Supplementary Information

January 16, 2017

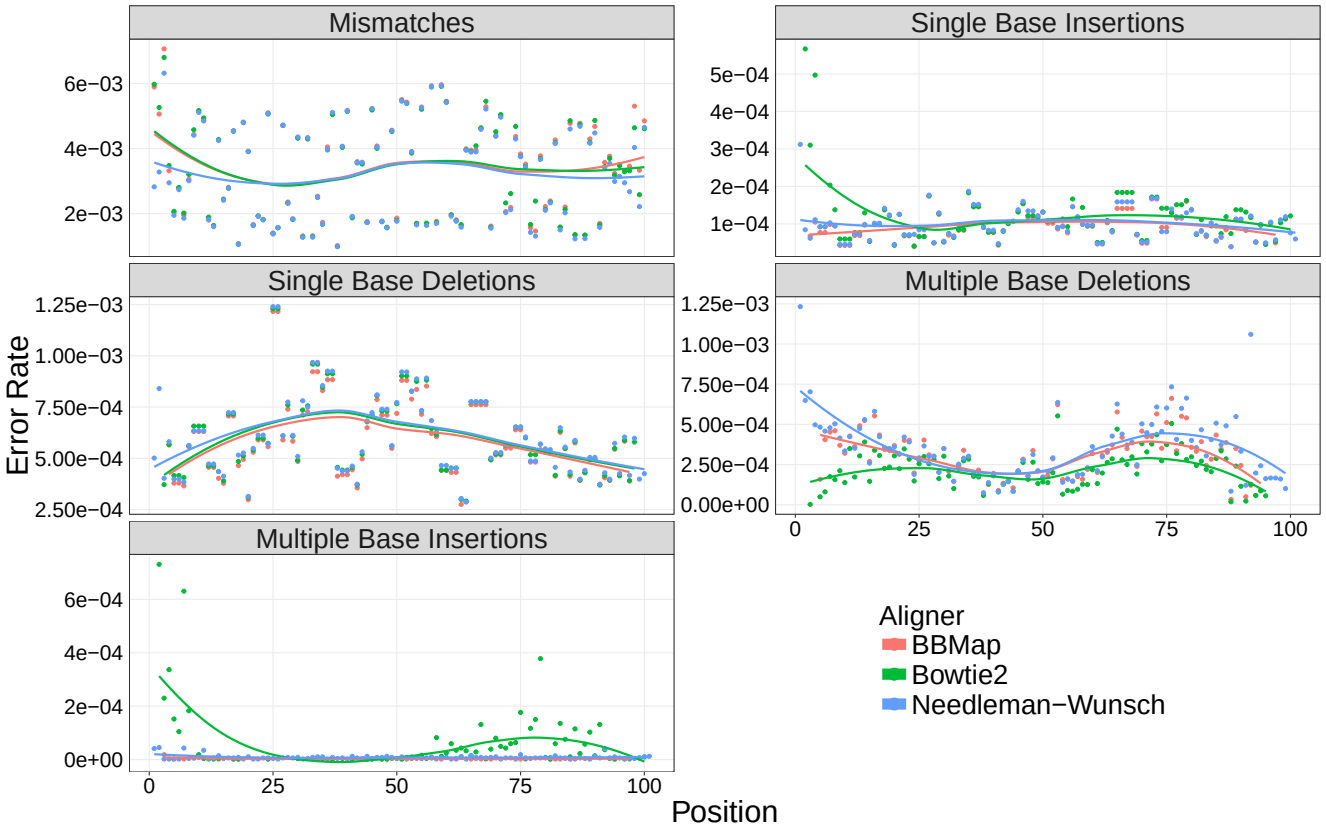

Figure S1: Effect of read aligner on error rates. Here we mapped reads from the standard IDT oligo with `BBMap` (red), `Bowtie2` (green), and our Needleman-Wunsch aligner (blue), and quantified the error rates with our pipeline. We see that the choice of aligner affects the resulting error rates, especially for detecting multiple-base deletions.

