

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

Feature	MYC	ESR1 ^c	ESR1 ^p	FOXA1	TRIM28	RELA	REST	TP53	TP63 ⁺	TP63 ⁻	STAT1 ^s	STAT1 ^u	USF1	USF2	Average
1. PhastCon Vertebrate	0.54	0.60	0.56	0.62	0.55	0.53	0.56	0.43	0.57	0.58	0.56	0.56	0.56	0.54	0.55
2. PhastCon Placental	0.55	0.61	0.57	0.63	0.52	0.54	0.57	0.44	0.57	0.58	0.57	0.57	0.56	0.53	0.56
3. Conservation Vertebrate Zero (r)	0.53	0.57	0.55	0.58	0.53	0.52	0.55	0.43	0.55	0.56	0.55	0.55	0.55	0.52	0.54
4. Conservation Placental Zero (r)	0.54	0.58	0.56	0.58	0.52	0.52	0.55	0.43	0.56	0.56	0.55	0.56	0.55	0.52	0.54
5. Missing Conservation Value (r)	0.50	0.52	0.51	0.51	0.50	0.50	0.52	0.49	0.51	0.51	0.50	0.48	0.51	0.50	0.50
6. InPhastCon Vertebrate Elem.	0.53	0.55	0.51	0.56	0.53	0.52	0.52	0.50	0.53	0.54	0.53	0.55	0.52	0.53	0.53
7. InPhastCon Placental Elem.	0.53	0.55	0.52	0.57	0.52	0.52	0.53	0.50	0.53	0.53	0.54	0.55	0.52	0.53	0.53
8. In an Indel Conserved Elem.	0.53	0.56	0.52	0.58	0.54	0.52	0.54	0.50	0.54	0.55	0.54	0.55	0.53	0.54	0.54
9. Dist.to PhastConVertebrateElem. (r)	0.62	0.70	0.60	0.71	0.63	0.58	0.71	0.51	0.65	0.66	0.67	0.69	0.69	0.74	0.65
10. Dist.to PhastConPlacentalElem. (r)	0.63	0.71	0.61	0.72	0.56	0.59	0.72	0.50	0.66	0.66	0.68	0.70	0.68	0.73	0.65
11. Dist.to Indel ConservedElem. (r)	0.59	0.68	0.60	0.70	0.60	0.57	0.67	0.49	0.64	0.65	0.63	0.63	0.64	0.68	0.63
12. Melting Temperature	0.70	0.70	0.59	0.55	0.70	0.62	0.74	0.60	0.57	0.62	0.67	0.73	0.75	0.76	0.66
13. GC Ratio	0.69	0.68	0.58	0.54	0.67	0.62	0.74	0.59	0.55	0.60	0.66	0.72	0.71	0.73	0.65
14. In CpG Island	0.58	0.50	0.49	0.50	0.51	0.53	0.54	0.51	0.50	0.50	0.56	0.61	0.55	0.58	0.53
15. In Repeat (r)	0.58	0.72	0.60	0.71	0.63	0.55	0.68	0.45	0.71	0.71	0.64	0.64	0.70	0.70	0.64
16. In Transcribed Region	0.57	0.53	0.50	0.54	0.50	0.52	0.51	0.50	0.54	0.54	0.51	0.50	0.52	0.50	0.52
17. In Coding Region (r)	0.47	0.48	0.50	0.47	0.51	0.49	0.51	0.50	0.48	0.48	0.51	0.53	0.53	0.54	0.50
18. In Exon	0.51	0.49	0.48	0.50	0.51	0.51	0.50	0.50	0.50	0.50	0.51	0.51	0.50	0.51	0.50
19. In Exon and Coding Region (r)	0.50	0.51	0.51	0.50	0.49	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
20. In Intron (r)	0.47	0.48	0.50	0.47	0.52	0.49	0.51	0.50	0.48	0.48	0.51	0.54	0.53	0.54	0.50
21. In 3'UTR	0.50	0.50	0.49	0.50	0.50	0.50	0.50	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.50
22. In 5'UTR	0.54	0.51	0.50	0.51	0.51	0.52	0.51	0.51	0.52	0.52	0.53	0.53	0.54	0.55	0.52
23. Distance to Nearest TSS (r)	0.74	0.59	0.54	0.56	0.62	0.59	0.62	0.56	0.61	0.62	0.67	0.70	0.83	0.82	0.65
24. In DNaseI Hypersensitive Region	0.61	0.52	0.50	0.53	0.54	0.56	0.55	0.52	0.54	0.54	0.64	0.72	0.71	0.75	0.59
25. Log of Number of CTCF Tags	0.61	0.55	0.52	0.55	0.52	0.56	0.57	0.53	0.53	0.53	0.68	0.77	0.66	0.69	0.59
26. Log of Number of H2A.ZTags	0.62	0.58	0.55	0.58	0.55	0.57	0.52	0.55	0.57	0.57	0.67	0.73	0.73	0.76	0.61
27. Log of Sum of Histone Tags	0.73	0.74	0.64	0.68	0.67	0.65	0.77	0.61	0.64	0.67	0.75	0.82	0.82	0.81	0.71
28. Sum of Log of Histone Tags	0.71	0.75	0.64	0.68	0.66	0.65	0.77	0.60	0.64	0.67	0.74	0.81	0.81	0.79	0.71
29. Log of RNA Polymerase Tags	0.66	0.56	0.53	0.57	0.53	0.59	0.56	0.52	0.57	0.57	0.67	0.76	0.74	0.76	0.61
GBP	0.75	0.84	0.69	0.79	0.72	0.67	0.82	0.56	0.75	0.78	0.81	0.87	0.90	0.90	0.78

