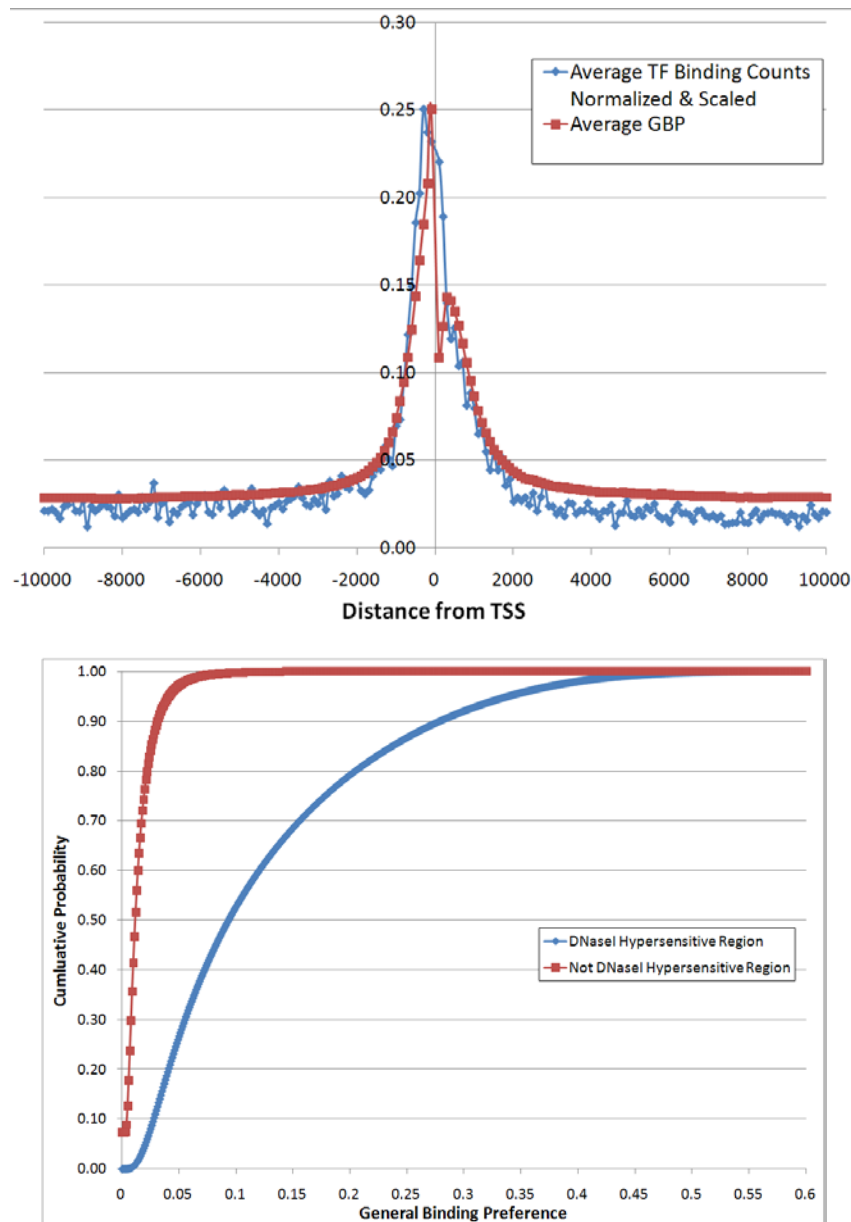
