Anik Mitra
Optional HW

Q2a

SELEX-seq:
It is a high through-put assay two determine the recognition sequence of the DNA binding proteins in vitro. The protein of interest is immobilized on a microarray chip and the random pool of oligonucleotides are run through the chip. The bound sequences are amplified and sequenced which is later used for finding the pattern.

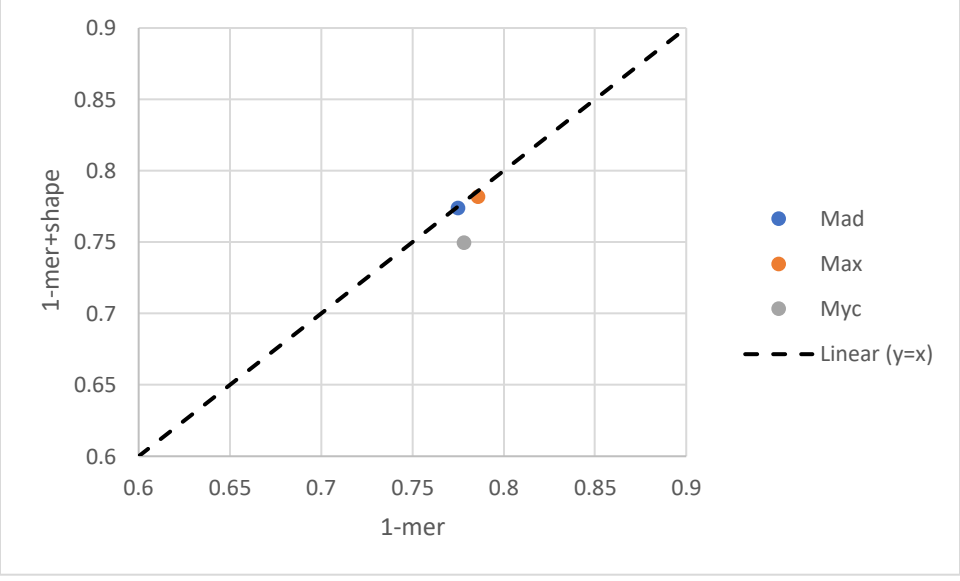
PBM:
It is also a high throughput assay used for the same purpose, but in this case ssDNA is immobilized on the microarray surface. And a primer is used to synthesize the dsDNA with the help of DNA polymerase. Then the protein of interest tagged with fluorophore is used to bind their recognition sequence and the chip is imaged for the fluorescence intensity.

Q2b.

ChIP-seq:
It is a high-throughput assay used in vivo to determine genome wide binding of the DNA binding proteins. In short, the protein is crosslinked and specific antibodies are used to immunoprecipitate the protein of interest, followed by reverse cross-linking. This leaves with strands of DNA bound to the protein of interest, which could be sequenced.

Q2c.

Q5a.



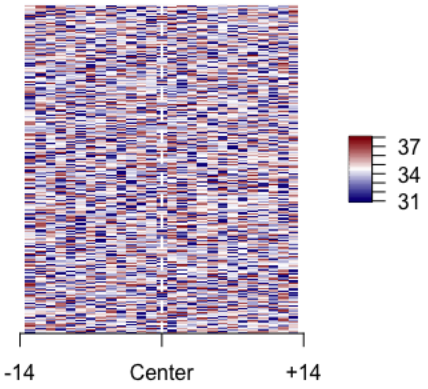
Q5b.

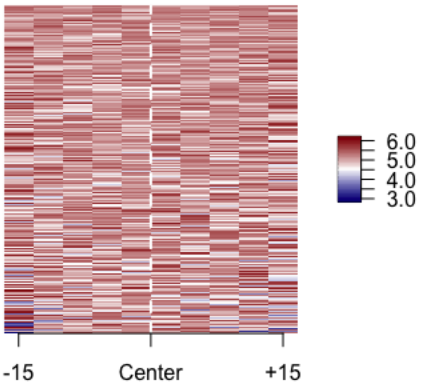
Anik Mitra
Optional HW

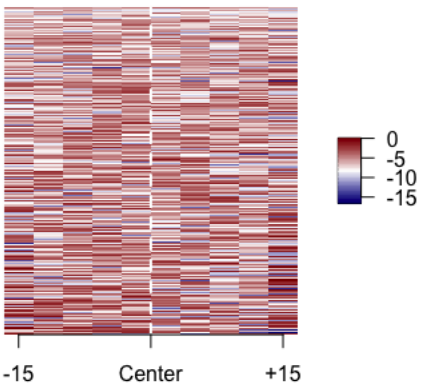
Not sure what I have done wrong, but it looks like 1-mer Rsquared is more compared to 1-mer+shape, which is in contradiction to Zhou et al. I have used glm to train the mlr to the best of my understanding, however not sure if that's the correct model for the training.

