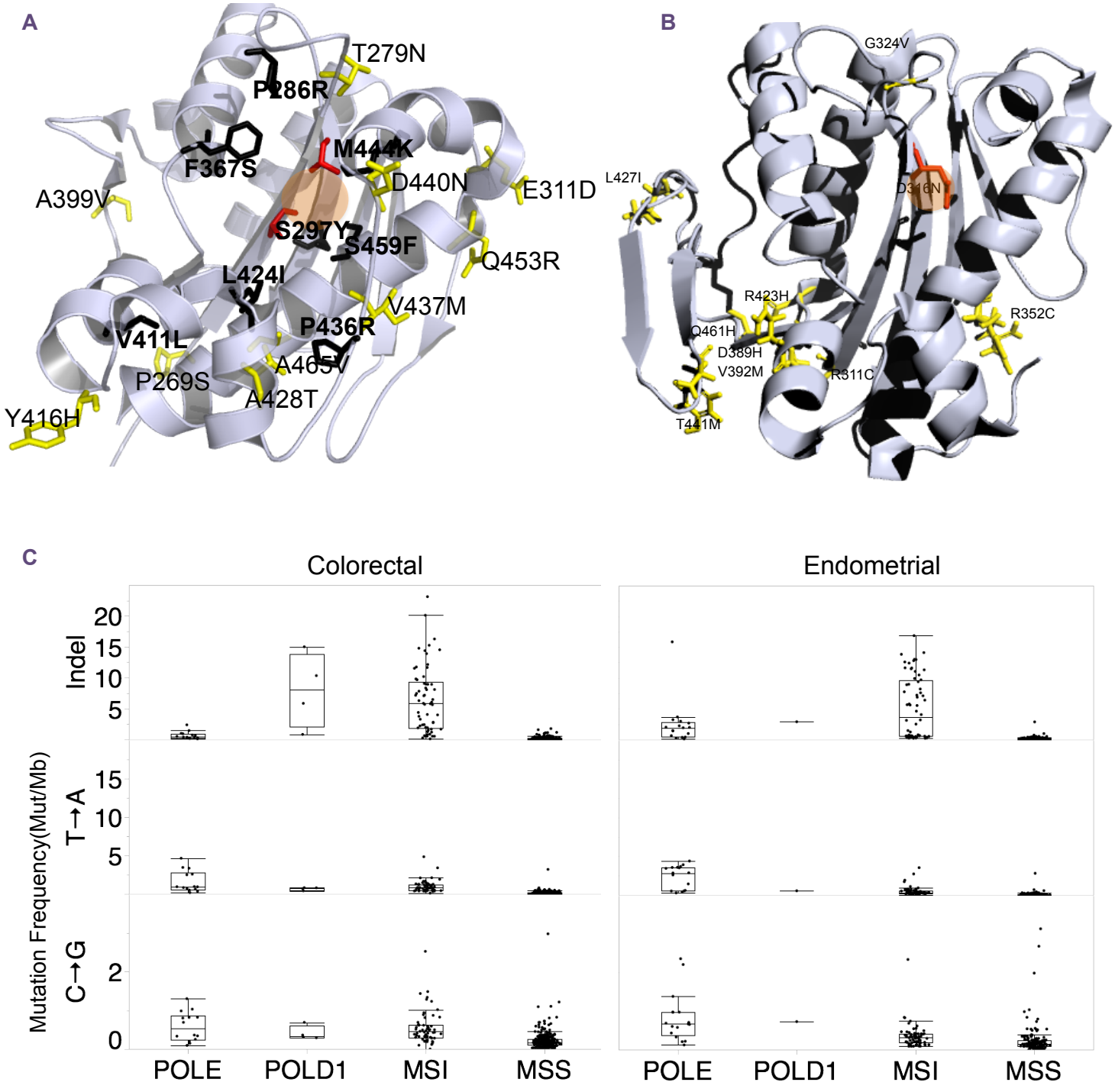
Figure S1



Supplemental Figure S1. Base context in which mutations occur in POLE-exo* in Colorectal and Endometrial tumors. (A) POLE-exo* GroupA **(B)** MSI tumors. CRC, Colorectal (left column, tallies from 14 patients); EEC, Endometrial (right column, tallies from 17 patients). The mutational base change is plotted as a stacked bar-graph, with the immediate mutational context 5 bases to either side. The central bar in each sub-panel indicates the mutation type. The 5 bars on either side show the relative frequencies of adjacent bases from -5 to +5. T, dark purple; A, dark green; C, light green; G, pink. The X-axis scale on each stacked bar-graph denotes to total number of mutations found in each patient group.