Supplementary Table 1: Comparison of the GBP with Individual Features for Each Data Set.

The table shows the AUC values when using each feature individually, and at the bottom the GBP when integrating the data sources using our method. Both ranking sites by the feature in increasing and decreasing order was considered with the value shown from the order which resulted in a larger AUC value across all datasets. If there is a (r) after the feature name, then a larger AUC value was obtained when the sites were ranked in increasing order of the feature. Ties in feature values were broken randomly. The highest value in each column is in red, while the second highest is in yellow. For 13 of the 14 data sets the AUC value of our method was highest. The last column is the average AUC value for the feature over all data sets, where all sites must be consistently ranked by the feature in either increasing or decreasing order, whichever gave a higher AUC value. Superscripts of 'c' and 'p' are used to specify the ChIP-chip or ChIP-PET experiment respectively for ERα, and superscripts of 's' and 'u' are used to specify the stimulated or unstimulated experiment respectively for STAT1.

Type of Features Excluded	MYC	ESR1 ^c	ESR1 ^p	FOXA1	TRIM28	RELA	REST	TP53	TP63 ⁺	TP63 ⁻	STAT1 ^s	STAT1 ^u	USF1	USF2	Average
1. Conservation (1-11)	.75	.82	.68	.76	.74	.66	.82	.56	.75	.77	.80	.86	.91	.90	.77
2. CG, CpG, Melting (12-14)	.74	.82	.68	.80	.70	.66	.81	.55	.75	.77	.81	.87	.89	.89	.77
3. Repeat Element (15)	.76	.81	.68	.75	.70	.67	.81	.60	.71	.74	.81	.87	.89	.89	.76
4. Refseq Annotations (16-23)	.73	.84	.68	.79	.72	.66	.82	.55	.74	.77	.81	.87	.88	.89	.77
5. CD4T Experimental Features (24-29)	.73	.82	.66	.76	.73	.64	.80	.52	.76	.78	.76	.78	.88	.89	.75
6. CD4T Experimental Features (24-29) replaced with ES Experimental Features	.73	.83	.66	.77	.74	.64	.80	.53	.75	.77	.77	.80	.89	.90	.76
7. No Features Excluded	.75	.84	.69	.79	.72	.67	.82	.56	.75	.78	.81	.87	.90	.90	.78

Supplementary Table 2: Impact on Removing Sets of Features on the AUC Value. This table compares the cross-validation AUC values that would result when removing selected groups of features as compared to all the features considered. The numbers in parenthesis in the first column indicate the features from Supplementary Table 1 that were removed. From this table we see that we had the largest drop when removing all the experimental features. We also observe that after removing any of these sets of features we are still able to achieve on average a higher AUC value than when working with any single feature. We also compare to the case in which features 24-29 based on CD4T data removed, but then replaced with features 27 and 28 now computed based on the three histone methylations profiled in human ES cells available from the supporting website of (Ku et al, 2008).

