Nate Evans

Christina Zheng

BMI650: Algorithms

12/7/18

Post-Presentation Report

This is a review of the recent Algorithms Final Projects presentations and reflection on my own project in light of Wednesday’s discussion. The goal is to answer:

1. How was my algorithm a naïve/brute-force approach? or how was it not?
2. What would I change about my own project?

In regards to the first question, my algorithm approach was not a naïve/brute-force approach because it lacked two significant components. First, while I understood the need to include variation data from only exon regions, I did not splice out the introns from my gene regions and therefore compared variation in the full gene region rather than only exons. Additionally, since we are working with RNA-seq data, and we know there are alternative splicing mechanisms for some genes, a naïve approach would include variation data from all possible exon splicing orders (e.g. exon order from each RefSeq transcript).

To answer the second question, I would change my algorithm in two ways. First, I would use the .gtf file to glean exon regions rather than using the entrez search feature to pull them from online databases, this would have made it infinitely easier to create an exon mask and thereby allowing me to compare PWK-B6 variation in only the exonic regions of chromosome 1 rather than in the full gene-regions as I currently have it implemented. Second, I would use the data from the .gtf file to produce all permutations of alternative exon spicing, and thereby compare all possible ways that the mRNA could be ordered and include variation data from all of these into my analysis.

I believe the changes detailed above and the incorporation of .gtf data (rather than using entrez) would be sufficient to classify my algorithm as a brute-force approach to calculating an optimal mismatch threshold for RNA-seq alignment.