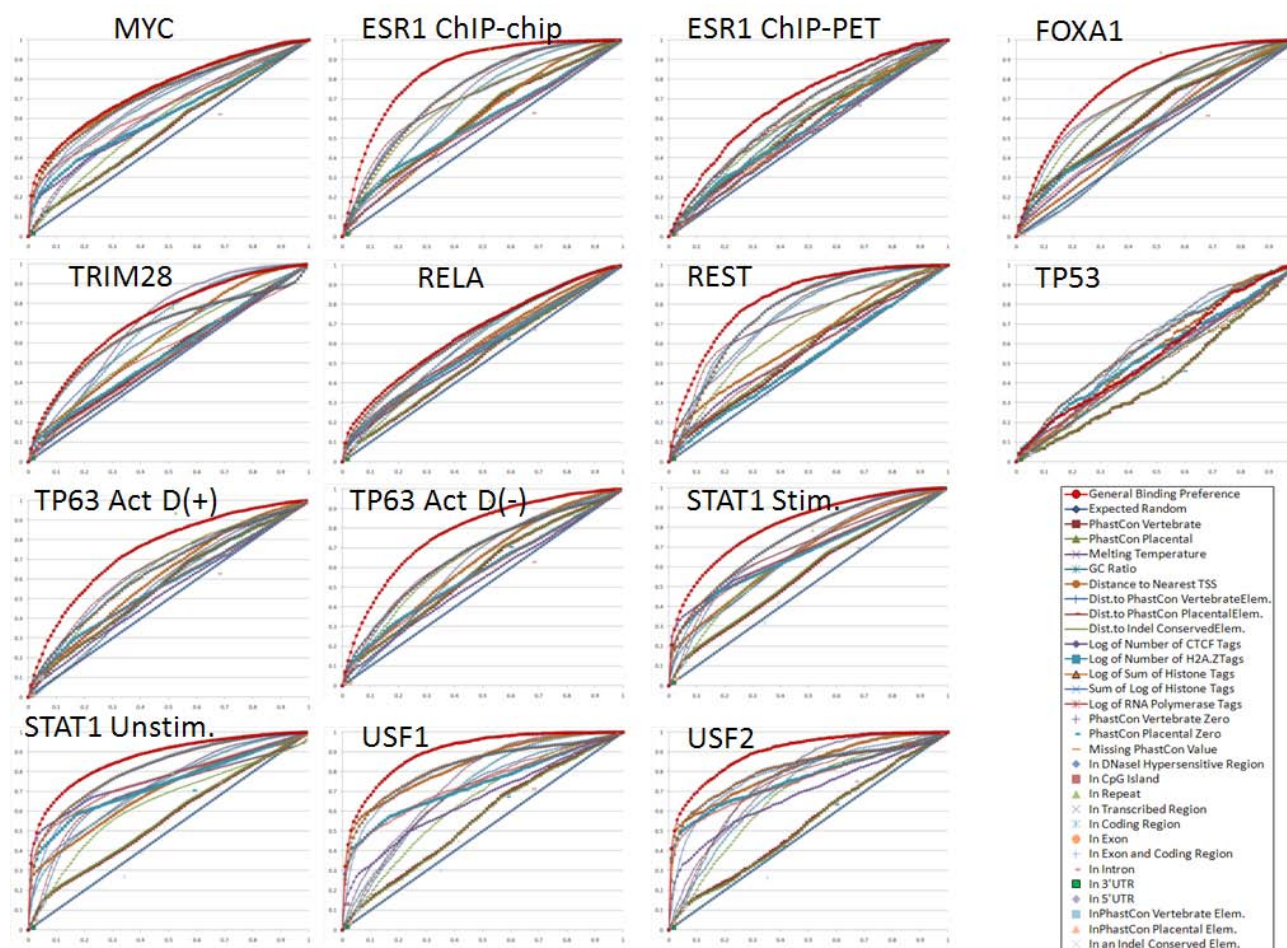