Regulator	Motif ID	Motif Accession
TP53	V\$P53_01	M00034
REST	V\$NRSF_Q4	M01028
USF1	V\$USF_Q6_01	M00796
USF2	V\$USF_Q6_01	M00796
TP63	V\$P53_DECAMER_Q2	M00761
MYC	V\$MYC_Q2	M00799
STAT1	V\$STAT_01	M00223
FOXA1	V\$HNF3ALPHA_Q6	M00724
ESR1	V\$ER_Q6	M00191
RELA	V\$NFKAPPAB_01	M00054

Supplementary Table 3: Table of Transfac Motifs. This table provides the identifying information for the motifs from the TRANSFAC database that we used in the analysis.

Data Set	# of Bound Regions	Max PWM Score	Average PWM Score	Max (NucPos x PWM Score)	Average (NucPos x PWM Score)	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBPx PWM Score)
MYC	2671	0.54	0.60	0.51	0.56	0.64	0.68	0.66	0.68
ESR1 ^c	1150	0.59	0.63	0.58	0.62	0.54	0.63	0.64	0.68
ESR1 ^p	209	0.64	0.64	0.60	0.63	0.50	0.56	0.65	0.65
FOXA1	2266	0.52	0.52	0.51	0.51	0.56	0.59	0.61	0.64
RELA	1660	0.57	0.59	0.56	0.58	0.60	0.64	0.62	0.63
REST	720	0.83	0.83	0.82	0.83	0.46	0.55	0.83	0.83
TP53	102	0.82	0.82	0.82	0.82	0.56	0.57	0.80	0.80
TP63 ⁺	911	0.53	0.50	0.54	0.54	0.56	0.58	0.57	0.57
TP63 ⁻	1487	0.53	0.54	0.53	0.55	0.56	0.59	0.58	0.59
STAT1 ^s	11504	0.53	0.51	0.50	0.46	0.77	0.73	0.65	0.70
STAT1 ^u	5399	0.51	0.48	0.48	0.45	0.73	0.69	0.61	0.65
USF1	1697	0.69	0.67	0.65	0.66	0.70	0.67	0.81	0.77
USF2	1099	0.67	0.63	0.62	0.63	0.71	0.68	0.78	0.75
Average		0.61	0.61	0.59	0.60	0.61	0.63	0.68	0.69

Supplementary Table 4: Table of AUC Values for Predicting if a RefSeq Transcription

Start Site has a Transcription Factor binding site within 10,000 bases. The highest value in each row is shown in red, while the second highest if there was no tie for the highest is in yellow. In the bottom row we show the average AUC values for all 13 test cases. The table compares AUC values for the Max and Average based on just using the motif, just using GBP, combining GBP with a PWM score, and combining the PWM score weighted by (1-predicted probability of an occupied nucleosome) (NucPos) (Kaplan et al, 2009; obtained from http://genie.weizmann.ac.il/software/nucleo_genomes.html). The GBP combined with motif information had the highest average score.

Background Model	Interval +/- around TSS	Max PWM Score	Average PWM Score	Max (Nuc Pos x PWM Score)	Average (Nuc Pos x PWM Score)	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBPx PWM Score)
Local 0	2000	0.64	0.64	0.62	0.63	0.63	0.64	0.71	0.73
Local 0	5000	0.60	0.61	0.59	0.59	0.61	0.63	0.68	0.70
Local 0	10000	0.58	0.59	0.57	0.56	0.61	0.63	0.66	0.68
Global 0	2000	0.65	0.64	0.64	0.64	0.63	0.64	0.71	0.72
Global 0	5000	0.62	0.62	0.61	0.61	0.61	0.63	0.69	0.70
Global 0	10000	0.61	0.61	0.59	0.60	0.61	0.63	0.68	0.69
Global 1	2000	0.62	0.63	0.62	0.62	0.63	0.64	0.67	0.68
Global 1	5000	0.60	0.61	0.59	0.60	0.61	0.63	0.66	0.67
Global 1	10000	0.60	0.60	0.58	0.59	0.61	0.63	0.65	0.66
Global 2	2000	0.63	0.64	0.63	0.64	0.63	0.64	0.68	0.69
Global 2	5000	0.61	0.62	0.60	0.61	0.61	0.63	0.67	0.68
Global 2	10000	0.60	0.61	0.59	0.60	0.61	0.63	0.66	0.67
Global 3	2000	0.64	0.65	0.63	0.65	0.63	0.64	0.69	0.70
Global 3	5000	0.62	0.63	0.60	0.62	0.61	0.63	0.67	0.68
Global 3	10000	0.60	0.61	0.59	0.60	0.61	0.63	0.66	0.68

