Different ways to improve imeter:

-account for degenerative sequences

-account for – and + strands

-set up a way to weigh the beginning of the intron vs the end of the intron (should work on finding a geometric decay in it)

-account for sequence repetition due to splice variants

-decoding, putting things into libraries (proper program in argparse, put functions into some shared library)

-trainer program

-decoder program

-scientific studies

-controls? Experimental data and controls ( R squared value)

-do certain kmers predict long introns? Long introns will contribute more kmers, if one is more common in long kmers, will be disproportionately weighted.

-could weight kmers on length of parent intron

-use expression as a response variable, like a checker.

+ incorporate 4-5 fold cross val, **eventually\***