



**Supplementary Figure S2.** Sensitivity of COMET to type known sequences with varying levels of noise (shown in the legend) and fragment sizes. A synthetic dataset was generated from reference sequences from the LANL HIV database, by randomly introducing mutations throughout the genome ("noise"). The sensitivity of COMET was calculated for varying degrees of noise (0-20%). Sequences of different lengths starting at different positions were submitted to COMET.