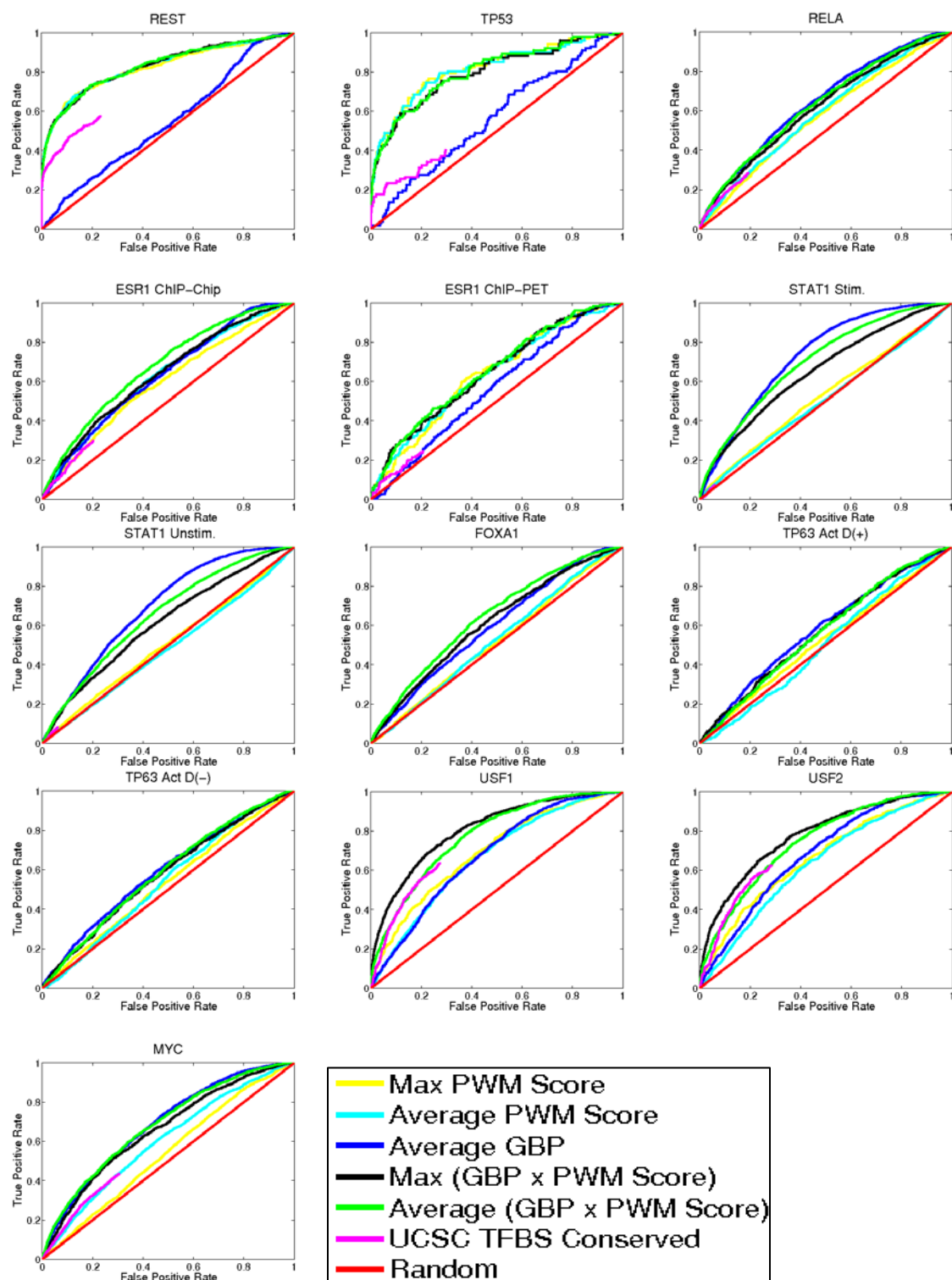
Supplementary Figures: Integrating Multiple Evidence Sources to Predict Transcription Factor Binding in the Human Genome