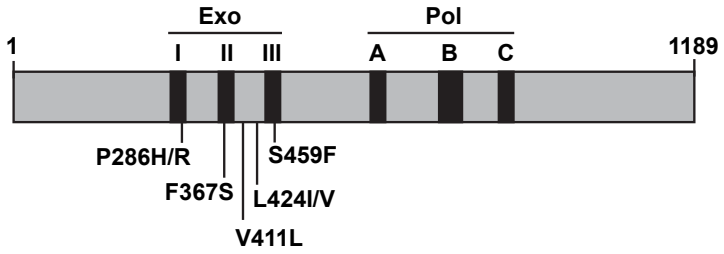
Figure S2


A
T279N
P286R
F367S
M444K
D440N
E311D
A399V
S297Y
S459F
Q453R
L424I
V437M
V411L
P436R
A465V
P269S
A428T
Y416H
B
G324V
L427I
D316N
R423H
R352C
Q461H
D389H
V392M
R311C
T441M
C
Colorectal
Endometrial
Indel
T→A
C→G
Mutation Frequency(Mut/Mb)
POLE
POLD1
MSI
MSS

Supplemental Figure S2. **Human DNA Pol exo mutations mapped to 3D structure of yeast Pol by homology.** All residues are labeled according to the human variant. Group A mutants black amino acid side chains with bold lettering, Group B mutants, yellow amino acid side chains, active site residues in red. **(A)** 3D structure of DNA polymerase epsilon from saccharomyces 3D (PDB accession 4M8O) (Hogg et al, 2014). **(B)** 3D structure of DNA polymerase delta from saccharomyces (PDB accession 3IAY) (Swan et al, 2009). The catalytic residue mutated in patient BR-6452 is black (TCGA gastric cancer, STAD, see Table 1B and Supplemental Table S1B).
(C) Absolute frequencies of indel and low-prevalence mutations in CRC and endometrial cancers. Upper panels, indel mutation frequencies are similar in POLE and MSS tumors and both are much lower than MSI tumors. Note, POLD1 tumors are similar to MSI tumors in distribution of absolute indel frequencies. Middle and lower panels show the lowest prevalence mutations in POLE (T->A and C->G, respectively, see Figure). The absolute frequencies of these mutations are similar across all groups. Compare with Figure 1, which displays relative frequencies for all mutation types, Figure 2A, which displays the highest prevelance mutations in POLE.

