Filter out CpG variants and high frequency variants ($AF > 0.1\%$)

Define neutral regions following the INSIGHT paper and filter out variants in non-neutral regions

Build a global mutation model:
$$log\left(\frac{P(obs)}{1-P(obs)}\right) = logit_g = b_g + W_g^T \cdot X$$
X: mutation type (7-mer context + derived allele), sequencing coverage, GC content, CpG island

Calibrate local mutation rate with a sliding window method: $logit_l = b_l + W_l \cdot logit_g$