Supplementary Table 5: Average AUC of the methods for the same 13 test cases in Supplementary Table 3 repeated with intervals +/-2kb, 5kb, and 10kb around the TSS and using global 0th, 1st, 2nd, and 3rd order background models and a local 0th order background model estimated based on the nucleotide frequency within the interval. In all cases combining GBP with PWM scores gives the highest average AUC values.

Background Model	Interval +/- around TSS	Max PWM Score	Average PWM Score	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBPx PWM Score)
Local 0	2000	0.18	0.19	0.08	0.10	0.25	0.25
Local 0	5000	0.16	0.16	0.07	0.09	0.22	0.23
Local 0	10000	0.13	0.14	0.08	0.10	0.19	0.20
Global 0	2000	0.19	0.19	0.08	0.10	0.23	0.23
Global 0	5000	0.17	0.15	0.07	0.09	0.21	0.21
Global 0	10000	0.15	0.14	0.08	0.10	0.20	0.19
Global 1	2000	0.15	0.14	0.08	0.10	0.17	0.17
Global 1	5000	0.13	0.13	0.07	0.09	0.15	0.16
Global 1	10000	0.13	0.12	0.08	0.10	0.15	0.15
Global 2	2000	0.16	0.16	0.08	0.10	0.18	0.18
Global 2	5000	0.14	0.14	0.07	0.09	0.16	0.17
Global 2	10000	0.13	0.13	0.08	0.10	0.15	0.16
Global 3	2000	0.17	0.17	0.08	0.10	0.19	0.20
Global 3	5000	0.15	0.15	0.07	0.09	0.17	0.18
Global 3	10000	0.14	0.13	0.08	0.10	0.16	0.17

Supplementary Table 6: Average normalized AUC of methods up to a maximum false positive rate of 10% for various background models and intervals around the TSS. The AUC values were normalized by multiplying by 10 so that a perfect AUC value would be 1, and thus an AUC value of 0.05 would be expected by chance. Again we see the highest average scores for combining the GBP with the PWM score.

Data Set	# of Bound Bins Exclude +/- 2bins	Max PWM Score	Average PWM Score	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBPx PWM Score)
MYC	4287	0.66	0.70	0.77	0.77	0.76	0.78
ESR1 ^c	5779	0.69	0.71	0.80	0.81	0.77	0.80
ESR1 ^p	1231	0.70	0.71	0.68	0.69	0.74	0.75
FOXA1	12896	0.65	0.65	0.73	0.75	0.76	0.78
RELA	5852	0.60	0.61	0.66	0.67	0.65	0.66
REST	1930	0.92	0.92	0.80	0.81	0.93	0.94
TP53	542	0.86	0.86	0.58	0.58	0.87	0.87
TP63 ⁺	3664	0.58	0.61	0.72	0.73	0.68	0.71
TP63 ⁻	5783	0.60	0.64	0.74	0.76	0.71	0.74
STAT1 ^s	41428	0.60	0.60	0.80	0.80	0.75	0.78
STAT1 ^u	10971	0.54	0.54	0.86	0.85	0.74	0.78
USF1	2506	0.80	0.83	0.90	0.91	0.89	0.91
USF2	1342	0.78	0.82	0.90	0.91	0.88	0.91
Average		0.69	0.71	0.76	0.77	0.78	0.80

Supplementary Table 7: AUC for predicting 1kb bins of the genome containing the center of a reported binding site. This was conducted genome-wide excluding gapped portions of the genome.