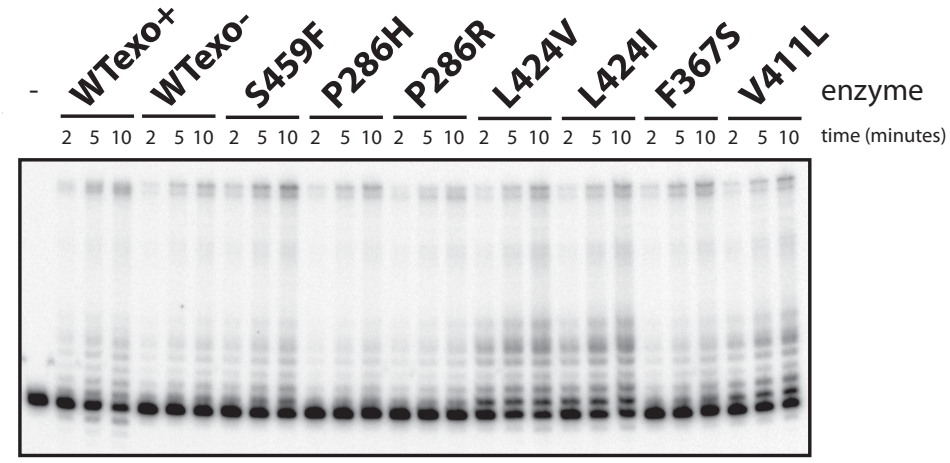
Figure S3

A



B

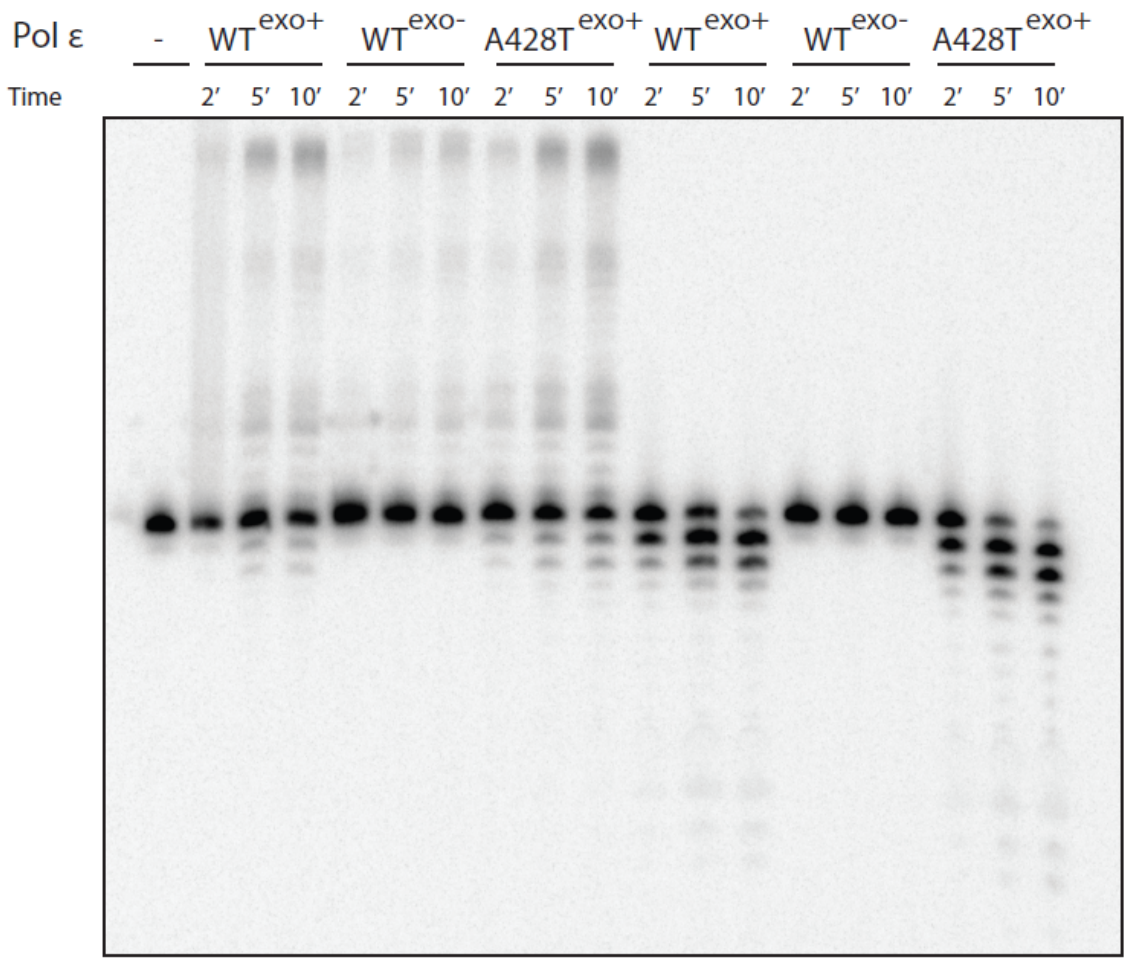
C



	Wild Type Exo+	D275A/ D277A	P286H	P286R	F367S	L424I	L424V	S459F
No. Plaques Counted	27,792	3482	6647	7690	5915	5713	4566	3370
No. Mutant Plaques	18	35	45	37	38	26	44	61
lacZ Mutant Frequency (x 10^{-4})	**6.5**	**100**	**68**	**48**	**64**	**46**	**96**	**180**
Fold increase over Wild Type Exo+ Pol ε	*1x*	*15x*	*10x*	*7.4x*	*9.8x*	*7.1x*	*15x*	*28x*
No. Mutant Plaques Sequenced	18	35	45	37	38	26	44	61
No. Mutations Sequenced	18	44	46	37	39	26	44	61