Supplementary Figure 1: Properties of the General Binding Preference: (Top) The plot shows the average GBP score at each 100bp interval distance relative to the transcription start site. The plot also shows the TF counts in each interval after normalizing by the total number of sites in an experiment, then averaging over experiments, and finally linearly scaling to have the same maximum as the GBP. This shows overall good agreement in the shape between the two. (Bottom) compares the cumulative distribution for DNaseI hypersensitive and not DNaseI hypersensitive regions as a function of general binding preference.

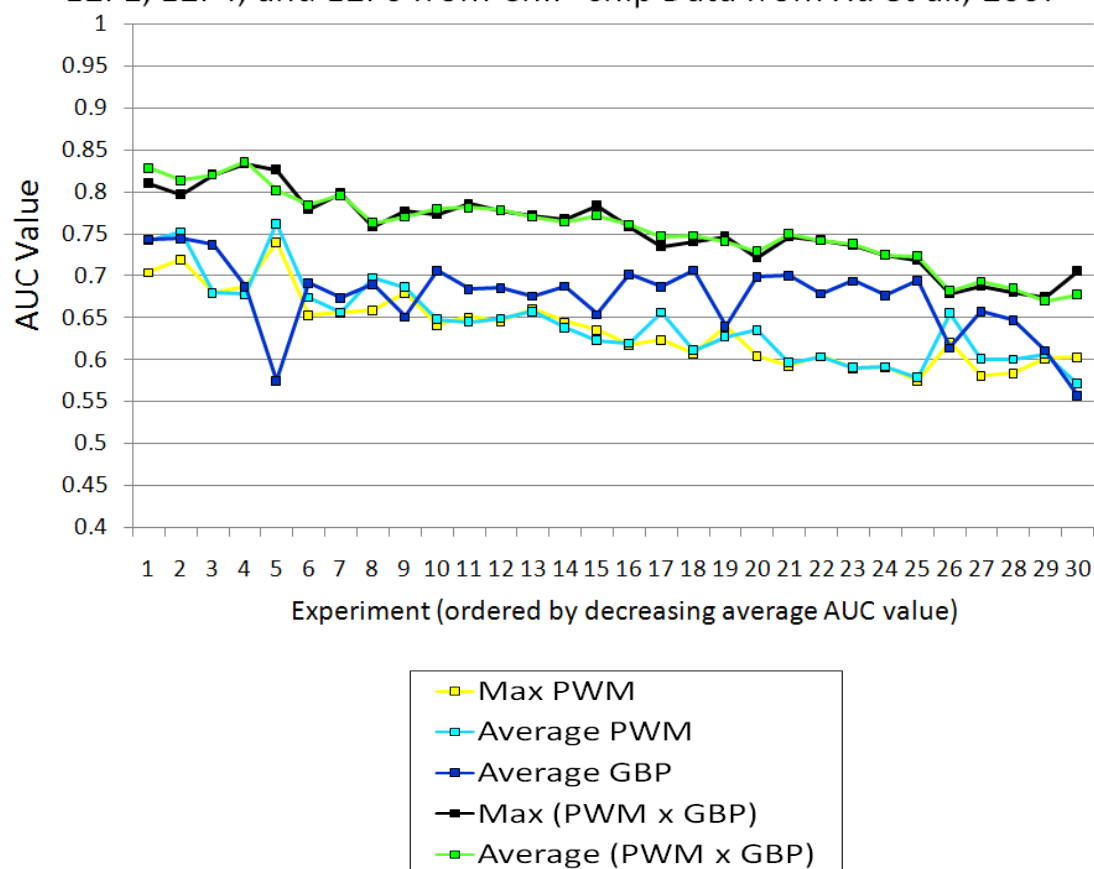


Supplementary Figure 2: This is an extended version of Figure 2 showing ROC values for each individual feature considered. The x-axis is the false positive rate and the y-axis is the true positive rate.

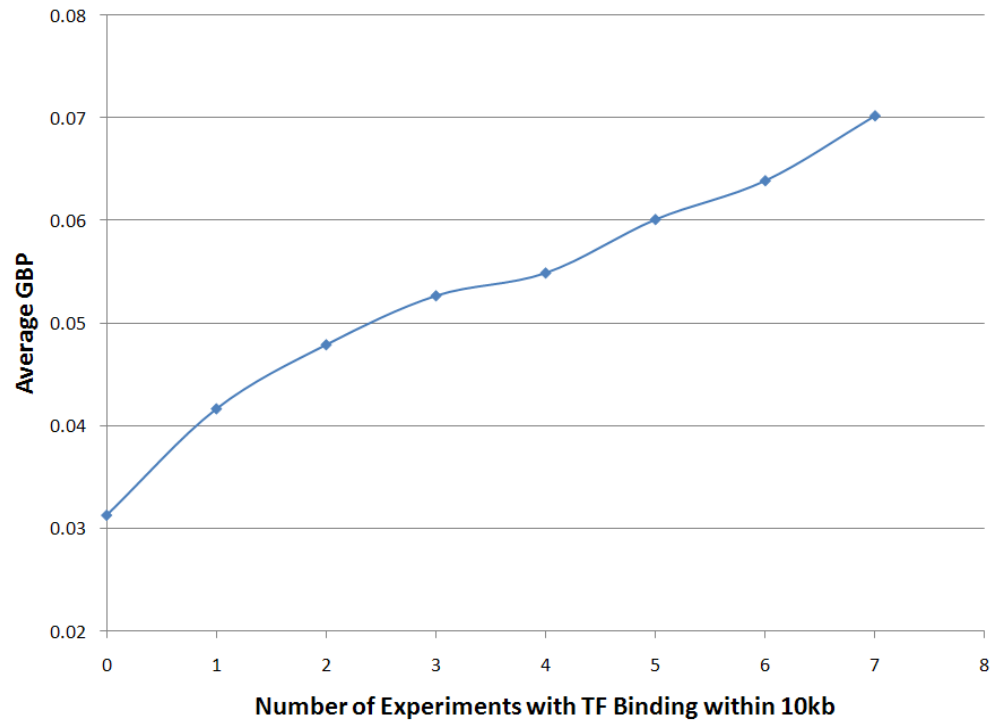


Supplementary Figure 3: ROC Curves for Identifying 20KB TSS Centered Regions Bound by a Transcription Factor. Here we compare on 13 data sets methods to prediction whether a transcription start site will have the transcription factor reported to bind within 10,000 bases. AUC values for these curves and curves of additional methods can be found in Supplementary Table 4.

