Supplementary Figure Legends

Figure S1, related to Figures 1 and 2. Summary of analysis using control sequences from flanking non-ChIP regions. (A) Comparison of GC-content under the two strategies; current strategy: using ChIP peaks of other DBPs, shown along X-axis; alternative strategy: using non-ChIP regions flanking the ChIP peaks of the same DBP, shown along Y-axis. All cases show high concordance, Pearson correlation coefficients are: 0.98 for positive training set, 0.97 for control training set, 0.98 for positive validation set, and 0.96 for control validation set. (B) Main results of our manuscript under the re-analysis using control sequences from flanking non-ChIP regions. Heatmap of negative-$\log_{10}$ transformed Bonferroni corrected p-values for the four types of shape-motifs of each TF. White cells indicate no significant motif. 'K' or 'G' after a TF's name denotes K562 or Gm12878 cells, respectively.

Figure S2, related to Figures 1, 2 and 4. (A)—(D) Main results of our manuscript under the re-analysis using control sequences from flanking non-ChIP regions. (A) Numbers of TFs with each of the four types of shape-motifs. (B) Length distribution of shape-motifs. (C) Boxplots of relative distances of sequence-, shape-, and overlapping-sites from ChIP-seq peak-summits. (D) The strategy of taking flanking non-ChIP regions as controls can still produce GC-content mismatches across positive and control sequences, and the extents are similar to the strategy of taking control sequences from ChIP-peaks of other DBPs. Scatterplots show GC-content of positive and control sequences in the validation sets across all DBP datasets. (E) Left panel shows boxplots of shape-feature values in the initial windows selected randomly by ShapeMF and in the windows that match the final shape-motif (blue and red boxplots, respectively). Right panel shows the final shape-motif and the sequence motif computed from the sequences underlying this shape-motif. The shown example is a HelT motif of MAX. (F) Barplots showing the faction (%) of shape-only peaks and sequence-based (sequence-only plus co-occurrence) peaks in different types of regulatory regions in the Gm12878 cell-line. Shape-only peaks are more enriched in putative enhancer regions than are sequence-based peaks; significant enrichments are marked with stars.

Figure S3, related to Figures 2 and 4. (A) Scatterplot showing $F_{1/3}$-scores of shape-motifs across training and validation datasets. (B) Distribution of recall of shape-motifs on validation datasets. (C) Scatterplot showing a monotonically increasing relationship between the fraction of overlapping sites of a DBP and the Information Content per Position (ICP) derived from sequence-logos underlying its shape-motifs. Each circle denotes a DBP and the mean ICP is the mean of ICP its different shape-motifs. (D) Histograms showing distributions of gkmSVM scores of sequence-only, overlapping, and shape-only sites (pooled across all DBPs). A site's score is the maximum gkmSVM score of any 10-mer within that site. (E) Scatterplots showing the % of shape-only sites and

sequence-sites of each DBP that have a minimum overlap (% of the site's length, not stipulated to be reciprocal) with sequence-sites of any other DBP.

Figure S4, related to Figures 2 and 4. (A) Histograms of Information Content per Position (ICP), GC-content, and length of shape- and sequence-motifs. The ICP and GC-content for a shape-motif is derived from the sequences underlying the occurrences of the shape-motif.

Figure S5, related to Figure 3. (A) Sequence-motifs underlying shape-motifs in HT-SELEX and ChIP-Seq are generally similar. For each panel, the shape-motifs are shown in the left and sequence-motifs underlying those shape-motifs are shown in the right (top motif is from ChIP-derived shape-motif, bottom motif is from SELEX-derived shape-motif). For 16/36 cases, the two sequence-logos have a very similar core (shown with double **). CTCF Roll motif (3rd column, 2nd motif from the bottom) and NRF1 Roll motif (4th column, 5th motif) are two such examples of extreme difference in length yet high similarity in some core. For 6/36 cases (shown with single *), such agreement is modest. (B) Statistical significance vs. IoU (Intersection over Union) for each pair of shape-motifs derived from HT-SELEX and ChIP-Seq.

Figure S6, related to Figure 4. Fraction of sequence-only, co-occurrence, and shape-only peaks within different deciles of ChIP-Seq intensity (10 = peaks with top 10% intensity, 1 = peaks with lowest 10% intensity). Shown is the K562 cell-line dataset.