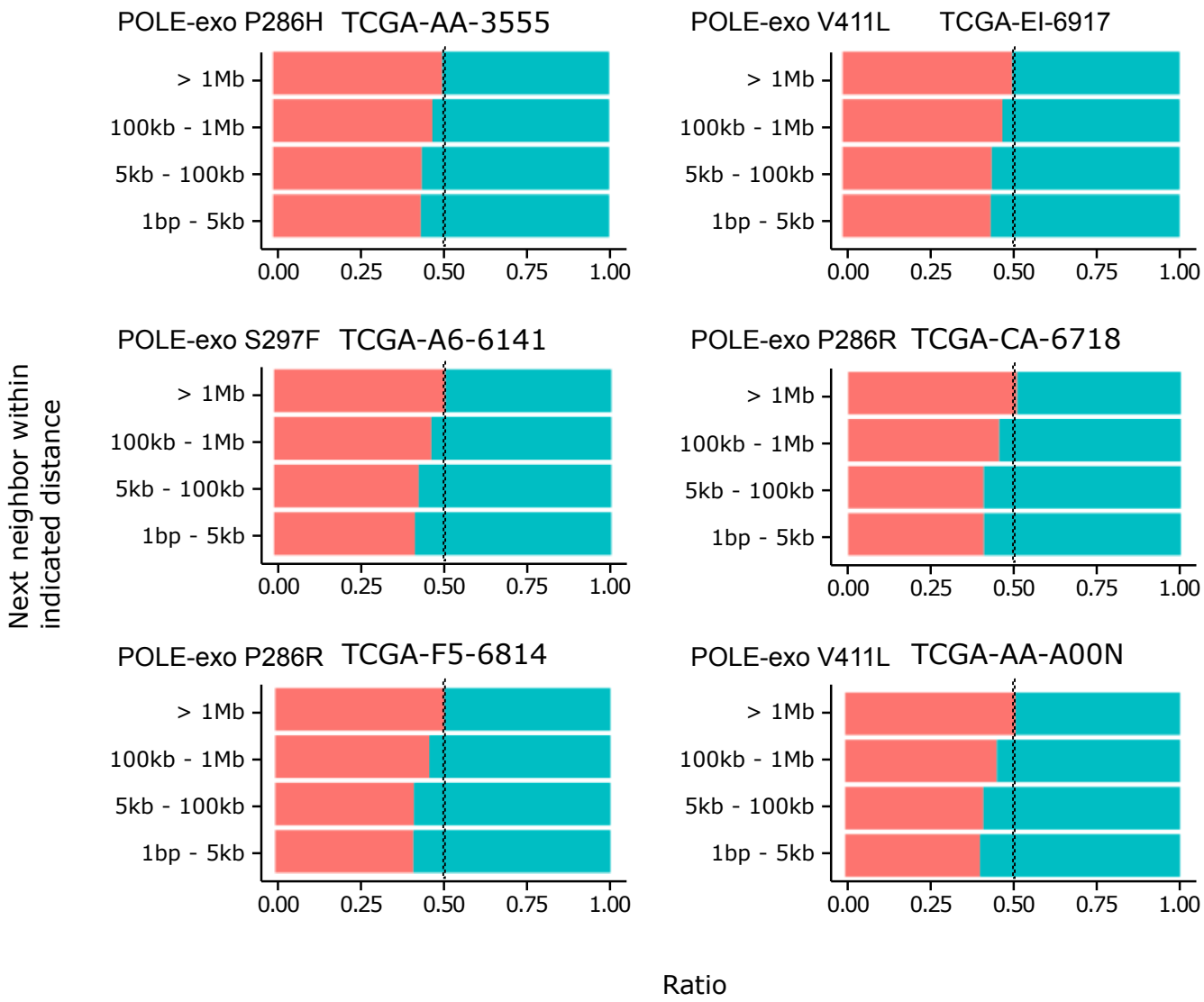
Supplemental Figure 3. Pol ε cancer mutations have little effect on DNA polymerization activity. (A) Schematic of the Pol ε construct used for *in vitro* assays in the current study. Locations of each cancer mutation with respect to the conserved exonuclease (Exo) and polymerase (Pol) motifs are shown. **(B)** The relative DNA polymerization activities of each Pol ε mutant allele were compared to those of wild type (WTexo +) and a previously defined exonuclease-deficient (D275A/E277A) mutant (WTexo-). Purified Pol ε constructs (1 nM) were incubated with all 4 dNTPs (25 μM each), 8 μM MgCl2 and a duplex DNA substrate containing 45-mer template oligonucleotide hybridized to a completely complementary 18-mer primer oligonucleotide. Reactions were carried out at 37°C and initiated by the addition of enzyme. Aliquots were removed at the indicated times and resolved by 12% denaturing PAGE. **(C)** *LacZ* mutant frequencies for the Pol ε cancer mutant alleles. The *lacZ* forward mutation assay was carried out as described in the Methods with each mutant construct at the same time in parallel. The indicated numbers of *lacZ* mutations were sequenced and used to calculate individual error rates.

