Relating Epigenetic Information to the Structure of DNA Using Deep Learning – Vanessa Job

In order for genes to be expressed, their encodings in DNA must be accessible. DNA is packaged around protein spools, forming a structure called chromatin, where DNA is wound or unwound from these spools depending on which regions are being accessed, (i.e., which genes are being expressed). Modifications called epigenetic tags attach to the spools; these tags influence which regions of DNA are exposed. The structure of chromatin is dynamic: different pieces of DNA may be exposed depending on the cell's environment. Some changes in structure are normal, for instance exposure of clock genes at different stages of the circadian cycle, while other changes in structure indicate disease such as cancer.

The goal of our research is to build a machine learning system to relate the structure of chromatin to epigenetic tags. To do this, we relate data from structural assays (which determine the current configuration of chromatin) to epigenetic assays (which indicate location of epigenetic tags on the genome).

Since assays require costly materials and the time of an expert to conduct, using a machine learning system to predict the result of one assay from another has potential to accelerate the pace of biological discovery. Additionally, such a system will enable virtual experiments which will elucidate the relationship between epigenetic tags and the structure of chromatin.