Meeting notes 10_2_2018
Johannes Soeding, Lukasz, Ariel

Lukasz: for some reason, ARG3 data is easy to predict but ARG1 data is harder.

Ariel: can charge be separated from other factors in the sequence? Johannes: no.

Difference between 99% and 95% ROC is a large differernce.

We should try the models using a more realistic situation where the negatives are at a 10/1 or 20/1 ratio compared to the positives. In practice, perhaps 1% to 0.1% of peptides have activator function (Ptashne and others).

We should calculate ROC for these other ratios.

This situation might be better to find unknown factors that influence the results.

Plans for future:

- Find other features that influence the model results.
- Make regression model (take into account the activator function of each peptide from GFP data)
- Think about using different sets of data to train models such as the top 1000 or 100 activators. Up to now, when we use all activators, most are probably weak activators. Perhaps the strong activators have different sequence requirements.
- Try to collaborate using google docs for paper outline and list of experiments.
- Lukasz: # filters says something about motif size. Probably filter width too but this discussion was not clear to me.

Next meeting: Oct 11, 9 am.