Figure S4


Pol ε
-
WT exo+
WT exo-
A428T exo+
WT exo+
WT exo-
A428T exo+
Time
2′ 5′ 10′ 2′ 5′ 10′ 2′ 5′ 10′ 2′ 5′ 10′ 2′ 5′ 10′ 2′ 5′ 10′

Supplemental Figure 4. **Group B Pol ε cancer mutations do not inhibit 3'→5' exonuclease activity.** The relative polymerase and 3'-5' exonuclease activities of the A428T variant was compared to that of wild type (WT^{exo+}) and exonuclease deficient enzyme (WT^{exo-}; D275A/E277A). For primer extension reactions (left), purified mutant Pol ε constructs (1 nM) were incubated with all 4 dNTPs (25 μM each), 8 mM $MgCl_2$ and a 20-mer primer/45-mer template duplex DNA substrate, with the primer fully complementary to the template. For excision reactions (right), reactions conditions were identical, except for the following changes. Each enzyme was used at 2 nM, the substrate contained a terminal C•T mismatch and dNTPs were withheld from the reaction. Reactions were carried out at 37°C and initiated by the addition of enzyme. Aliquots were removed at the indicated times and resolved by 12% denaturing PAGE.

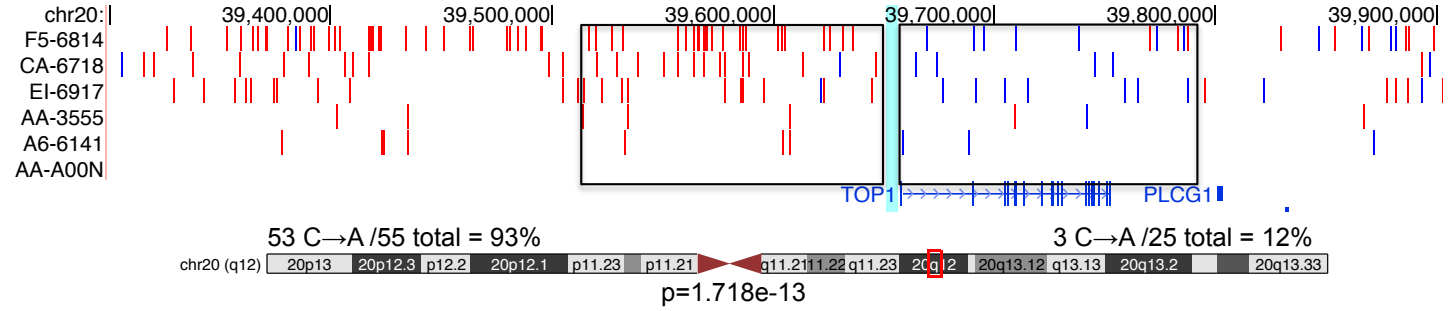
Figure S5


POLE-exo P286H TCGA-AA-3555
POLE-exo V411L TCGA-EI-6917
POLE-exo S297F TCGA-A6-6141
POLE-exo P286R TCGA-CA-6718
POLE-exo P286R TCGA-F5-6814
POLE-exo V411L TCGA-AA-A00N
> 1Mb
100kb - 1Mb
5kb - 100kb
1bp - 5kb
0.00
0.25
0.50
0.75
1.00
Next neighbor within indicated distance
Ratio