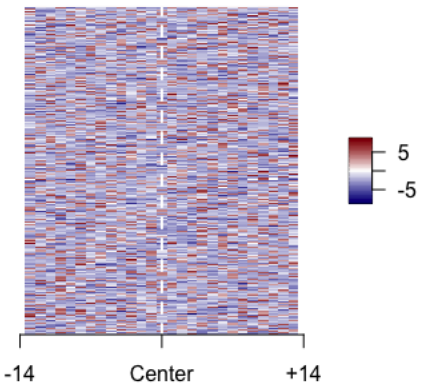
Q7

Bound.fa

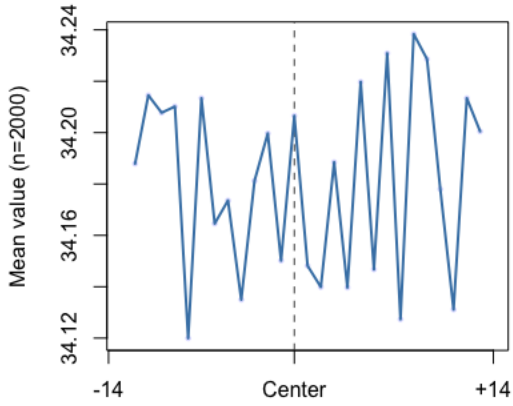


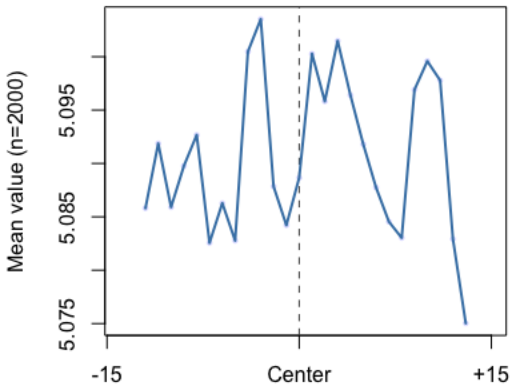


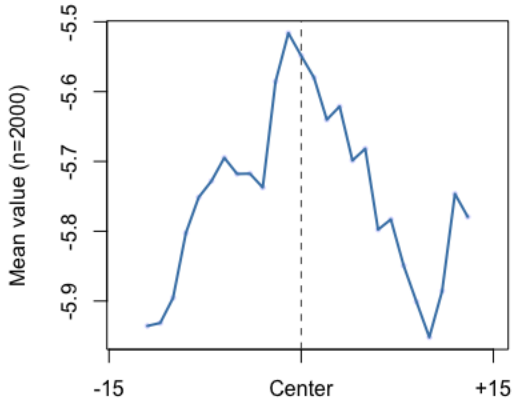


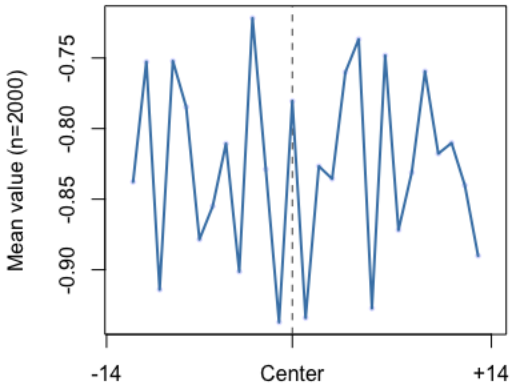


Anik Mitra
Optional HW



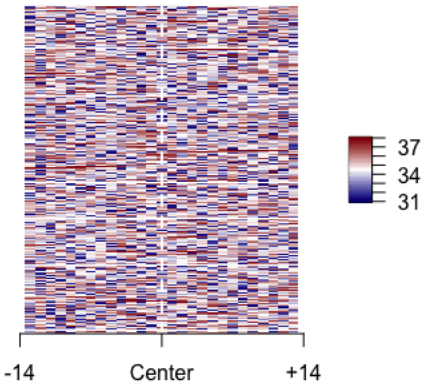


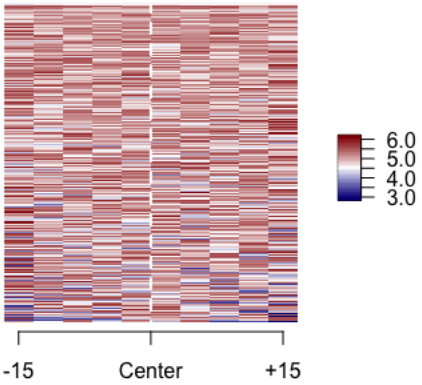


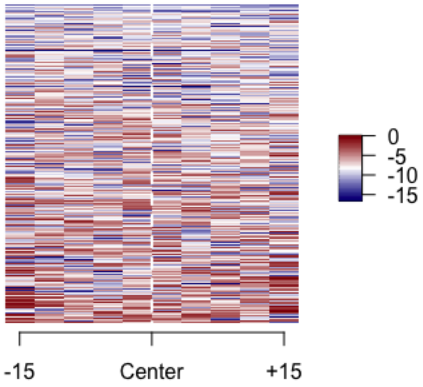


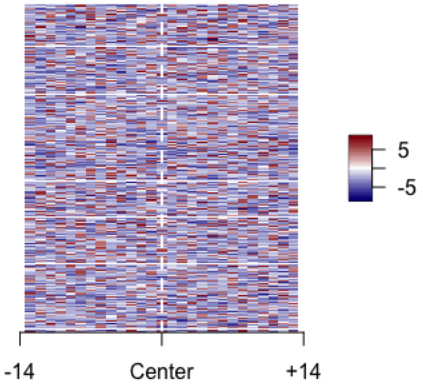
Anik Mitra
Optional HW

Unbound.fa









Anik Mitra
Optional HW