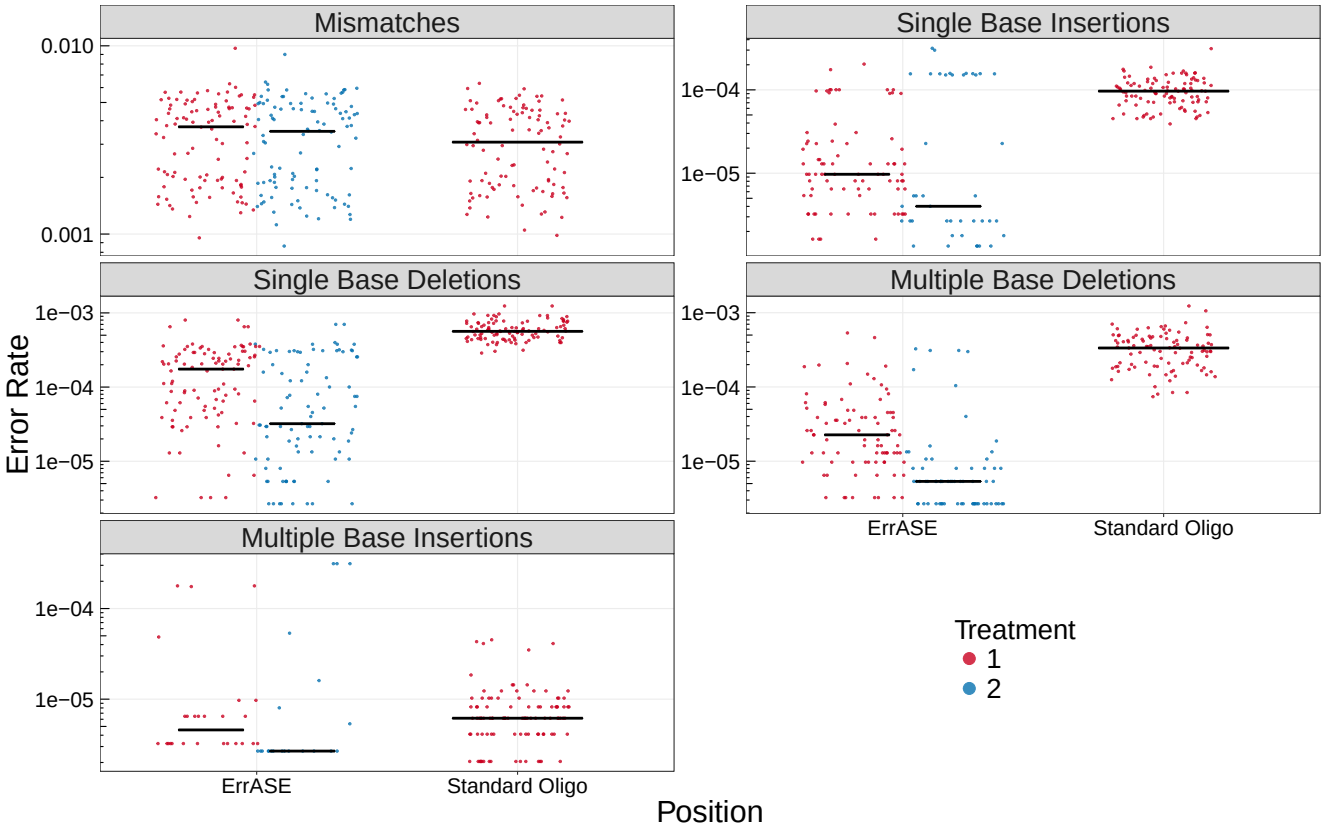

Figure S2: Distributions of error rates per position for the standard oligo assembly before and after ErrASE treatment. We were unable to detect a significant change between the median error rate after two treatments for mismatches. **Note:** black bar is median value.