Supplemental Figure 5. POLE-exo* context mutations are spatially correlated in all WGS POLE-exo* samples. For each TCT→TAT or AGA→ATA variant we collected the next variant in positive genomic position within the indicated distances (1bp-5kb, 5-100kb, 100kb-1Mb, and greater than 1Mb). The relative amounts of same context pairings to same context is shown. POLE-exo* samples show a characteristic enrichment of same context mutation pairs when the pairs are close in genomic space, this effect diminishes with increased genomic distance.

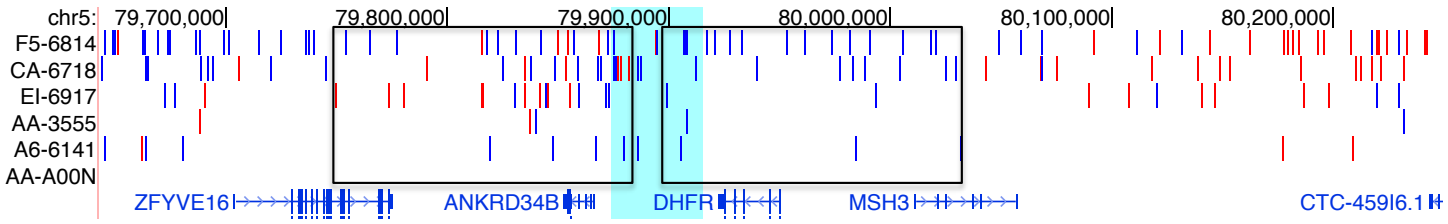
Figure S6

A



p=1.718e-13

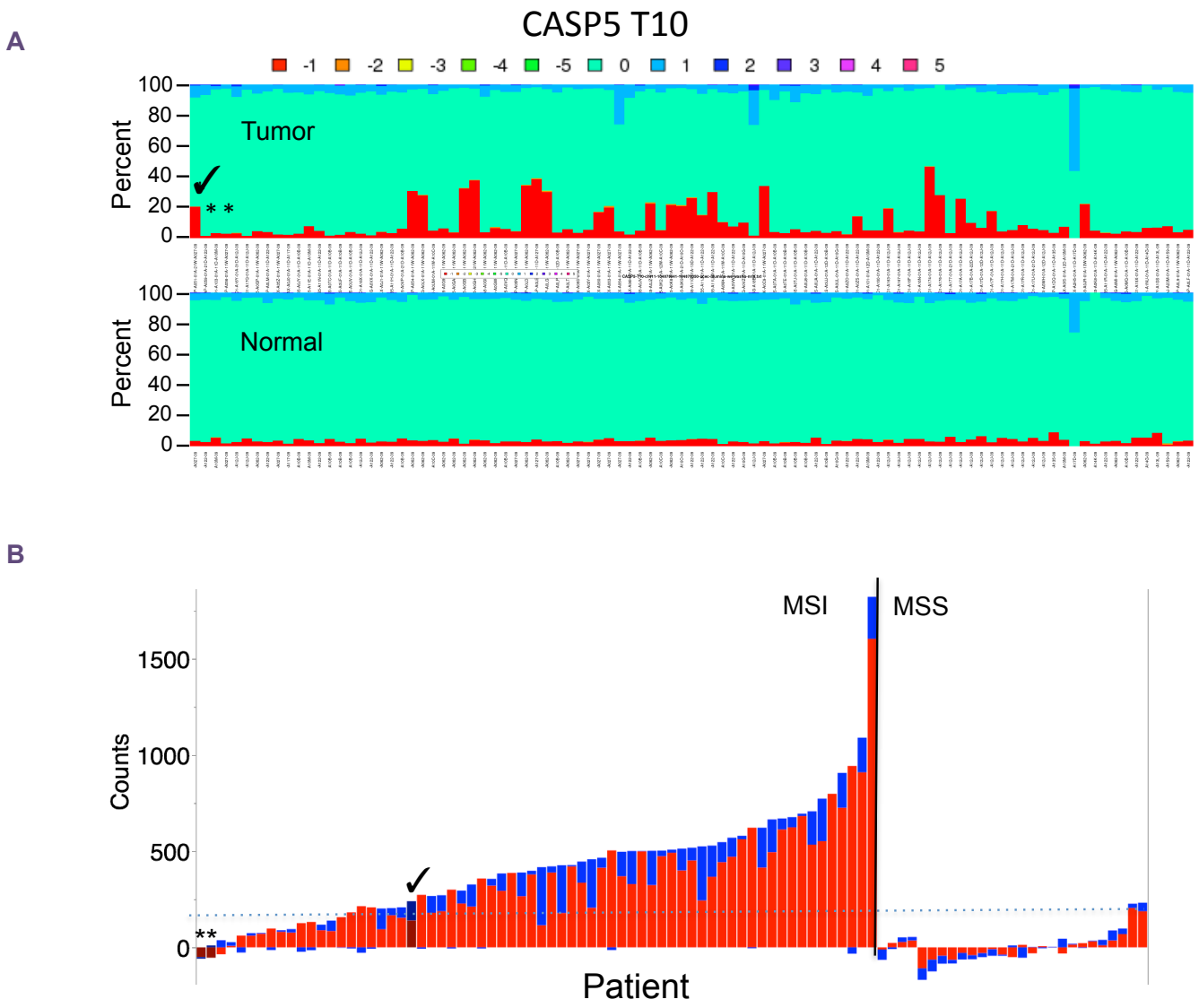
B



p=0.000208

Supplemental Figure 6. Mutations generated by POLE-exo* are mapped to known ORI regions in the UCSC Genome Browser. Mutations of POLE-exo* group A mutants in genome coordinates. Whole genome sequences of POLE-exo* mutant tumors (TCGA-F5-6814, TCGA-AA-A00N, TCGA-EI-6917,TCGA-CA-6718, TCGA-AA-3555, TCGA-A6-6141) were used to count TCT→TAT(red) and AGA→ATA (blue) mutations.
The location of the reported ORI for the given gene is shown by a shaded turquoise bar. The boxes are drawn to define regions of putative strand-specific mutation associated with these origins. (**A**) *TOP1* ORI (Keller et al. 2002). (**B**) *DHFR* ORI (Looney and Hamlin 1987; Altman and Fanning 2001). The ORI for *DHFR* is not defined as a single site, but a "zone of initation" (Hamlin et al. 2010) containing multiple initiation sites. We selected a region adjacent to the *DHFR* gene to count TCT→TAT or AGA→ATA mutations, but note there are multiple regions containing runs of C→A adjacent to G→T mutations in this genomic region. The patterns of mutations (red and blue banding) in this genomic region show potential multiple origins.

