Sophia Han

HW 3

2a. The advantages of in-vitro experiments SELEX-seq and PBM are that they are qualitative, done outside the cell in an environment that can be easily manipulated, and easy to use for large-scale production or experimentation. The disadvantages of these experiments are that they are not quantitative and may be difficult to perform outside of the cell and replicate in a test tube due to special cell environment conditions that could be required.

2b. The advantages of in-vivo experiments ChIP-seq are that they are quantitative and can be done in the cell, thus providing more information about the physical and chemical properties inside the cell. The disadvantages of these experiments are that they are not qualitative and may have too many parameters and variability inside the cell that cannot be tightly controlled.

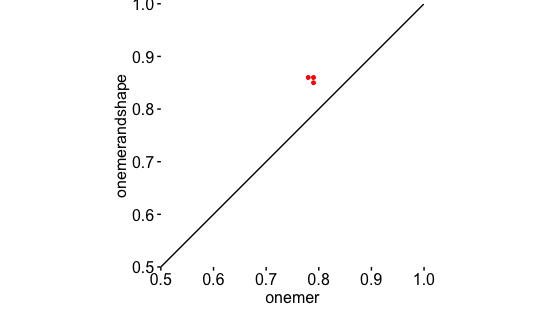
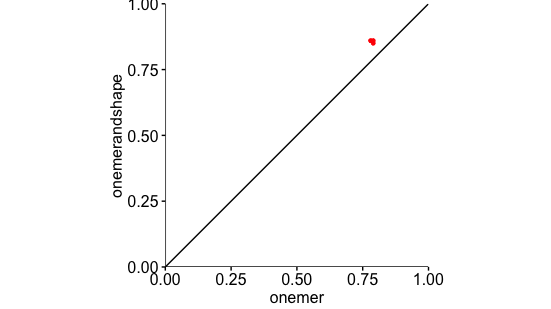
4. Average R2 values for these models with respect to Mad, Max, and Myc

- Mad – 1mer+shape (0.8631381), 1mer (0.7752976)

- Myc – 1mer+shape (0.8545169), 1mer (0.7780511)

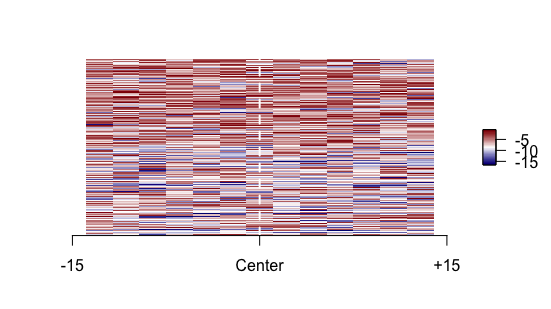
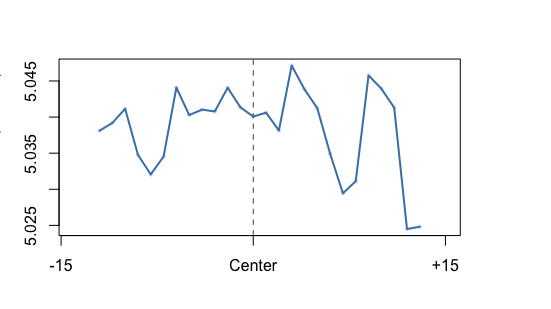
- Max – 1mer+shape (0.8643579), 1mer (0.7854098)

5a. Zoomed out plot of 1mer vs. 1mer+shape Zoomed in plot of 1mer vs. 1mer+shape



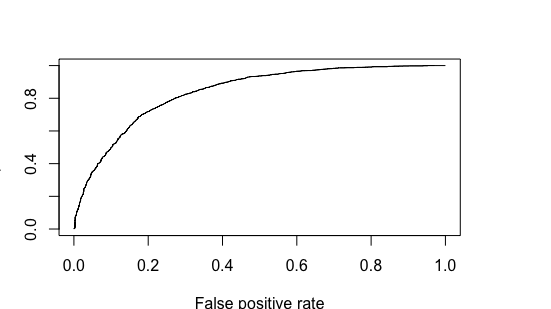
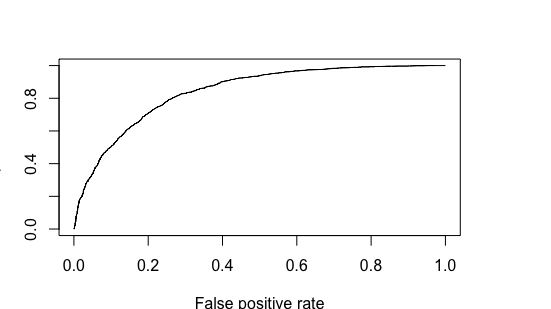
5b. The plot comparison of the two different 1mer and 1mer+shape models helps to visualize and learn the significance of the R2 values that correspond with each model. It provides a graphical representation of how close the values fall in comparison to the linear R2 =1 line and if the values fall either above or below this line. Through the graphs and R2 values, it is evident that shape and configuration are both important for protein binding as evidenced by the difference in values of around 0.86 for 1mer+shape and 0.78 for 1mer and its position above the curve.

7a. Plot Shape Heat Shape



7b. The plotShape and heatShape function plots of DNAshapeR provide in-vivo data analysis of the bound and nonbound data sets. It is evident through the plots that these two data sets are fairly similar, thus suggesting that the different shape parameters do not play a great role in the binding affinity of the transcription factor protein.

8a. 1mer (AUC = 0.8411668) 1mer + shape (AUC = 0.8393292)



8b. Both the 1mer and 1mer+shape feature models have similar logistic regression plots and similar AUC scores of around 0.84. Because the AUC score is fairly high and close to 1, it is considered to be good at predicting the outcome for the model for in-vitro data used. In general, it can be assumed that the prediction model would be good at separating the data from these two different groups, 1mer and 1mer+shape.