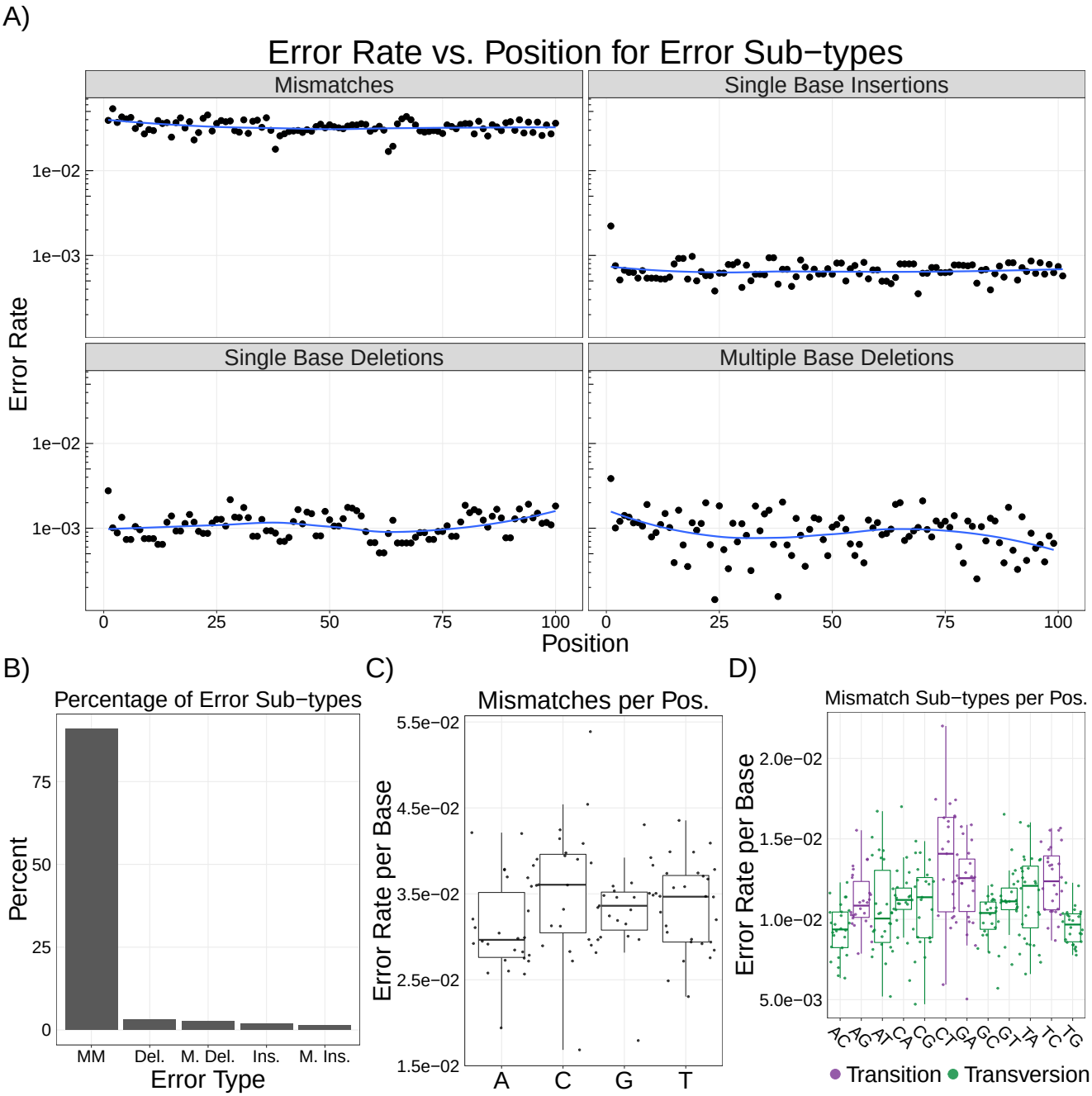

Figure S3: In-depth analysis of error-doped assemblies. Here, we find no significant difference in the median mismatch rate of all four bases, median transition or transversion rate, or rate of single-base deletions for each base (all tests were Mann-Whitney U, NS, Holm-corrected). **Note:** here we performed the same analysis as Figure 2 in the main text with the error-doped assembly.

