Definitions

Define a "mutation class" to be a single base substitution in trinucleotide context, e.g. AAA > ACA AAA > AGA, AAA > ATA, CAA > CCA, …, TTT > TGT.

Let *μ* denote a mutation class and let be the set of all mutation classes.

Let denote the source trinucleotide of , e.g. s(AAA > ACA) = AAA.

Let denote the count of trinucleotide in the human genome and let be the set of all trinucleotides.

Let be the relative proportion of trinucleotide t in the genome,

.

Define C\_g(mu) as the observed count of mu in a human genome.

Define O to be the observed mutations per million source trinucleotides in a human genome; O(mu) = 10^6 \* C\_g(mu)/G(s(mu) ).

Define the term "***standard representation of a mutational spectrum***" to be

. It is the mutational spectrum we would observe in a genome in which all trinucleotides had equal frequency.

xxxxx

To convert a "PCAWG" or "COSMIC" genome-count mutational signature, in which relative proportions are based on observed genome counts, , to a "standard representation" (a.k.a. "flat") mutational signature, do

To convert a standard representation mutational signature to a "PCAWG" or "COSMIC" or genome-count mutational signature, in which relative proportions are based on observed genome counts do

Let be the relative proportion of trinucleotide t in promoters,

. To convert a standard representation mutational signature to a promoter-count mutational signature, ,in which relative proportions are based on observed promoter counts do