

COM SCI C121 Week 1

Aidan Jan

April 2, 2024

Biology Review

Central Dogma of Biology

- DNA is transcribed to RNA, which is then translated to proteins.
- During transcription, splicing may occur, so one section of DNA can produce multiple different strands of RNA, which is then translated into different proteins.
 - Occurs often in more complex humans (e.g., not bacteria)
- The definition of a gene is complicated, since one gene may have multiple exons, which may be spliced into different RNAs. (How do you quantify the isoforms?)
- Difference in splicing, translation, and regulation are part of what defines cell types.
 - This means that molecular mapping needs to be done for all different contexts
 - Computationally, we *need* fast, accurate, and space-efficient algorithms.

21st Century Biology Revolution

- High throughput DNA sequencing has revolutionized modern biology
- Can sequence billions of DNA fragments for relatively cheap (~\$1000)
- Many biological questions can be reduced to sequencing experiments
 - e.g., RNA-Seq, ChIP-Seq, Methyl-Seq, RIP-Seq, CNV-Seq
- Currently, hundreds (thousands?) of experiments (since ~2008)
- If you can reduce your experiment to a sequencing experiment, you can essentially do **thousands** of experiments at once.

What is DNA?

There are many types of biomolecules. (e.g., carbohydrates, lipids, proteins, and nucleic acids).

- DNA is a type of nucleic acid.
- DNA stores all the genetic information that a particular organism needs to live.
- DNA is stored in nearly every human cell. DNA inside chromosomes, inside nuclei, in cells.

DNA, genes, RNA, and proteins

- DNA contains coding and non-coding regions.
 - Coding regions are referred to as *exons*.
 - Non-coding regions are referred to as *introns*.
 - There are non-coding regions outside of these two groups, but are not discussed in this class.
- Introns exist to allow the same DNA section to code for multiple different proteins
 - Introns of some proteins may be exons of a different protein.

DNA Strands

DNA has two strands - the forward and reverse strands. Which one is forward strand is arbitrary - someone just picked it.

- The forward strand goes from 5' to 3' (these are names for the ends); the numbers represent the direction transcription occurs - transcription always occurs from 5' to 3'.
- 5' and 3' are named based on how the carbons are bonded.

Random Useful Facts about DNA

- A human "genome" stores about 3.1Gb (just one side of a double helix)
- Humans are 99.9% genetically identical
- A great overestimate of a person's variability is 3M genetic variants
- If we take the union of all single nucleotide variants, it's only ~8M (> 5% allele frequency)