(a)

Data Set	# of Bound Bins Exclude +/- 2 bins	Max PWM Score	Average PWM Score	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBPx PWM Score)
MYC	3052	0.65	0.68	0.70	0.71	0.71	0.73
ESR1 ^c	5399	0.68	0.71	0.80	0.82	0.76	0.80
ESR1 ^p	1178	0.70	0.71	0.69	0.70	0.74	0.75
FOXA1	12126	0.65	0.65	0.74	0.75	0.76	0.78
RELA	5166	0.59	0.60	0.64	0.64	0.62	0.63
REST	1580	0.91	0.92	0.78	0.79	0.93	0.93
TP53	511	0.86	0.86	0.58	0.57	0.87	0.87
TP63 ⁺	3295	0.58	0.60	0.71	0.72	0.67	0.70
TP63 ⁻	5160	0.60	0.63	0.73	0.75	0.70	0.73
STAT1 ^s	32868	0.61	0.61	0.77	0.77	0.72	0.75
STAT1 ^u	7585	0.53	0.53	0.81	0.80	0.68	0.72
USF1	1207	0.82	0.85	0.84	0.85	0.87	0.90
USF2	630	0.75	0.78	0.82	0.84	0.83	0.85
Average		0.69	0.70	0.74	0.75	0.76	0.78

(b)

Data Set	# of Bound Bins Exclude +/- 5 bins	Max PWM Score	Average PWM Score	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBPx PWM Score)
MYC	2815	0.65	0.68	0.70	0.70	0.71	0.73
ESR1 ^c	5222	0.69	0.71	0.81	0.82	0.77	0.80
ESR1 ^p	1148	0.70	0.71	0.69	0.70	0.74	0.75
FOXA1	11816	0.65	0.65	0.74	0.75	0.76	0.78
RELA	4993	0.59	0.60	0.64	0.64	0.62	0.63
REST	1515	0.91	0.92	0.78	0.79	0.93	0.93
TP53	498	0.86	0.86	0.58	0.57	0.87	0.86
TP63 ⁺	3196	0.58	0.61	0.71	0.72	0.67	0.70
TP63 ⁻	5010	0.60	0.64	0.74	0.75	0.70	0.73
STAT1 ^s	31685	0.61	0.61	0.77	0.77	0.71	0.75
STAT1 ^u	7324	0.53	0.53	0.81	0.80	0.67	0.71
USF1	1105	0.82	0.85	0.84	0.85	0.88	0.90
USF2	596	0.75	0.78	0.82	0.84	0.82	0.85
Average		0.69	0.70	0.74	0.75	0.76	0.78

(c)

Data Set	# of Bound Bins +/- 10bins	Max PWM Score	Average PWM Score	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBPx PWM Score)
MYC	2568	0.65	0.68	0.69	0.70	0.71	0.73
ESR1 ^c	5010	0.69	0.71	0.81	0.83	0.77	0.80
ESR1 ^p	1111	0.70	0.71	0.69	0.70	0.74	0.75
FOXA1	11442	0.65	0.65	0.74	0.76	0.76	0.78
RELA	4776	0.59	0.60	0.64	0.64	0.62	0.63
REST	1437	0.91	0.92	0.78	0.79	0.93	0.93
TP53	481	0.86	0.86	0.58	0.57	0.87	0.86
TP63 ⁺	3088	0.58	0.61	0.71	0.72	0.67	0.70
TP63 ⁻	4805	0.60	0.64	0.74	0.75	0.70	0.73
STAT1 ^s	30021	0.61	0.61	0.77	0.77	0.71	0.75
STAT1 ^u	6948	0.53	0.53	0.81	0.80	0.67	0.71
USF1	1029	0.82	0.85	0.84	0.85	0.88	0.90
USF2	561	0.75	0.78	0.82	0.84	0.83	0.85
Average		0.69	0.70	0.74	0.75	0.76	0.78

Supplementary Table 8: AUC for predicting 1kb bins of the genome containing the center of a reported binding site excluding regions near an annotated RefSeq TSS. This analysis was done genome-wide excluding gapped portions of the genome and bins containing an annotated TSS or within (a) +/-2 bins, (b) +/- 5 bins and (c) +/-10 bins of a bin containing an annotated TSS.