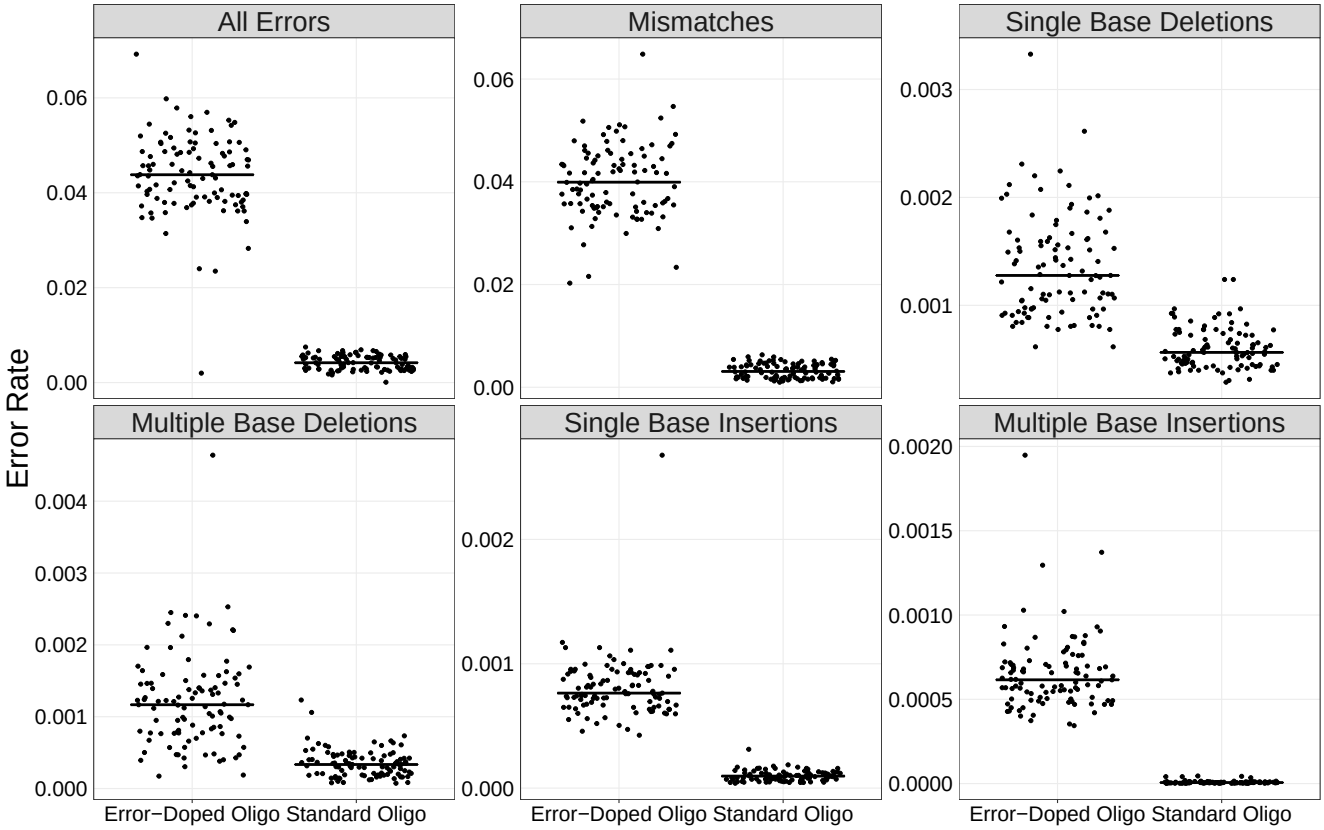

Figure S4: Comparison of measured error rates from error-doped and standard oligos. Here we plot the distribution of error rates per position and see that for every error sub-type the error rates are significantly higher for the error-doped oligos than those produced by the standard process (Mann-Whitney U Test, all $p << 0.001$). **Note:** Black bar is the median value.

