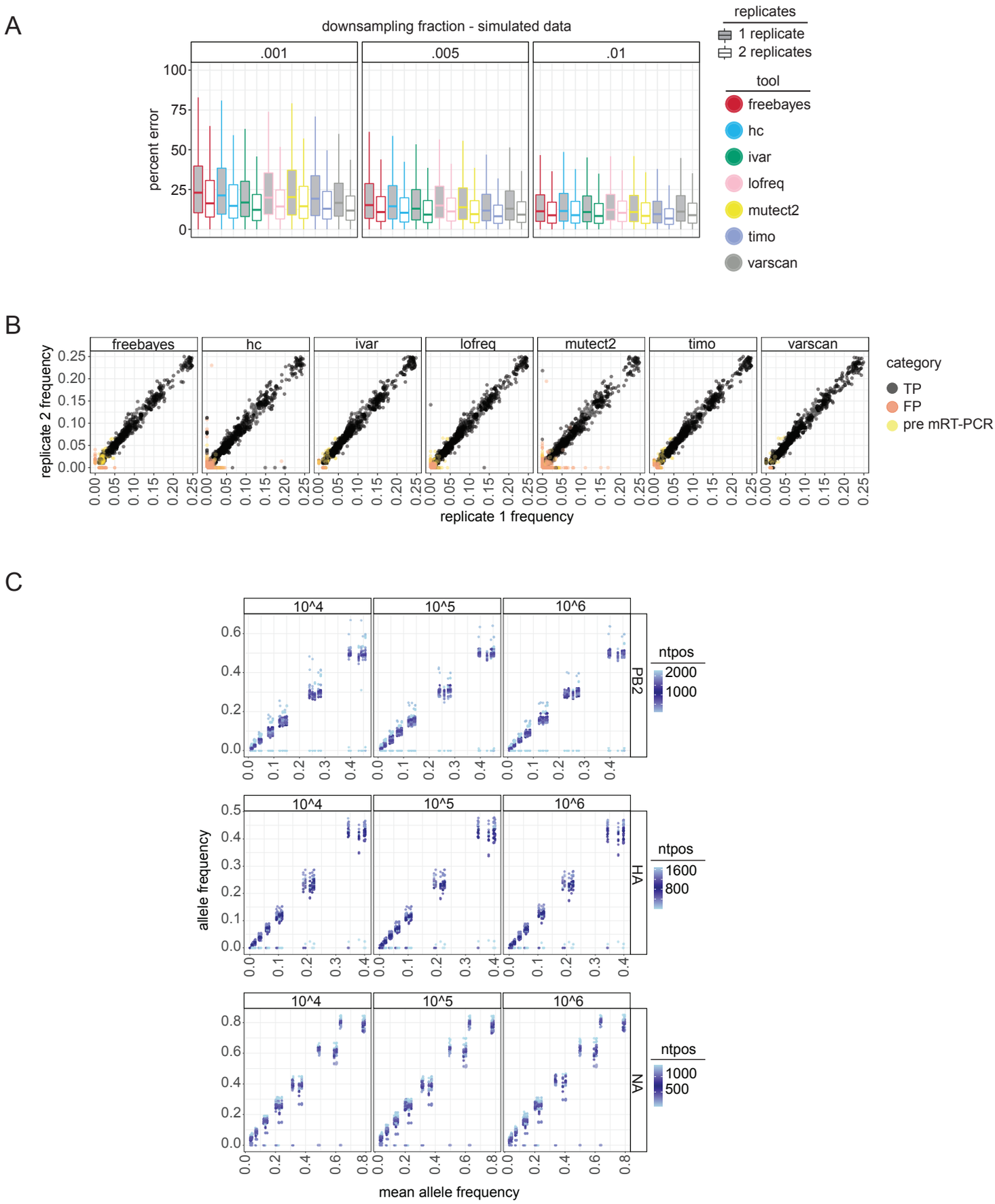

Figure S5. Effect of sequencing replicates and genome location on the accuracy of allele frequency estimation. **(A)** Percent error ((|(observed-expected)|)/expected*100) of single replicate (grayed boxes) or merged replicate (open boxes) SNVs in simulated influenza A data across downsampling fractions (0.001, 0.005, and 0.01). Color represents the variant caller used. **(B)** Scatter plot showing the frequency of variants in synthetic influenza data across sequencing replicates with frequency in replicate 1 on the x-axis and frequency in replicate 2 on the y-axis. Color represents the SNV type. **(C)** Concordance between the output allele frequency (y-axis) and mean allele frequency (x-axis) across nucleotide position and copy number (10^4 – 10^6) on each synthetic influenza gene. Output from both replicates is shown. Color represents nucleotide position, with darker colors near the ends of the gene segments.