Justin Kim

BISC481

Assignment #3

2.

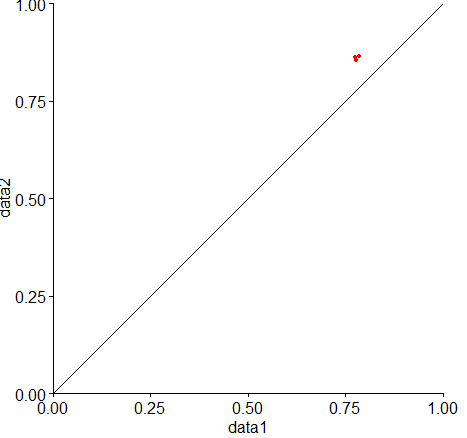
a.) In Vitro experiments such as SELEX-seq and PBM provide quantitative data on protein binding and affinity. SELEX-seq is a combinatorial chemistry technique that produces oligonucleotides of either single stranded DNA or RNA that specifically bind to a target ligand. PBM is a high throughput method used to track the interactions and activities of proteins in order to determine their function.

b.) In Vivo experiments such as ChIP-seq provide qualitative data on protein binding and tell us whether or not it is binding. ChIP is a method that analyzes protein interactions with DNA by combining chromatin immunoprecipitation with parallel DNA sequencing to identify the binding sites of DNA-associated proteins.

c.) An advantage of In Vivo experiments is that it is performed in the cell, which helps take into account the environment and conditions of the cell. However, this can also be a disadvantage because researches are then unable to control those conditions, leading to potentially skewed results and biases. On the other hand, In Vitro experiments are able to determine the definite DNA binding specificities of many transcription factors. A disadvantage is that the process takes place outside the cell, and thus, neglects the other factors that contribute to the DNA binding.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4.  b) Average R^2   |  |  |  | | --- | --- | --- | |  | 1mer + shape | 1 mer | | Mad | 0.8633932 | 0.775409 | | Max | 0.8641879 | 0.7854495 | | Myc | 0.8547479 | 0.7779904 | |  |  |  | |
| |  | | --- | |  | |

5. a) 1mer vs. 1mer+shape

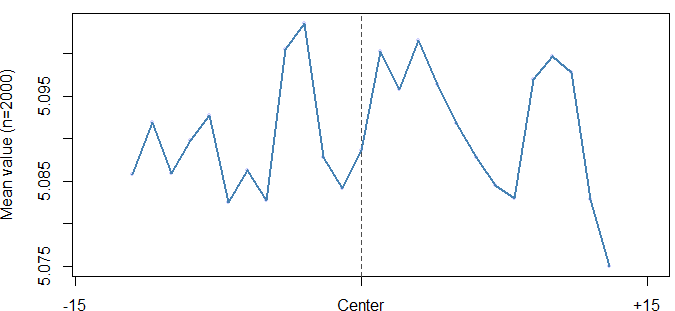


b.) Since the points are deviating from the line(model), this suggests that DNA shape does contribute to the binding. On the other hand, if the points were located on or closer to the line, that observation would suggest that shape is not a significant factor in binding.

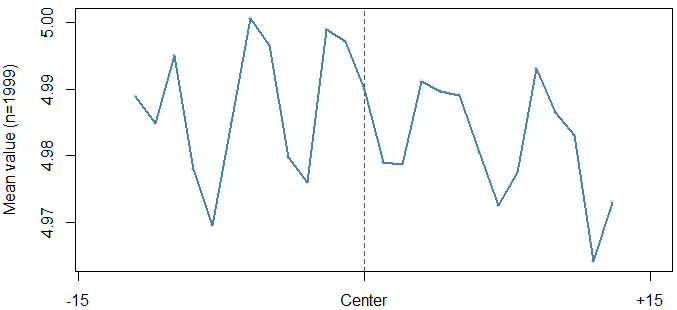
7. a)

Minor Groove Width

Bound:

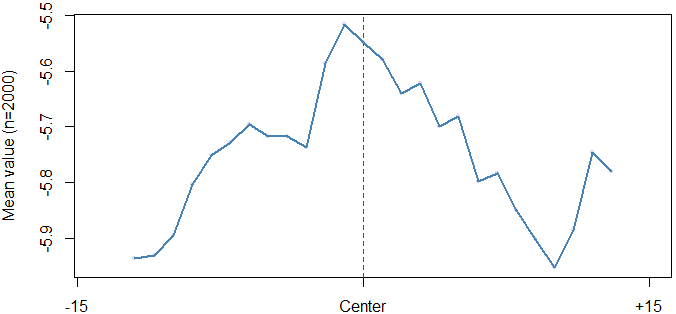


Nonbound:

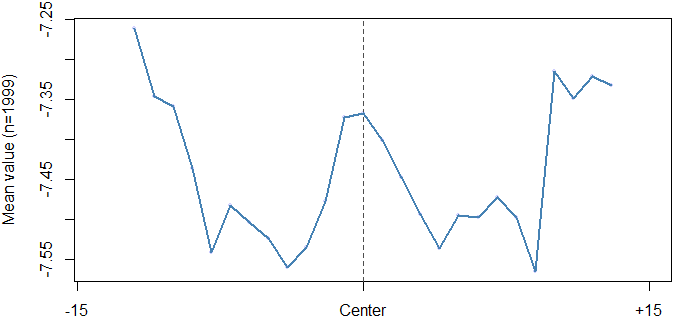


Propeller Twist

Bound:

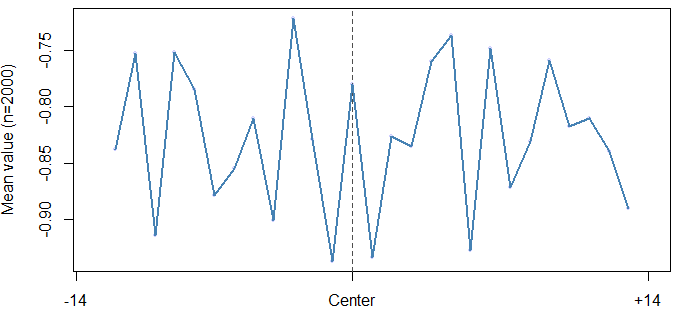


Nonbound:

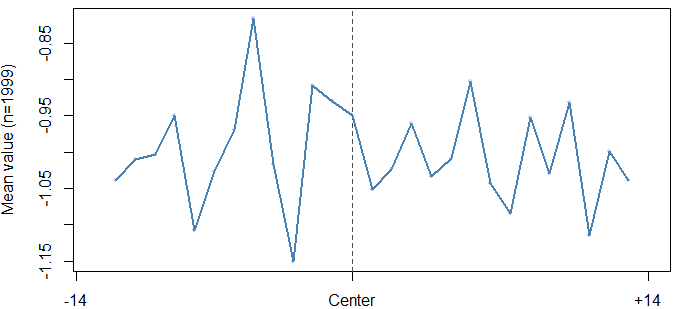


Roll

Bound:

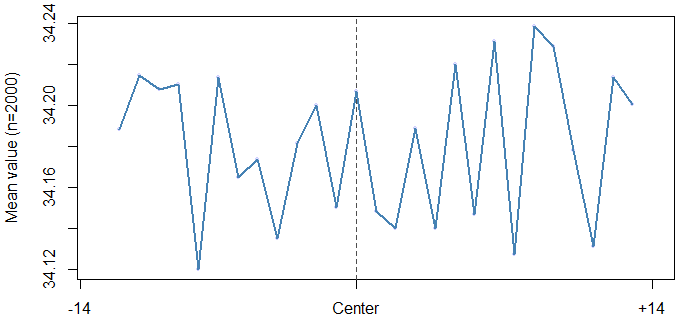


Nonbound:

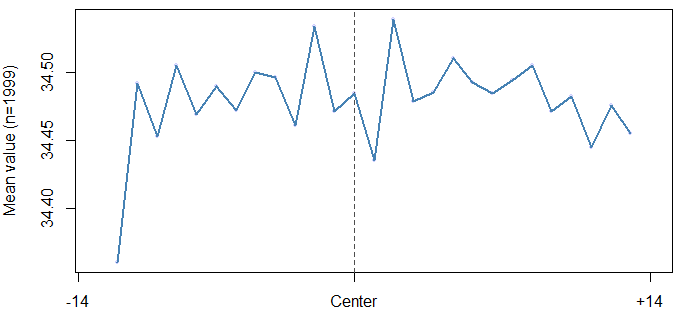


Helix Twist

Bound:

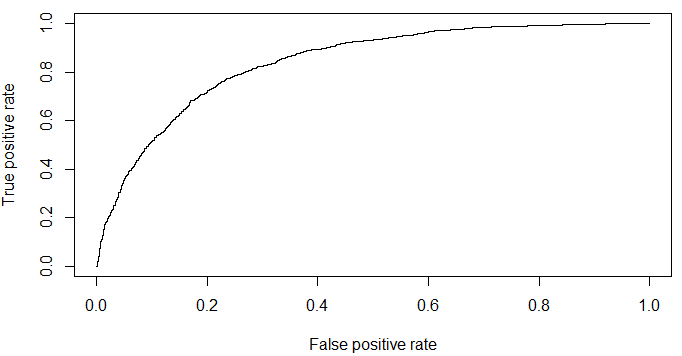


Nonbound:



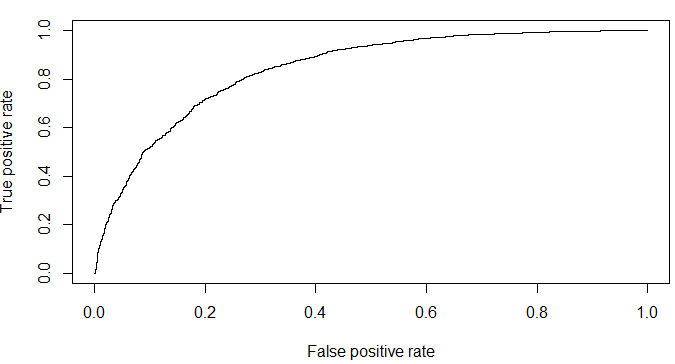
b) The bound and nonbound plots for the DNA shape parameters are generally similar, suggesting that these shape parameters do not play a significant factor in the binding affinity for the CTCR protein.

8. a) 1mer+shape



AUC Score: 0.83906

1mer



AUC Score: 0.8415583

b.) The ROC curve is created by plotting the true positive rate against the false positive rate. The AUC value is used in classification and determines the validation of the prediction result. A higher AUC value means that the overall performance of the test is better, and allows us to predict if the test is positive or negative. For our results, the AUC values were relatively similar suggesting that 1mer and 1mer+shape are not significantly different. This implies that we cannot distinguish if shape contributes significantly for protein binding in the CTCR protein.