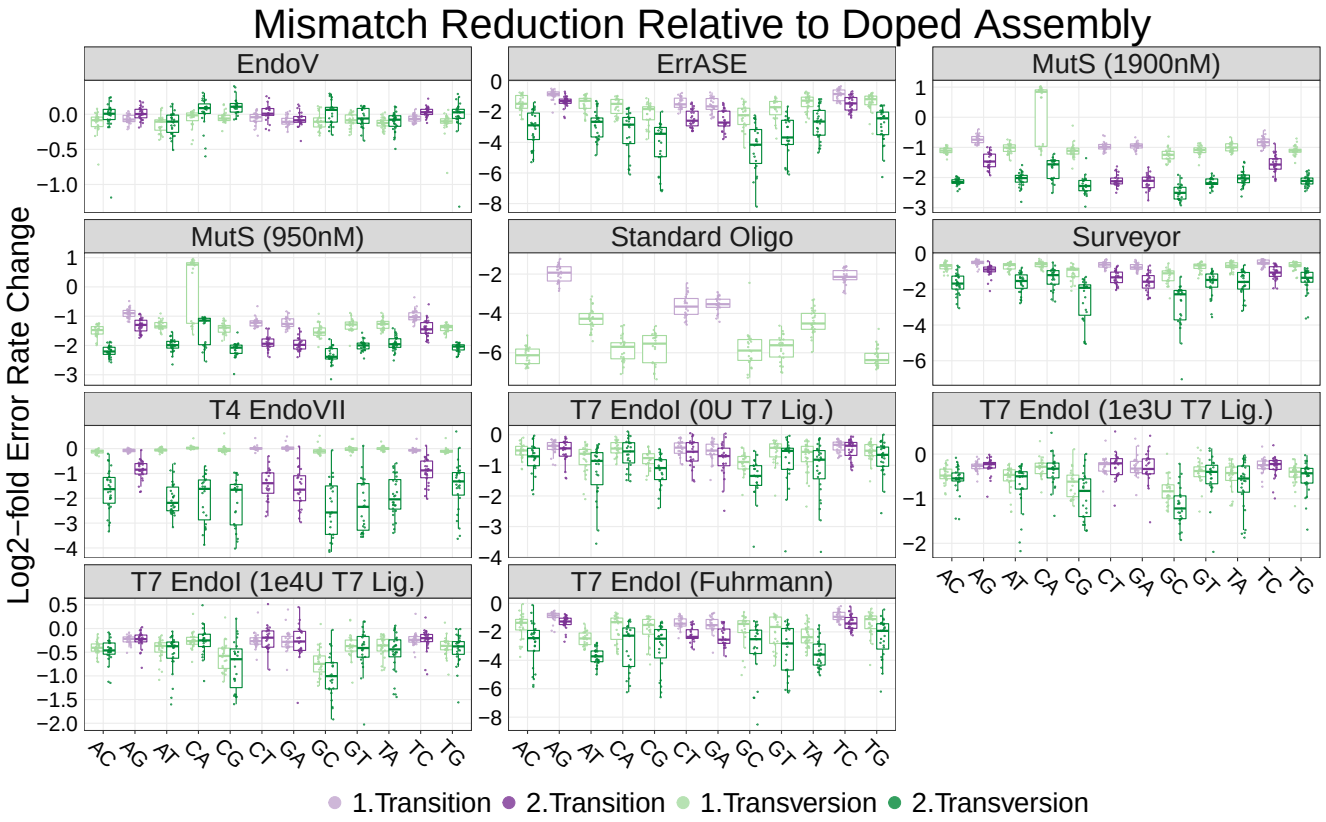

Figure S5: Mismatch correction preferences relative to the error-doped oligo for every enzyme across two consecutive treatments. Error rates are plotted as the $\log_2$-fold-change in error rate relative to the error-doped template. **Note:** box plots are first and third quartile for hinges, median for bar, and 1.5× the inter-quartile range for whiskers.