TF	Cell Treatment	Array ID	# of Targets	Max PWM Score	Average PWM Score	Average GBP	Max(GBP x PWM Score)	Average (GBP x PWM Score)
E2F2	Late G1	4501	311	0.64	0.64	0.63	0.76	0.75
E2F2	None	4375	505	0.62	0.62	0.66	0.74	0.74
E2F2	Late G1	4503	407	0.62	0.62	0.64	0.75	0.74
E2F2	S-phase	4666	487	0.62	0.62	0.64	0.74	0.73
E2F2	None	4376	428	0.62	0.62	0.65	0.74	0.73
E2F4	None	4685	476	0.62	0.61	0.63	0.75	0.73
E2F4	None	4683	430	0.63	0.60	0.63	0.74	0.73
E2F4	None	4684	501	0.62	0.61	0.63	0.75	0.73
E2F4	None	4668	482	0.62	0.60	0.63	0.74	0.72
E2F2	S-phase	4667	323	0.62	0.61	0.63	0.72	0.71
E2F4	None	4669	583	0.61	0.59	0.62	0.73	0.71
E2F4	None	4670	459	0.61	0.59	0.62	0.72	0.71
E2F4	None	4671	310	0.61	0.59	0.61	0.72	0.71

Supplementary Table 9: Table of AUC Values for Predicting bound Promoters in new E2F ChIP-chip data.

TF	Cell Treatment	Array ID	# of Targets	Max PWM Score	Average PWM Score	Average GBP	Max(GBP x PWM Score)	Average (GBP x PWM Score)
E2F1b	GM06990	82997	224	0.70	0.74	0.74	0.81	0.83
E2F1c	GM06990	82031	128	0.72	0.75	0.75	0.80	0.81
E2F6c	GM06990	85703	152	0.68	0.68	0.74	0.82	0.82
E2F6c	MCF7	74971	47	0.69	0.68	0.69	0.83	0.84
E2F6b	MCF10A	70614	16	0.74	0.76	0.58	0.83	0.80
E2F1b	MCF7	77440	472	0.65	0.67	0.69	0.78	0.78
E2F4c	GM06990	82770	198	0.66	0.66	0.67	0.80	0.80
E2F1b	Ntera2	88783	205	0.66	0.70	0.69	0.76	0.76
E2F4a	Ntera2	81752	346	0.68	0.69	0.65	0.78	0.77
E2F4c	HelaS3	104455	1218	0.64	0.65	0.71	0.77	0.78
E2F4b	GM06990	82998	656	0.65	0.65	0.68	0.79	0.78
E2F1a	MCF7	71262	736	0.65	0.65	0.69	0.78	0.78
E2F4b	Ntera2	88877	809	0.66	0.66	0.68	0.77	0.77
E2F6b	GM06990	85359	200	0.65	0.64	0.69	0.77	0.76
E2F4b	MCF7	77447	264	0.64	0.62	0.65	0.78	0.77
E2F4b	HelaS3	90989	2402	0.62	0.62	0.70	0.76	0.76
E2F1c	HelaS3	92000	478	0.62	0.66	0.69	0.74	0.75
E2F1b	HelaS3	91860	2120	0.61	0.61	0.71	0.74	0.75
E2F6a	MCF7	74981	249	0.64	0.63	0.64	0.75	0.74
E2F1b	MCF10A	70246	399	0.61	0.64	0.70	0.72	0.73
E2F4b	MCF10A	70613	3330	0.59	0.60	0.70	0.75	0.75
E2F4a	MCF7	63104	1859	0.60	0.60	0.68	0.74	0.74
E2F4a	MCF10A	70610	3965	0.59	0.59	0.69	0.74	0.74
E2F6c	HelaS3	91178	1244	0.59	0.59	0.68	0.72	0.73
E2F6b	HelaS3	91995	3633	0.58	0.58	0.70	0.72	0.72
E2F1a	Ntera2	88875	416	0.62	0.66	0.61	0.68	0.68
E2F6a	Ntera2	88867	5567	0.58	0.60	0.66	0.69	0.69
E2F6b	Ntera2	85202	4399	0.58	0.60	0.65	0.68	0.69
E2F1a	MCF10A	70609	894	0.60	0.61	0.61	0.67	0.67
E2F6a	MCF10A	72084	28	0.60	0.57	0.56	0.71	0.68

Supplementary Table 10: Table of AUC Values for Predicting bound Promoters in E2F ChIP-chip data from (Xu et al, 2007).

