## 2024.12.1

1. 这两天写得头大，把gapped kmer的各个方程写完了，并且跑起来了。使用这个方法可以让每个长度（l）的总频数大致相同，然后再使用4^l的normalize方法，这样得到的性能最好，scc在0.22但是还是不如kmer unigram。（使用长度增益的方法，基本上出来的都是3mer片段，很显然没有达到要求）

2. 明天试一下在kmer unigram的情况下，把每个长度下的总频数先归一化，像gapped kmer一样，然后再用4^l的方法。