Figure S7, related to Figure 4. Fraction of sequence-only, co-occurrence, and shape-only peaks within different deciles of ChIP-Seq intensity (10 = peaks with top 10% intensity, 1 = peaks with lowest 10% intensity). Shown is the Gm12878 cell-line dataset.

Figure S8, related to Figures 2 and 4. Shape-motifs and shape-specific binding occur similarly across sequence-specific TFs and non-specific DBPs. (A-C) Density histograms showing density of TFs and non-specific DBPs binned according to the fraction of their genome wide peaks: sequence-only (A), shape-only (B), and co-occurrence peaks (C). Of note, the densities are unimodal and the modes of the two groups of proteins are very close. The generality of our results would not hold in case the densities for the two groups were skewed in opposite directions. (D) Fraction (%) of sequence-specific TFs and other non-specific DBPs with each type of shape-motif. (E) Distributions of information content per position (ICP) of shape-motifs from the two groups of proteins. Each box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data.

Figure S9, related to Figure 6. (A) Histogram of the number of overlapping peaks between co-binding DBP pairs. Only the pairs with at least 100 overlapping peaks are included. (B) Shape-zinger DBPs (nodes) and the fraction (color-coded) of ChIP-Seq peaks of other DBPs with their shape-sites. For simplicity, edges connecting shape-zinger pairs

are omitted, but most zingers have large numbers of shape-sites in each other's peaks. (C) Histogram of inter-site distances at which DBP pairs show significant bias for co-binding. (D) Shape-motifs of TBX5 and NKX2-5 for MGW and Roll, and the corresponding sequence-motifs created from the underlying sequences.

Figure S10, related to Figure 6. Heatmaps showing the extent of shape-only binding for each DBP F_1 (row-names) in the peaks where it co-binds with another DBP F_2 (column-names). The extent is quantified as the log-transformed ratio of the number of shape-only peaks to sequence-based peaks of F_1 in the co-binding peaks. The colored bars (row and column aligned) show the same statistic for the corresponding DBP computed from its entire set of peaks.

Figure S11, related to Figure 7. Shape-motifs of bHLH proteins showing that DBPs within the same family have different shape-motifs (H, M, P, and R denote motifs for HelT, MGW, ProT, and Roll, respectively). A feature is not shown for a DBP if the DBP does not have a shape-motif for that feature. All DBPs in the bHLH class for which ChIP-Seq assays were performed by ENCODE in the K562 cell-line and we could find at least one shape-motif are shown.

Figure S12, related to Figure 4. MITOMI-based ranks of sequence-only, overlapping, and shape-only sites.

Supplementary Tables

Table S1, related to Figure 4. Average number of sequence, shape, and overlapping sites per peak.

Table S2, related to Figures 2 and 4. Mean information content per position (ICP) of shape-motifs.

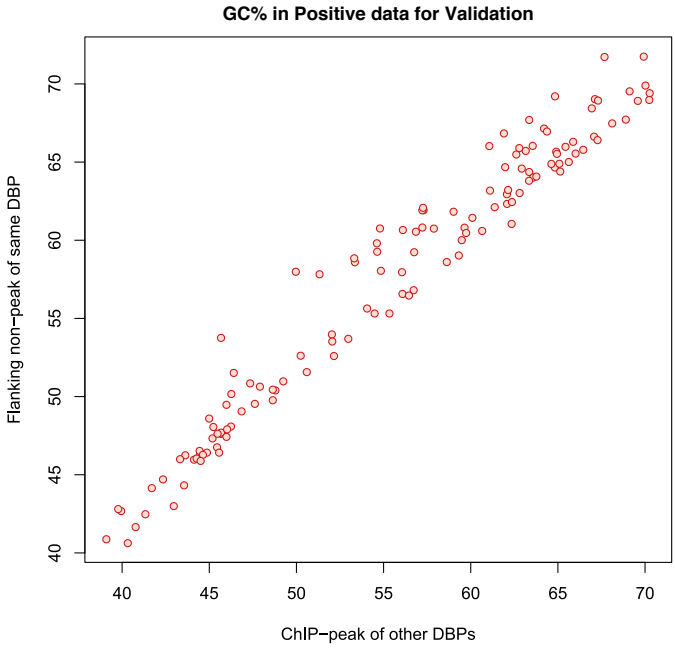
Table S3, related to Figure 4. Fractions of top peaks that are sequence-only, shape-only, and co-occurrence. Every cell shows two values in the format $f_s/f_{sh}/f_{co}$, where f_s, f_{sh}, and f_{co} denote the fraction of sequence-only, shape-only, and co-occurrence peaks, respectively.