Experiment Set	Corrected p-value	Fold-enrichment
E2F2 (this study)	4×10^{-16}	3.2
E2F4 (this study)	1×10^{-19}	3.3
E2F1 (Xu et al., 2007)	5×10^{-37}	3.5
E2F4 (Xu et al., 2007)	3×10^{-59}	2.6
E2F6 (Xu et al., 2007)	6×10^{-22}	1.7

Supplementary Table 11: Cell Cycle GO Enrichment for Genes that were a bound target in at least two experiments within the experiment set.

>=Num Experiments	# Targets	Max PWM Score	Average PWM Score	Average GBP	Max(GBP x PWM Score)	Average (GBPx PWM Score)
1	1553	0.59	0.59	0.65	0.71	0.71
2	975	0.61	0.60	0.64	0.74	0.72
3	737	0.61	0.60	0.64	0.74	0.73
4	582	0.63	0.61	0.63	0.75	0.73
5	475	0.64	0.62	0.63	0.76	0.75
6	391	0.65	0.63	0.63	0.77	0.75
7	299	0.65	0.64	0.62	0.77	0.75
8	221	0.67	0.65	0.62	0.78	0.76
9	167	0.68	0.66	0.62	0.79	0.77
10	132	0.68	0.67	0.61	0.79	0.77
11	93	0.67	0.64	0.58	0.78	0.75
12	58	0.66	0.65	0.58	0.76	0.74
13	19	0.77	0.70	0.62	0.79	0.79

Supplementary Table 12: AUC Values for Predicting Targets Bound in Multiple Experiments for the new E2F ChIP-chip data

Experiments	#predictions	Max PWM Score	Average PWM Score	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBP x PWM Score)	PREM Database	Random Expected
MYC	1575	235	298	253	470	405	509	286	197.9
ESR1 ^c	1653	167	188	69	171	203	223	131	89.4
ESR1 ^p	1653	30	39	14	17	50	40	18	16.3
FOXA1	3335	393	364	383	499	556	633	590	355.5
RELA	822	89	109	83	171	149	162	129	64.2
REST	385	257	255	12	19	245	246	116	13
TP53	195	25	25	2	2	27	26	7	0.9
TP63 ⁺	998	53	26	36	59	65	50	63	42.8
TP63 ⁻	998	79	52	70	128	118	102	108	69.8
STAT1 ^s	1575	1013	946	1294	1245	1297	1327	945	852.7
STAT1 ^u	1575	450	404	730	707	669	729	500	400
USF1	1845	388	285	268	276	661	508	443	147.3
USF2	1845	233	147	214	185	413	318	303	95.4

Supplementary Table 13: Comparison with the predictions in the PREM database

(Blanchette et al, 2006) on Predicting Binding within +/-10kb of a RefSeq TSS in Genome-wide binding data. Predictions were based on all the predictions in the PREM database and the equivalent number of top predictions based on the other methods. The table reports for each method the number of predicted intervals which were bound. The UCSC genome browser liftover tool was used to convert the PREM coordinates from hg17 to hg18. Predictions were based on the same TRANSFAC motif as listed in Supplementary Table 1 except M00256 was used for REST since predictions for M01028 were not available.

