Computational Genomics

Lecture 1

Learn the fundamentals of biological Sequence analysis

A close up of a paper

Description automatically generated

Central Dogma of Biology:

* DNA contains genetic code blah blah blah
* mRNA is an intermediary for protein
* protein does work in the cells

Differences in splicing translation and regulation are part of what defines cell types, and due to this we need to do computation molecular sampling for each one of them.

You don’t need genomic sequencers to sample dna, rather you can just use computational molecular analysis to do that for you for relatively cheap.

Basis: one of the A C G or T’s

Primers bind with each other, the 5’ binds to the 3’, and this is for both the forward and reverse strands. Strands have non-coding and coding regions, which are called exons and introns, the coding parts of which are used for protein synthesis instructions.

The forwards and reverse strands bind with each other using a single connection between both the proteins, which gives the double helix shape that the strands of the DNA.

Chromosomes contain two chromatids, one from the father and one from the mother, which are basically two double helix strands that encode genetic information.

For now, current technology can only sequence a few hundred nucleotides at once in a reasonable time, and we would be specifically focusing on illumina genome sequencing technologies(which is a specific company product). Anything that usually ends with a seq is usually using an illumina model.

We are gonna assume that we did genome sequencing already. We’re gonna assume these protein sequences are just data. These reads are going to be 100 nucleotides long.