Khan Inan Professor Truong BI-GY 7683

- 1. Alignment-free RNA sequencing methods essentially involve using k-mer counts to find the maximum likelihood estimates of relative transcript abundance. The goal is to find the relative abundance of transcripts and isoforms because each time we find a read it can be associated with different transcripts of a gene. How this differs from the previous alignment based RNA-seq analysis is that instead of comparing the sequence of the read to a reference genome, we are using statistical analysis to calculate probabilities and how reach read might have originated from a particular transcript.
- 2. In terms of using the same alignment-free method for DNA sequencing, I can imagine it being most useful in cases where a sequence is divergent and reliable alignments cannot be used. Additionally, on a practical level, I can see alignment-free methods being most useful for small-scale labs that don't have a large amount of sequence data and are limited in their computational resources. More specifically, in a hypothetical scenario if people were trying to DNA sequence an organism that lived prior to the mass extinction event in the Cretaceous period. I say this because there would be no reliable reference sequence to align with, since there would be no living relatives to an organism that went extinct that long ago.
- 3. The conceptual differences between aligning reads obtained from sequencing RNA versus DNA comes down to the fact that DNA sequencing tools and the ability to work with DNA molecules is more developed. In fact, RNA sequencing itself can be considered DNA sequencing in a way because RNA is converted to cDNA via reverse transcription and then the sequencing is done. Enzymes that process RNA in its base form are said to have low fidelity, which means that they cannot replicate a template accurately in most cases. To make things even worse, RNA is its base form is less stable than DNA, and when sequencing efforts can take long periods of time it's good to have samples with a long shelf-life. On a more fundamental level, DNA provides an idea of what a cell can do or become and how they might progenate, while RNA sequencing is a snapshot of the current state of the cell/organism. It is because of this reason that DNA-sequencing efforts are more widespread, because an organism's "genetic future" is more interesting to researchers and scientists compared to an organism's present state of being.