TF Factor	ID	Max PWM Score	Average PWM Score	Average GBP	Max (GBP x PWM Score)	Average (GBP x PWM Score)	PREM database	Random Expected
E2F2	4501	92	97	47	142	132	98	49.1
E2F2	4375	139	137	98	200	189	155	79.7
E2F2	4503	121	118	77	176	164	130	64.2
E2F2	4666	131	132	83	201	188	144	76.8
E2F2	4376	112	116	87	174	166	136	67.5
E2F4	4685	134	130	80	200	177	131	75.1
E2F4	4683	118	117	69	179	166	124	67.8
E2F4	4684	138	135	74	214	195	141	79
E2F4	4668	128	128	74	196	178	136	76.1
E2F2	4667	93	92	50	131	117	94	51
E2F4	4669	147	140	86	225	200	148	92
E2F4	4670	121	107	64	176	155	125	72.4
E2F4	4671	78	74	40	124	107	78	48.9

Supplementary Table 14: Comparison with the predictions in the PREM database (Blanchette, 2006) on the E2F data. Predictions were based on the 2845 predictions in the PREM database and the same number of top predictions based on the other methods. The table reports the number of predicted spots which were bound. The UCSC genome browser liftover tool was used to convert the PREM coordinates from hg17 to hg18. The common TRANSFAC E2F motif M00050 was used for this analysis.

Data Set	# of Bound Regions	Max Log Sum Histone Feature	Average Log Sum Histone Feature	Max GBP	Average GBP	Max (PWM x Log Sum Histone Feature)	Average (PWM x Log Sum Histone Feature)	Max (GBP x PWM Score)	Average (GBP x PWM Score)
MYC	2671	0.66	0.67	0.64	0.68	0.64	0.65	0.66	0.68
ESR1 ^c	1150	0.56	0.61	0.54	0.63	0.63	0.66	0.64	0.68
ESR1 ^p	209	0.52	0.56	0.50	0.56	0.67	0.66	0.65	0.65
FOXA1	2266	0.56	0.58	0.56	0.59	0.56	0.58	0.61	0.64
RELA	1660	0.63	0.63	0.60	0.64	0.61	0.62	0.62	0.63
REST	720	0.45	0.55	0.46	0.55	0.83	0.83	0.83	0.83
TP53	102	0.56	0.55	0.56	0.57	0.82	0.82	0.80	0.80
TP63 ⁺	911	0.56	0.57	0.56	0.58	0.54	0.53	0.57	0.57
TP63 ⁻	1487	0.55	0.57	0.56	0.59	0.55	0.56	0.58	0.59
STAT1 ^s	11504	0.76	0.70	0.77	0.73	0.61	0.63	0.65	0.70
STAT1 ^u	5399	0.72	0.67	0.73	0.69	0.57	0.58	0.61	0.65
USF1	1697	0.69	0.65	0.70	0.67	0.72	0.69	0.81	0.77
USF2	1099	0.70	0.65	0.71	0.68	0.69	0.66	0.78	0.75
Average		0.61	0.61	0.61	0.63	0.65	0.65	0.68	0.69

Supplementary Table 15: Comparison with Log Sum Histone Feature. We compare with just using the log sum histone feature and multiplying the PWM by this feature instead of the GBP. Combining GBP with motif information gives the highest average AUC values.

Experiments	# 1kb bin predictions	Max PWM Score	Average PWM Score	Max GBP	Average GBP	Max(GBP x PWM Score)	Average (GBP x PWM Score)	Random
MYC	4287	21	54	230	256	303	292	6.4
ESR1 ^c	5779	213	216	38	49	329	317	11.7
ESR1 ^p	1231	42	43	4	1	41	37	0.5
FOXA1	12896	177	112	213	257	663	769	58.2
RELA	5852	58	95	156	171	256	274	12
REST	1930	698	699	12	8	710	708	1.3
TP53	542	96	99	0	1	84	84	0.1
TP63 ⁺	3664	7	23	15	21	49	38	4.7
TP63 ⁻	5783	27	44	57	65	104	85	11.7
STAT1 ^s	41428	2033	2627	9335	9292	5133	6182	600.5
STAT1 ^u	10971	125	146	2146	2026	640	870	42.1
USF1	2506	112	119	158	134	339	355	2.2
USF2	1342	40	31	69	71	165	172	0.6

Supplementary Table 16: Number of correctly predicted bins when the number of predictions is set to the number of bound bins. The table reports the number of 1kb bins that would be correctly predicted when making the same number of predicted as are positive in the evaluation set.