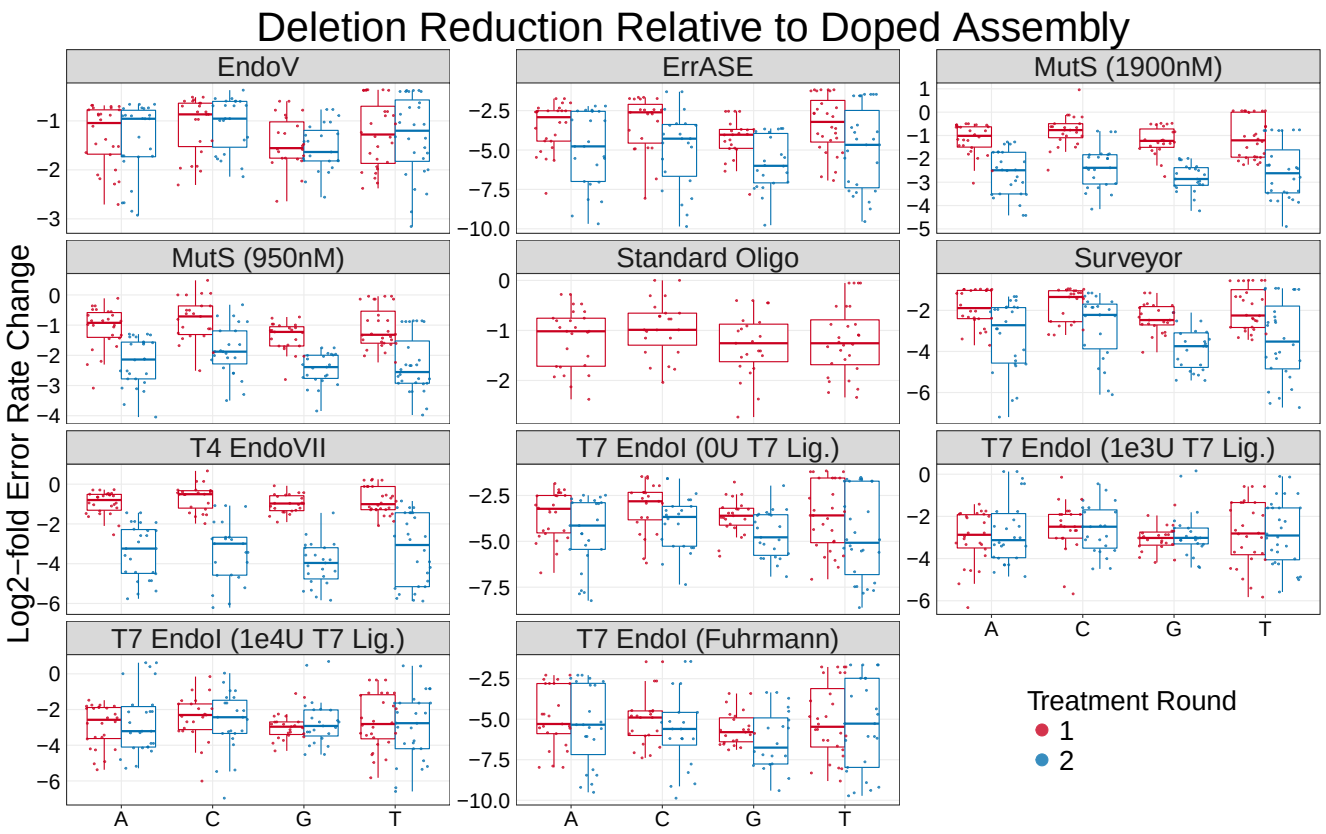

Figure S6: Single-base deletion correction preferences relative to the error-doped oligo for every enzyme across two consecutive treatments. Error rates are plotted as the $\log_2$-fold-change in error rate relative to the error-doped template. **Note:** box plots are first and third quartile for hinges, median for bar, and 1.5× the inter-quartile range for whiskers.

