



## Figure S2

**Trinucleotide sequence is more impactful on background substitution type than more distant surrounding bases.**

**a**, Shown for each of the 6 possible substitution types are the relative frequencies at which each of the various nucleotides are found at a given position surrounding the variant nucleotide (plotted as means with standard deviation between individuals as error; n = 22). Within each of the subplots, the variant is not shown and exists at position 0. It appears that the greatest skewing of nucleotide representation occurs at positions -1 and 1, suggesting that they have the greatest impact on how a base will change when it suffers a substitution in vitro. Note that for C changes, underrepresentation of G at position 1 is expected based on low representation of CpGs in the captured regions. **b**, As seen in Fig 2a-b, substitutions tend to exist within an upper or lower region of allele frequencies. To understand if flanking nucleotide sequence plays a role in this, the populations were analyzed separately for each of 6 base changes at Cs and Ts. Suggesting that the actual mutation plays a role in the resulting VAF, most substitutions exist disproportionately in either the upper or lower population rather than being equally distributed between the two. For some comparisons, this resulted in larger error within one of the populations, rendering some comparisons not feasible. For C>T changes, the flanking base sequence was largely conserved between the two populations. Other substitutions show differences in flanking sequence when they exist at higher or lower VAFs, as observed for T>A at positions -1 and 1 and C>A at position 1.