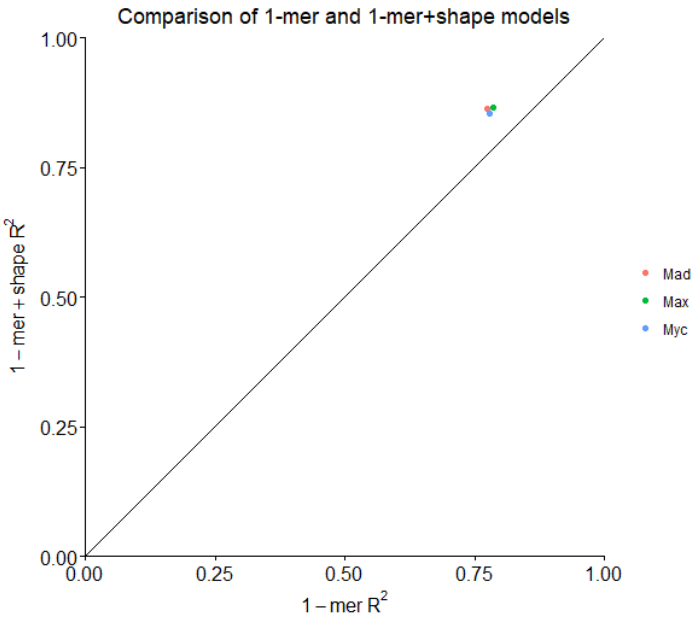
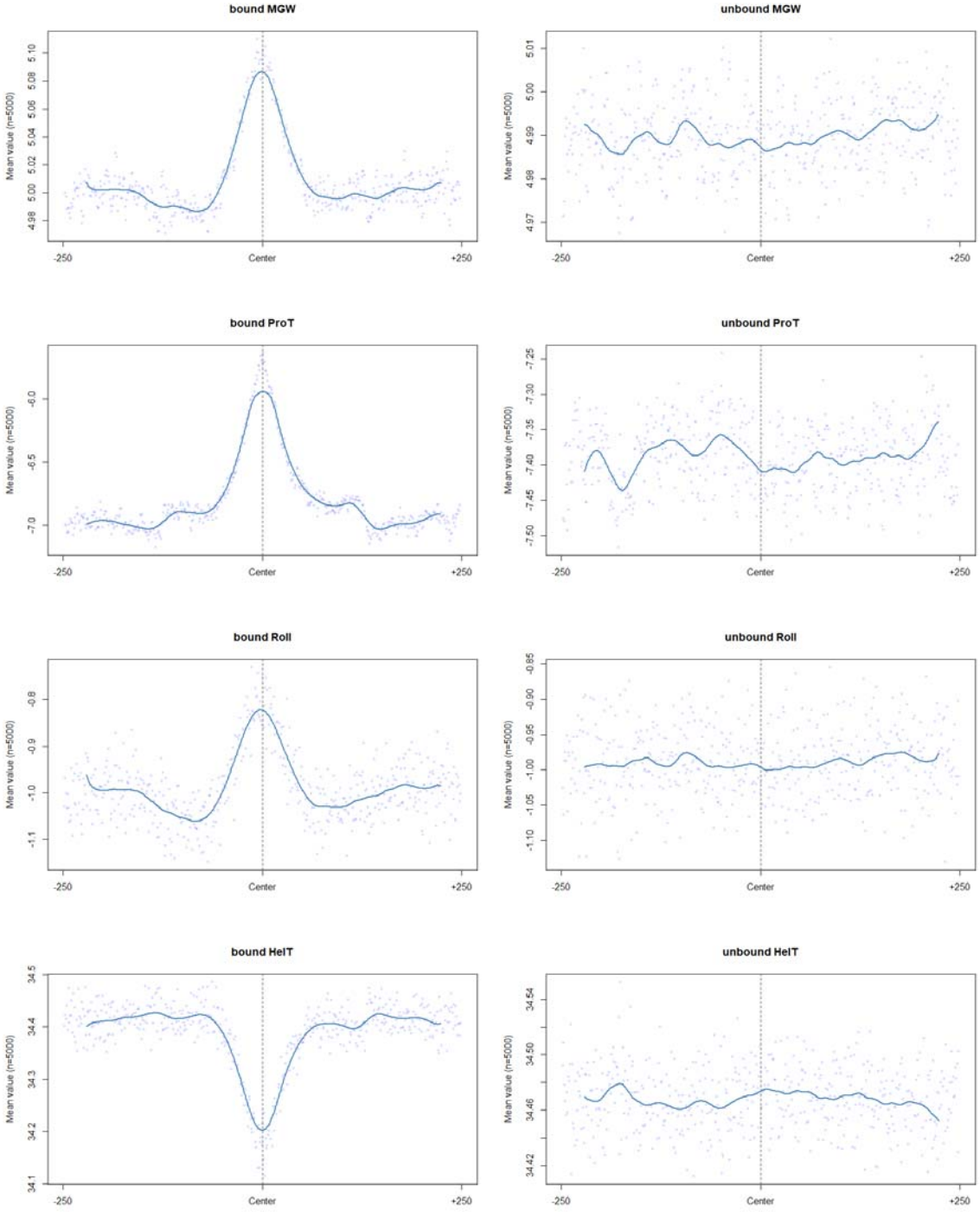
2. a. SELEX-seq: A DNA library is exposed to purified TF. Then, bound and unbound DNA sequences are separated. The bound fraction can be amplified and the process can be repeated to select only high affinity sites, which can then be sequenced.
   PBM: All possible 10 base long binding sites are used on one array. A TF is added to the array. The amount of protein binding to any sequence is quantified with a fluorescent antibody against that protein.
   b. ChIP-seq: Crosslinked DNA-protein complexes are sheared, followed by detection of target protein using an antibody attached to a bead. The complex is precipitated, the protein is unlinked, and the DNA is sequenced.
   c. SELEX-seq can determine important aspects of binding specificity (such as consensus sequence). It also has large dynamic range and allow binding sites of any length. Requires purified TFs. Low specificity TFs may require multiple rounds of selection.
   PBM tends to have higher backgrounds but still produces strong signals. PBM also requires dsDNA and purified TFs. Equipment is commonly available for microarrays, so PBM is cost effective.
   ChIP-seq has relatively low resolution of binding sites (within 100 bp). However, ChIP-seq can provide added information: for example, if a high affinity binding site is not bound by a TF, then that region of the genome may be inaccessible to the TF due to some other factors like chromatin organization.
4. R^2 values for L2-regularized models for "1-mer" and "1-mer+shape" features

	1-mer	**1-mer+shape**
Mad	0.7168	0.7328
Max	0.7028	0.7206
Myc	0.7208	0.7303

5. Since the R^2 values for 1-mer+shape are generally higher than that of 1-mer, we can see that including shape produces better predictions for affinity.



7. The data suggest that the protein tends to bind to sites with larger than average minor groove width, propeller twist, and roll. Additionally, the protein tends to bind to sites with smaller than average helix twist.



8. Adding shape to the model does not appear to improve prediction using ChIP-seq data.