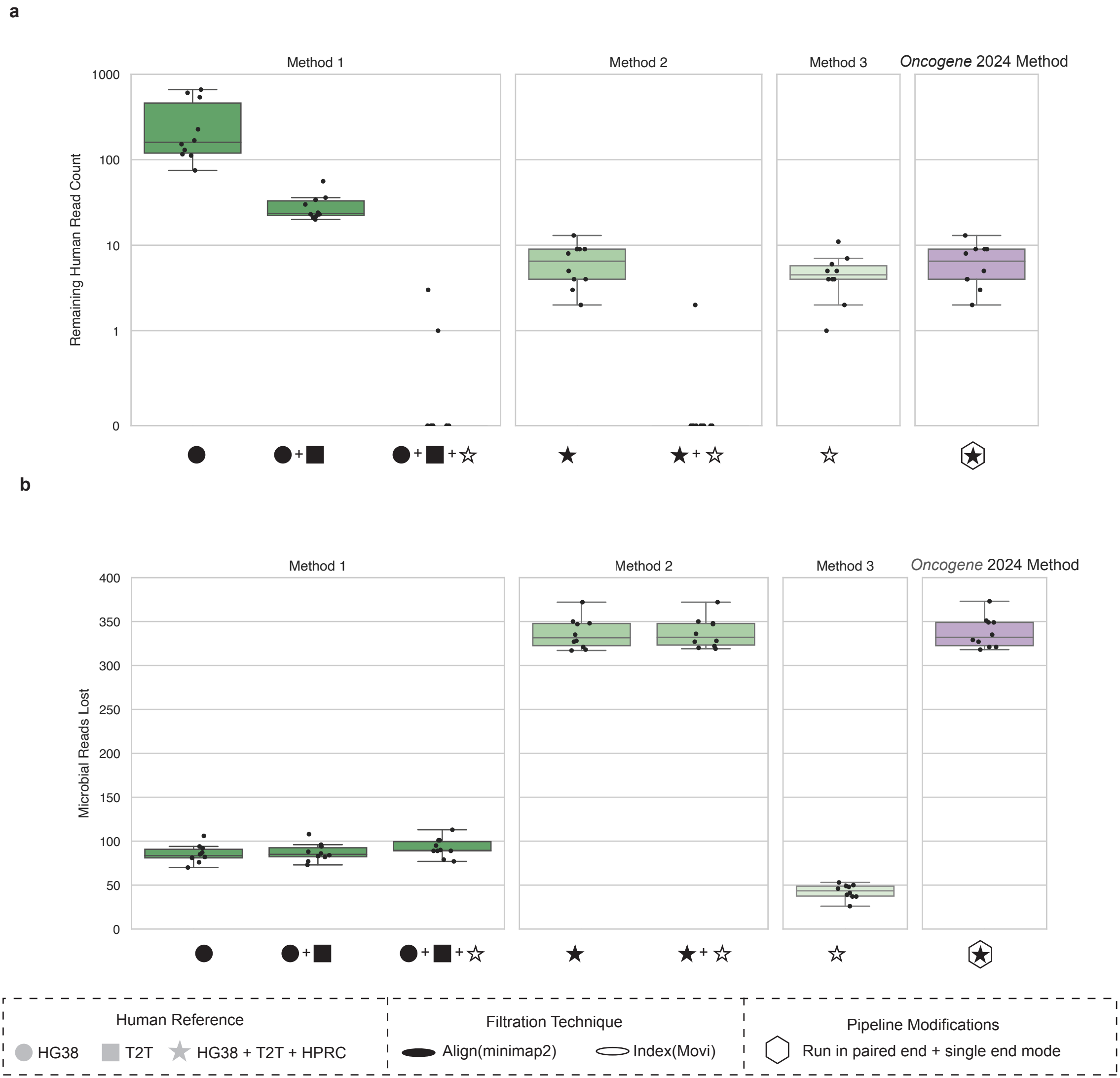

Supplementary Figure 3. Host filtration pipeline simulated data validation Using the 10 simulated datasets of 1 million reads as described in Figure 2b (a) calculated the number of human reads remaining, and (b) number of microbial reads remaining, for host filtration Methods 1-3 as well as the method used in Sepich-Poore,G. et al., *Oncogene* 2024 which used minimap2 in both paired end and single end mode (HPRC host filtration done excluding the 10 pangenomes simulated on). HG38 : GRCH38.p14, T2T: T2T-CHM13v2.0, HPRC: Human Pangenome Reference Consortium 2024 release.