Analyzing shape (mgw) across MEF2A datasets

It is known that MEF2A binds to a YTA(A/T)4TAR, and since shape readout seemed to be important from structural studies, we hypothesized that the shape at the central region could be important for achieving binding specificity. Given the dataset from Jaspar and the raw ChIP-seq data from the Husdon Lab, I wanted to investigate the DNA shape binding preferences of MEF2A at that central region, since that would be a first step to lead to designing some sequences that could be tested.

The MEF2A data from Jaspar (top) was plotted just by binning every 10 sequences. The y axis here is NOT binding affinity because that information was not available from Jaspar. For the raw ChIP-seq data (bottom), I used narrowpeak files, got sequences from galaxy based on genomic coordinates, and then searched for the regions containing the mef2a motif (MA0052.1) using FIMO and a default p-value of 1e-4, getting 6393 motifs with flanks.

For the Jaspar data (top panel), there is a wide range of shape variations in the central region, which could be implicated with binding affinity. However, when I analyzed the ChIP-seq data (bottom panel), no clear differences between the higher binding vs lower binding affinities were observed. This could be due to the p-value cutoff used on FIMO, which allows for more sequences to actually be selected, and also from more noise present in the in vivo experiment. Binning into larger groups of sequences did not change the shape pattern at the central region.