Figure S7


A
CASP5 T10
-1 -2 -3 -4 -5 0 1 2 3 4 5
Percent
Tumor
Normal
B
MSI
MSS
Counts
Patient

Supplemental Figure 7. Homopolymer analysis of sequence reads to determine MSI status. **(A)** Histogram showing analysis of insertions and deletions (blue, 1 bp insertion; dark blue, 2 bp instertion; red, 1 bp deletion) mutations or reference (green) across a set of patients. Tumors are in the top panel normal tissue in the bottom. We used a panel of 46 genes with known homopolymer runs (Supplemental Table 4) and tallied the reads supporting insertion or deletion mutation. Tallies were generated for each of the 46 homopolymer containing genes. Only the T10 homolymer region within the *CASP5* gene is shown graphically. Using this method, we re-evaluated the previouly defined MSI status of the UCEC samples. **(B)** Bar graph summarizing the cumulative somaitc insertion delelion burden across all sites in 46 homopolymer-containing genes in each tumor. The counts of insertion or deletion mutations in each homopolymer site (see Supplemental Table 4) were tallied in the tumor and normal. The normal count was subtracted from the matched tumor counts from the same homopolymer site and resulting totals across all sites in all genes were plotted on a bar graph. Samples with insertion and deletion counts below the dotted line were considered to be MSS. Samples above the dotted line were designated MSI. The symbol "✓" marks POLE-exoV424I, patient AP-A051, that remained MSI. The symbol "*" marks POLE-exo* patient AP-A059 and D1-A103 in panel A and B that were changed to MSS status.