- What is the objective of this work, and what kind of data are they using? Why is the prediction problem important?

The goal of this work is to study how SNP effect genomic accessibility. This is an important task because the genome changes which parts are and are not accessible as a function of cell type and environment. Being able to predict how SNPs will affect accessibility could give insight into how these SNPs cause changes in gene expression and give rise to diseases. To train their model they use DNase sequencing which works by cleaving non-protected regions of the genome (areas without nucleosomes) and then sequencing those regions. These regions will be where transcription factors bind and give rise to gene expression.

- How is this work different than DeepBind? How is the Basset architecture different than the one used for DeepBind?

DeepBind goal is to predict binding motifs of different transcription factors. Basset goal is to predict weather a region will be accessible in a target cell. Really interestingly that some of the convolutional filters in Basset were found to be motifs to transcription factors. The two architectures are similar, both using CNN to linear layers to predict a final output. The major differences are in the outputs, DeepBind predicts a binding score where as Basset predicts a vector of which cells the sequence will be open and closed in.

- What are the paper's contributions (in your opinion and in the author's opinion)? This can come in the form of ideas or insights, methods, or software.

The author says that this paper gives researchers a way to, with a single sequencing experiment, get knowledge on DNA accessibility. They did this by taking a common tool and applying it in a unique way. I think that the paper does a good job of exploring the network and figuring out what it is learning. Allowing for more information to be extracted by the tool.

- Do you consider this an important paper? If so, what makes it important? How has it advanced the field?

I do consider this an important step in the field. Being able to predict DNA accessibility is a very difficult task. The problem is that genomic architecture changes constantly under different conditions. They were able to find a base set of cells and predict how that is affected but if you were to stimulate those cells with a condition, like a drug or stress, many of those predictions would no longer be accurate. So, I do believe this paper has very useful information that could be expanded further to target future challenges.

- What questions are you left with? What questions would you like to raise in an open discussion of the work? What do you find difficult to understand? List as many as you can.

I found it difficult to figure out where many of their measurements are coming from (figure 4 as an example). They also started on cell-to-cell accessibility, but then later they were studying SNPs. Were these SNPs studied at a cell-to-cell level or globally. Some of their scoring metrics were difficult to understand where they were coming from.