



Supplemental Fig S5. Frequency of mutation candidate or duplex disagreement detection as a function of distance to ligation site.

(a) The histogram (associated with the left y-axis) shows the distribution of ligation site distances for mutations passing all Q2 thresholds *other than the threshold to do with ligation site distance*. For comparison, the line (associated with the right y-axis) shows the distribution of ligation site distances for wildtype duplexes passing the same Q2 thresholds. The threshold distance used in the main analysis to separate excluded and retained mutations was chosen at the base of the left peak in mutation frequency (see Supplemental Table S1 for numerical values of all thresholds). (b) Same as (a), but showing duplex disagreement rates instead of mutation frequencies and distinguishing between mated and unmated samples. Disagreements were also found to peak close to ligation sites. Error bars represent SEM across samples. Data in both panels are derived by aggregating all DNA-seq samples reported in this manuscript. (c) Average Phred score as a function of distance to ligation site, for duplexes passing Q1 thresholds.