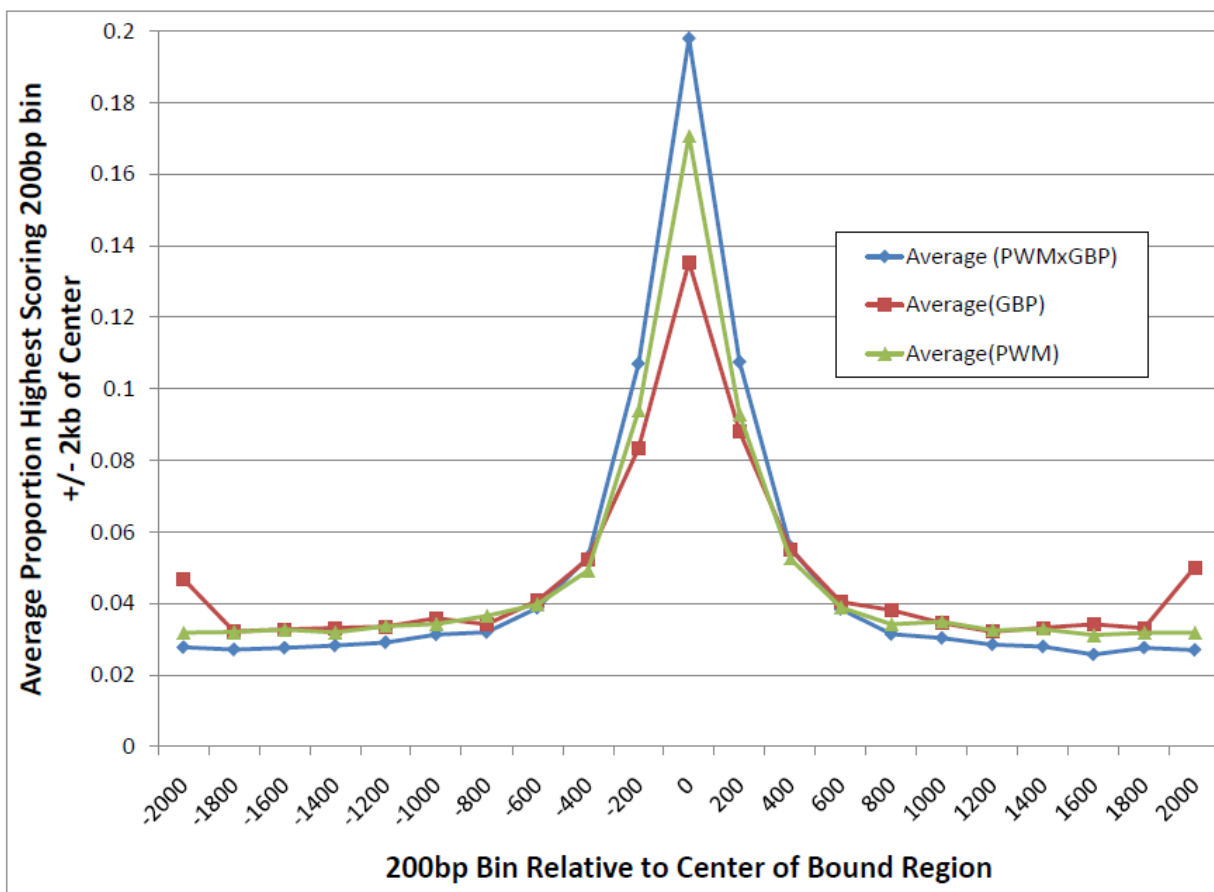
Comparison of Methods in Predicting Bound Promoters in E2F1, E2F4, and E2F6 from ChIP-chip Data from Xu et al., 2007



Supplementary Figure 4: Results for Predicting Targets of the E2F Family of Transcription Factors. The chart shows a comparison of five methods for the task of predicting gene targets of the E2F family of transcription factors based on 30 different ChIP-chip experiments (Xu et al., 2007). For each method, an AUC value was computed for each of the 30 experiments. The y-axis shows the AUC value, the x-axis shows the experiment ordered by decreasing average AUC values across the five methods. The plot shows that the methods that use jointly the prior and motif have a higher AUC value at each rank position than methods that only use the prior or only the motif information, thus performing better in our evaluation.



Supplemental Figure 5: GBP Increases with the Number of Genome-wide Experiments with Report Binding within 10kb of the TSS. x-axis shows the number of experiments with reported TF binding within 10kb of the TSS, and the y-axis reports the average GBP of the region.



Supplementary Figure 6: Peak center agrees with local predictions of most likely binding location. For the center of each reported bound region in the genome-wide experiments, the 200bp bin predicted most likely to contain the source of the TF binding out of all bins within 2kb of the reported center based on the Average (PWMxGBP), Average(GBP), and Average(PWM) prediction methods was computed. The proportion assigned to each of the 21 bins was then determined for each experiment, and the resulting graph is averaged over 13 experiments.