# Lecture 11

Gidon Rosalki

2025-11-25

## 1 Reminder

Last lecture, we (allegedly) discussed how genomes are made of sequences of DNA, spooled into nucleosomes, which prevent the DNA from damage, and from becoming tangled, since we do not want the DNA to be full length.

## 2 Packaging proteins

We have various modifications of the packaging proteins, such that we can bind them to nucleosomes, rather than to individual base pairs. Since we can bind to nucleosomes, we can start reading the DNA/genes as groups of nucleosomes, rather than as individual base pairs. We can bind in H3K4me5 in solution, and we will observe a peak corresponding to a start point in the encoding of the genome. We can then bind in H3K36me3, which binds across the entire section that is read. This appears to be a great way to find active parts of the gene, until we established that it also binds to other parts that are inactive. Ah well.

We can model all this through HMMs (who would have guessed?!). The hidden state is what is happening in the genome. This can then be translated into appearances in the peaks of the concentration graphs. To enable discretisation, we split the genome into sections of 200 base pairs, which means in the resultant peak graphs, we can choose easily if each area is a 0 or 1. This is mildly arbitrary (though probably designed to neatly align with the approximate size of nucleosomes), and a good way to bastardise the data, but does make it easier to analyse.