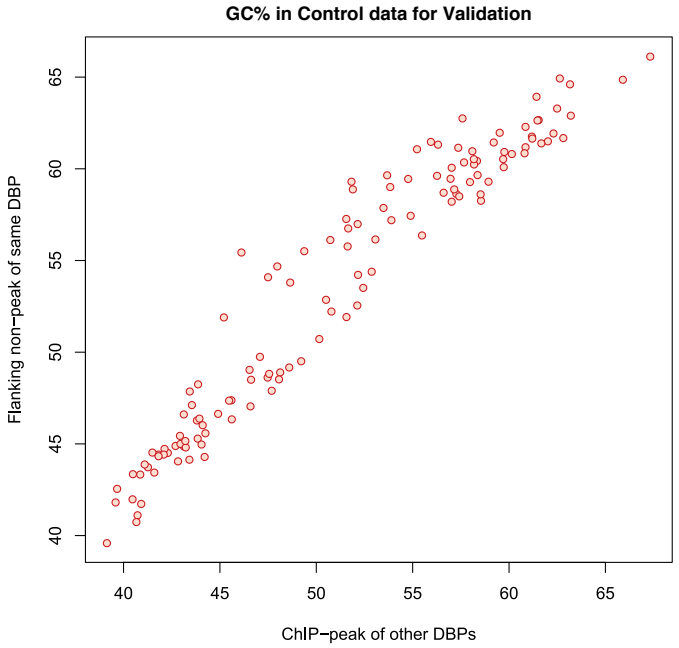
Table S4, related to Figure 4. Number of datasets where the highest fraction of peaks is sequence-only, co-occurrence, or shape-only across different affinity bins

(A)

