New distribution over base pairs (ATCG)

* Generate a long sequence (ATCG)
  + Chop up randomly (10 parts)
  + Generate transcripts randomly (3-5 parts of the 10 parts)
  + Some decreasing exponential distribution
* Plot distributions over base pairs (some transcripts are shorter than others)

Distributions -> likelihood of getting the reads

Reads are equivalence classes

Make a weight or attribute telling which equivalence class it is coming from

Randomly sample per each base pair? For distribution

Recall they overlap so they should have similar distributions

2 models

1. truncate the distribution (still sum to 1)

2. stretch or shorten the distribution to fit the transcript length

genes -> isoforms

dude gg