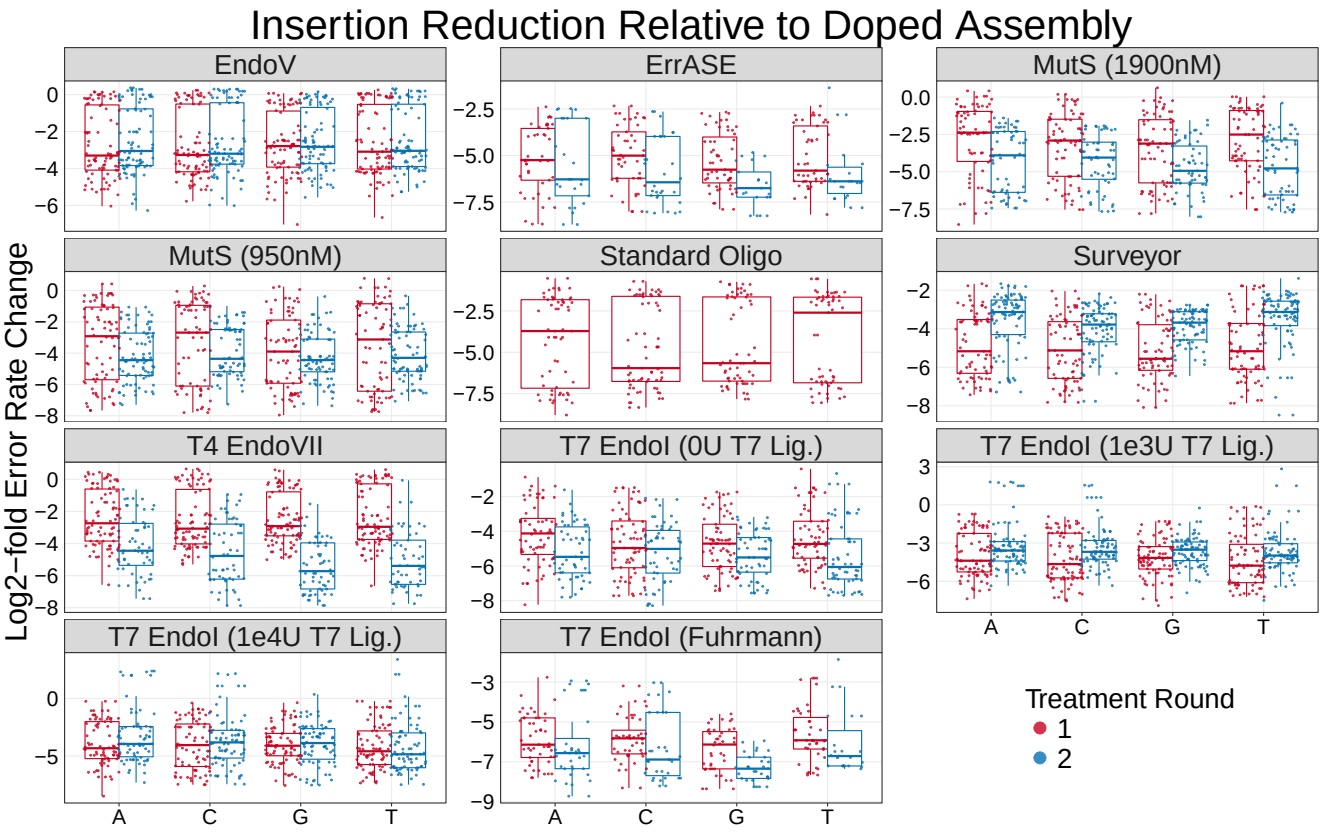

Figure S7: Single-base insertion correction preferences relative to the error-doped oligo for every enzyme across two consecutive treatments. Error rates are plotted as the $\log_2$-fold-change in error rate relative to the error-doped template. **Note:** box plots are first and third quartile for hinges, median for bar, and 1.5× the inter-quartile range for whiskers.

Table S1: Examples of where various aligners fail. Here _ are padding for visualization, * are soft-trimming, and lower-case bases are inserts.

Aligner:	Ideal	Needleman-Wunsch	Bowtie2	BBMap
Reference:	_GCTGCCGATTTCCA...	_GCTGCCGATTTCCA...	G_CTGCCGATTTCCA...	*GCTGCCGATTTCCA...
Read:	aGCTGCCGATTTCCA...	aGCTGCCGATTTCCA...	aGCTGCCGATTTCCA...	*GCTGCCGATTTCCA...
Reference:	__GCTGCCGATTTCCA...	__GCTGCCGATTTCCA...	G__CTGCCGATTTCCA...	**GCTGCCGATTTCCA...
Read:	aaGCTGCCGATTTCCA...	aaGCTGCCGATTTCCA...	aaGCTGCCGATTTCCA...	**GCTGCCGATTTCCA...
Reference:	GCTGCCGATTTCCA...	GCTGCCGATTTCCA...	GCTGCCGATTTCCA...	GCTGCCGATTTCCA...
Read:	GCT___GATTTCCA...	GCT___GATTTCCA...	___GCTGATTTCCA...	___GCTGATTTCCA...
Reference:	GCTGCCGATTTCCA...	GCTGCCGATTTCCA...	GCTGCCGAT_TTCCA...	GCTGCCGATTTCCA...
Read:	GCTG_____TTCCA...	GCTG_____TTCCA...	______GCTGTTCCA...	GCTG_____TTCCA...
Reference:	...TGTTGTATATATCG_	...TGTTGTATATATCG_	...TGTTGTATATATC_G	...TGTTGTATATATCG*
Read:	...TGTTGTATATATCGa	...TGTTGTATATATCGa	...TGTTGTATATATCaG	...TGTTGTATATATCG*
Reference:	...TGTTGTATATATC__G	...TGTTGTATATATC__G	...TGTTGTATATATC__G	...TGTTGTATATATCG**
Read:	...TGTTGTATATATCatG	...TGTTGTATATATCatG	...TGTTGTATATATCatG	...TGTTGTATATATCa**
Reference:	...TGTTGTATATAT__CG	...TGTTGTATATA__TCG	...TGTTGTATATA__TCG	...TGTTGTATATATCG**
Read:	...TGTTGTATATATgtCG	...TGTTGTATATATgtCG	...TGTTGTATATATgtCG	...TGTTGTATATATgt**
Reference:	...TGTTGTATATATCG	...TGTTGTATATATCG	...TGTTGTATATATCG	...TGTTGTATATATCG
Read:	...TGTTGTATATA__G	...TGTTGTATATA__G	...TGTTGTATATAG__	...TGTTGTATATAG__
Reference:	...TGTTGTATATAT__CG	...TGTTGTATATA__TCG	...TGTTGTATATA__TCG	...TGTTGTATATATCG**
Read:	...TGTTGTATATATgtCG	...TGTTGTATATATgtCG	...TGTTGTATATATgtCG	...TGTTGTATATATgt**