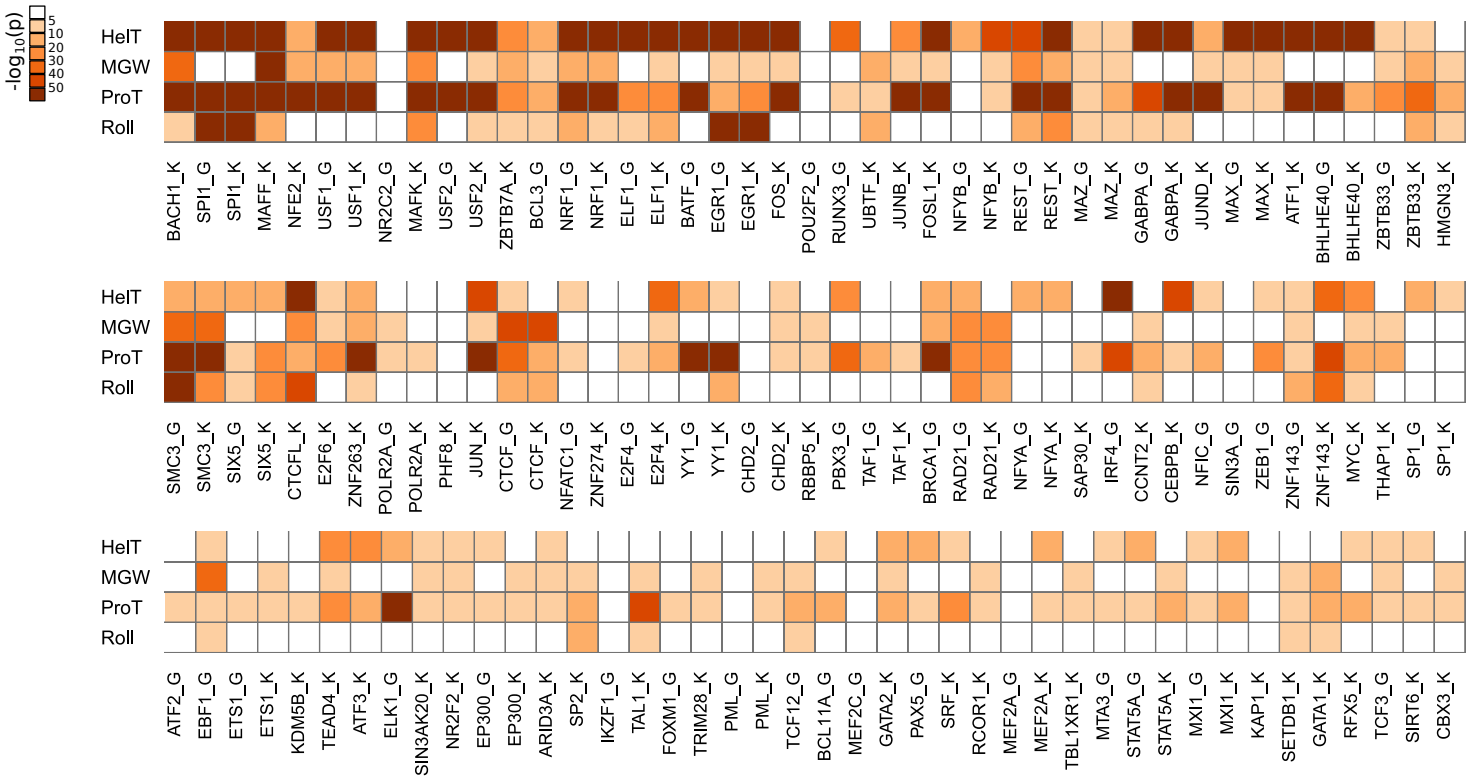

GC% in Positive data for Training
Flanking non−peak of same DBP
ChIP−peak of other DBPs

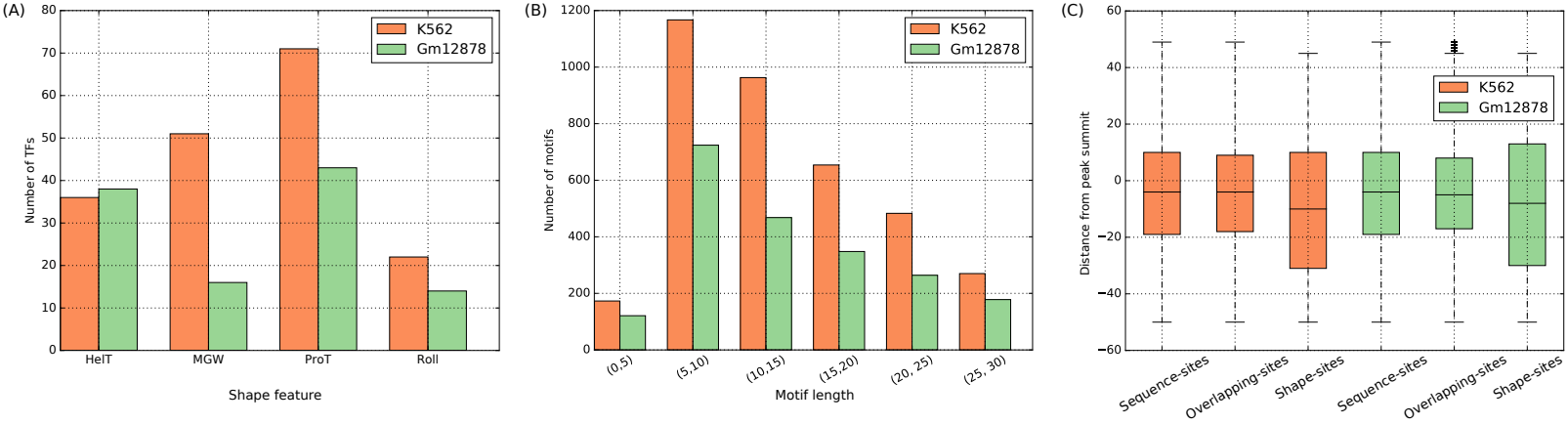

GC% in Control data for Training
Flanking non−peak of same DBP
ChIP−peak of other DBPs

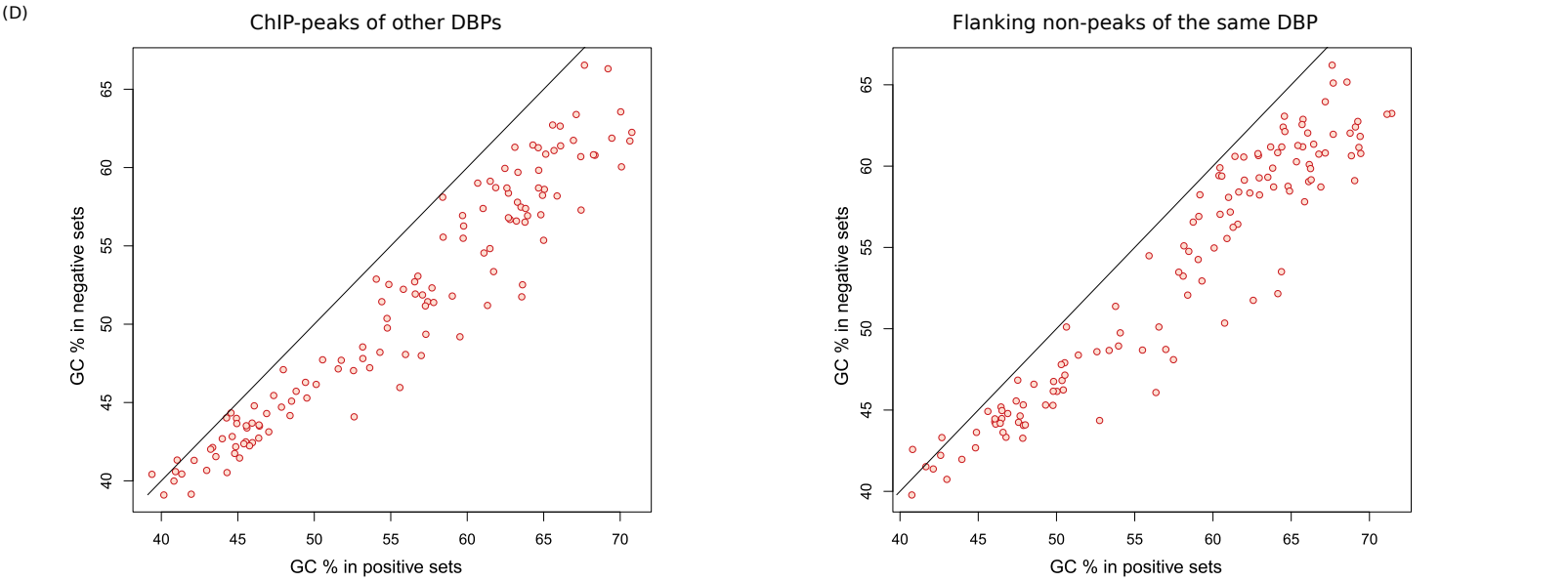

GC% in Positive data for Validation
Flanking non−peak of same DBP
ChIP−peak of other DBPs

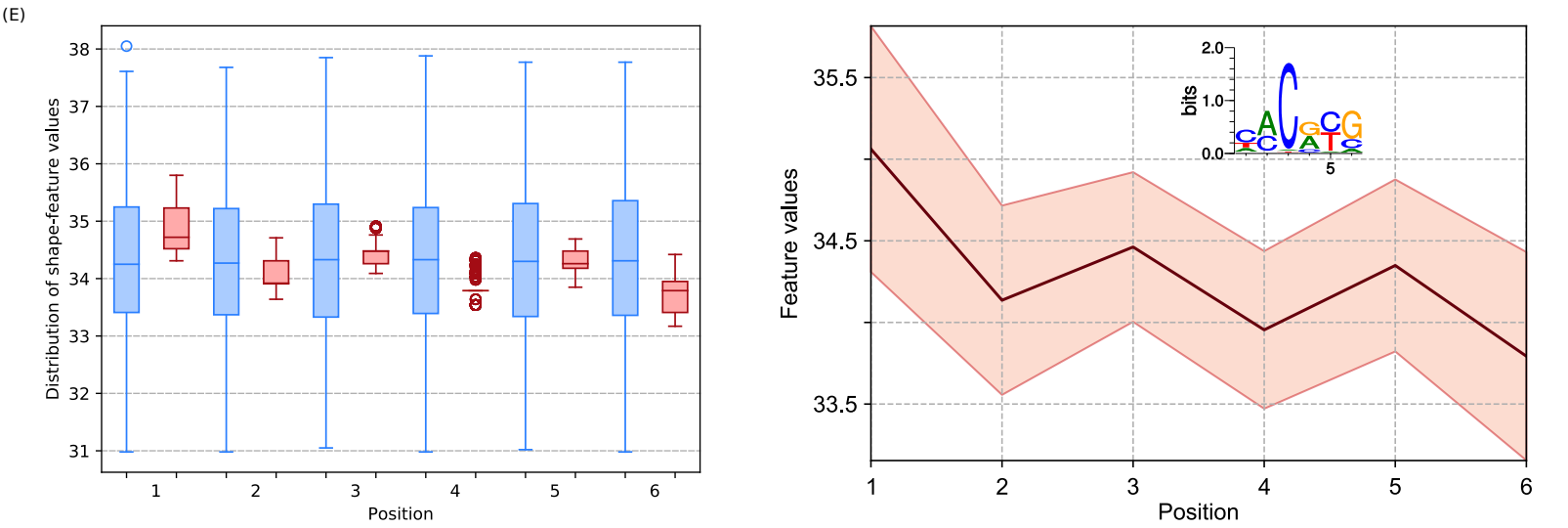

GC% in Control data for Validation
Flanking non−peak of same DBP
ChIP−peak of other DBPs

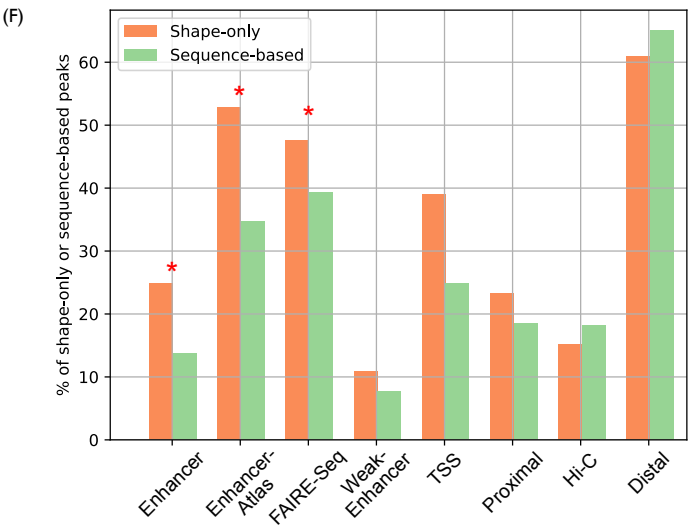
(B)

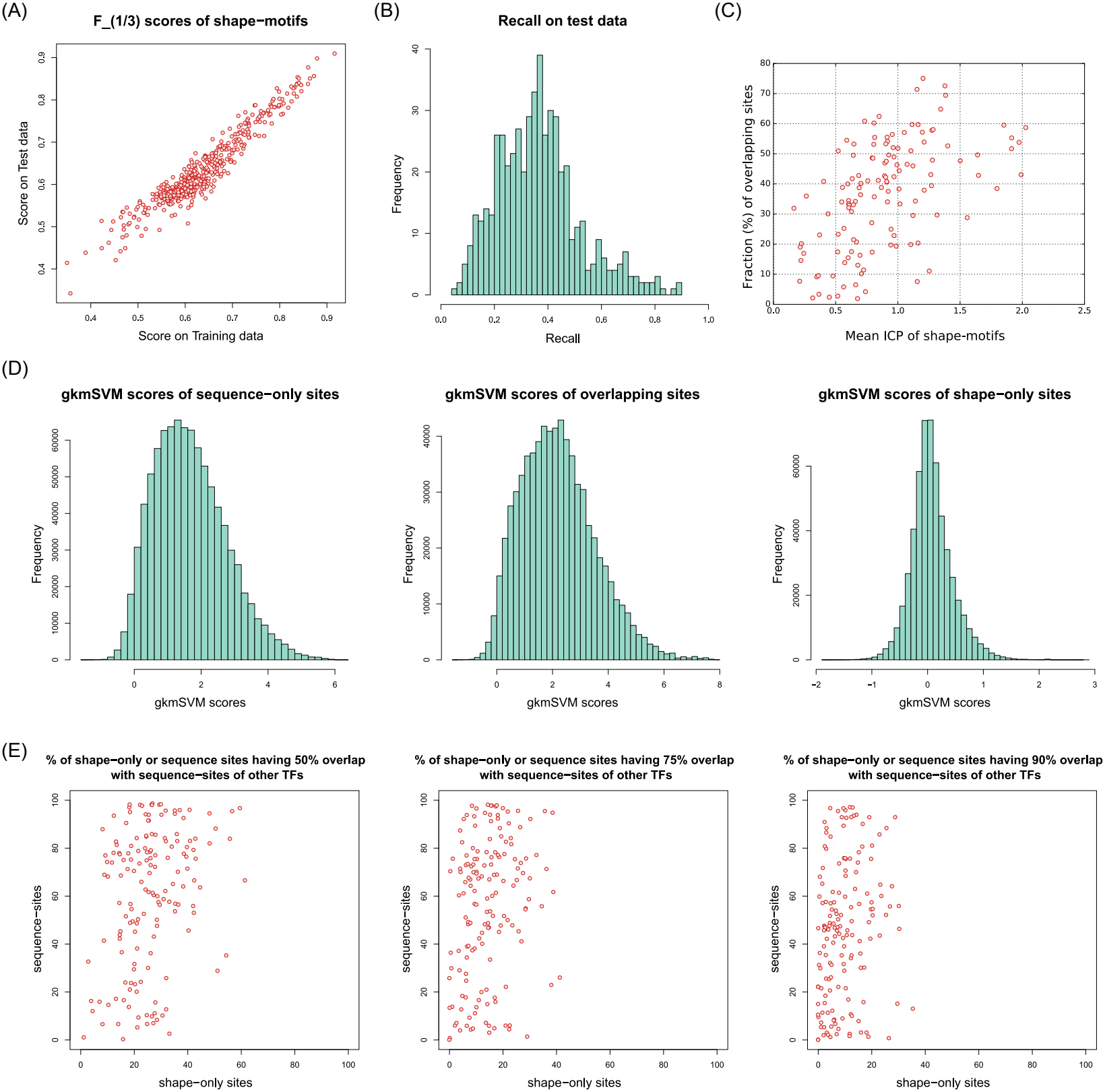

-log10(p)
5
10
20
30
40
50
HelT
MGW
ProT
Roll
BACH1_K
SPI1_G
SPI1_K
MAFF_K
NFE2_K
USF1_G
USF1_K
NR2C2_G
MAFK_K
USF2_G
USF2_K
ZBTB7A_K
BCL3_G
NRF1_G
NRF1_K
ELF1_G
ELF1_K
BATF_G
EGR1_G
EGR1_K
FOS_K
POU2F2_G
RUNX3_G
UBTF_K
JUNB_K
FOSL1_K
NFYB_G
NFYB_K
REST_G
REST_K
MAZ_G
MAZ_K
GABPA_G
GABPA_K
JUND_K
MAX_G
MAX_K
ATF1_K
BHLHE40_G
BHLHE40_K
ZBTB33_G
ZBTB33_K
HMGN3_K
SMC3_G
SMC3_K
SIX5_G
SIX5_K
CTCFL_K
E2F6_K
ZNF263_K
POLR2A_G
POLR2A_K
PHF8_K
JUN_K
CTCF_G
CTCF_K
NFATC1_G
ZNF274_K
E2F4_G
E2F4_K
YY1_G
YY1_K
CHD2_G
CHD2_K
RBBP5_K
PBX3_G
TAF1_G
TAF1_K
BRCA1_G
RAD21_G
RAD21_K
NFYA_G
NFYA_K
SAP30_K
IRF4_G
CCNT2_K
CEBPB_K
NFIC_G
SIN3A_G
ZEB1_G
ZNF143_G
ZNF143_K
MYC_K
THAP1_K
SP1_G
SP1_K
ATF2_G
EBF1_G
ETS1_G
ETS1_K
KDM5B_K
TEAD4_K
ATF3_K
ELK1_G
SIN3AK20_K
NR2F2_K
EP300_G
EP300_K
ARID3A_K
SP2_K
IKZF1_G
TAL1_K
FOXM1_G
TRIM28_K
PML_G
PML_K
TCF12_G
BCL11A_G
MEF2C_G
GATA2_K
PAX5_G
SRF_K
RCOR1_K
MEF2A_G
MEF2A_K
TBL1XR1_K
MTA3_G
STAT5A_G
STAT5A_K
MXI1_G
MXI1_K
KAP1_K
SETDB1_K
GATA1_K
RFX5_K
TCF3_G
SIRT6_K
CBX3_K


(A)
Number of TFs
K562
Gm12878
HelT
MGW
ProT
Roll
Shape feature
(B)
Number of motifs
K562
Gm12878
(0,5)
(5,10)
(10,15)
(15,20)
(20, 25)
(25, 30)
Motif length
(C)
Distance from peak summit
K562
Gm12878
Sequence-sites
Overlapping-sites
Shape-sites
Sequence-sites
Overlapping-sites
Shape-sites

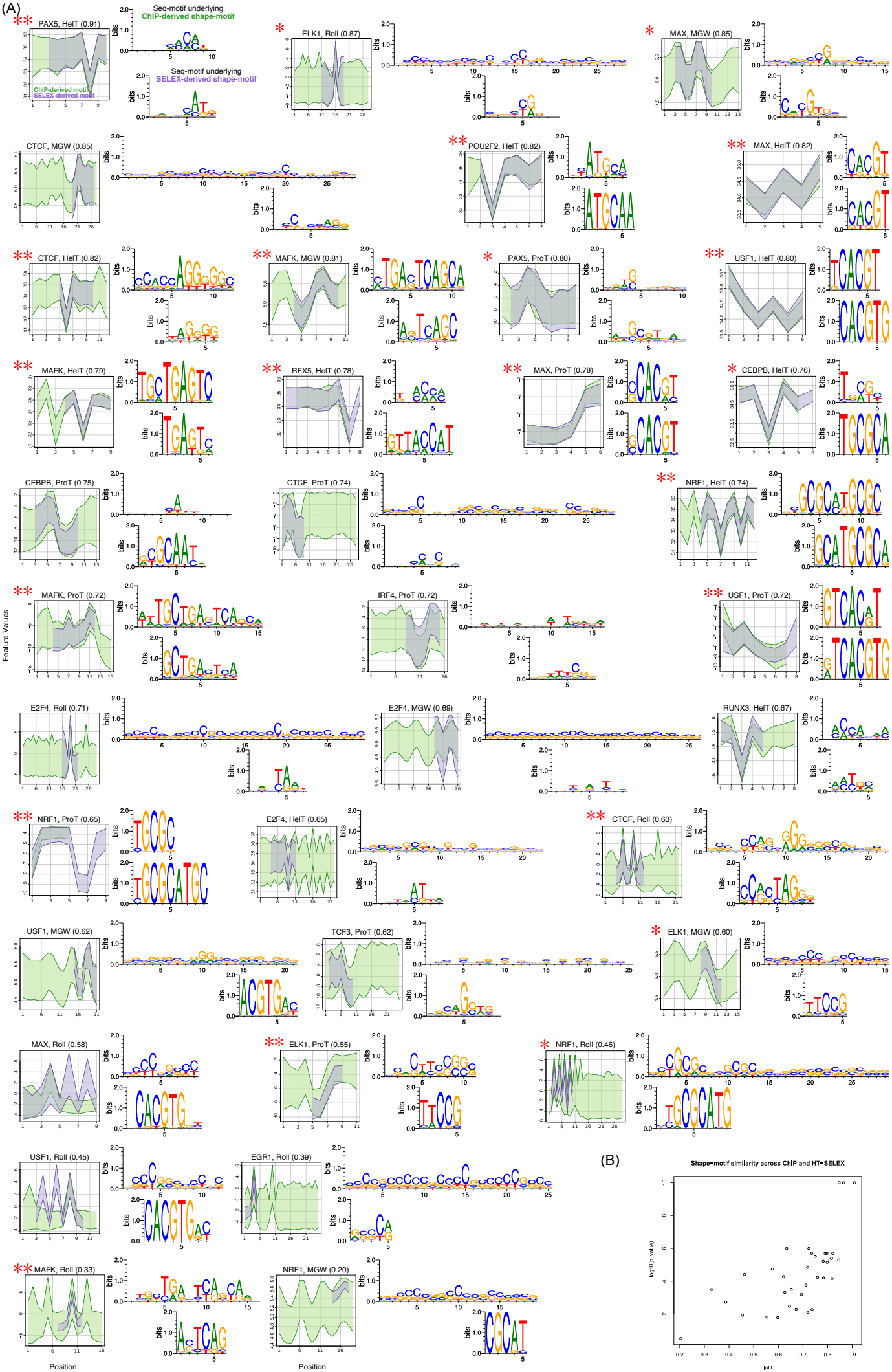

(D)
ChIP-peaks of other DBPs
GC % in negative sets
GC % in positive sets
Flanking non-peaks of the same DBP
GC % in negative sets
GC % in positive sets


(E)
Distribution of shape-feature values
Position
Feature values
bits
Position


(F)
Shape-only
Sequence-based
% of shape-only or sequence-based peaks
Enhancer
Enhancer-Atlas
FAIRE-Seq
Weak-Enhancer
TSS
Proximal
Hi-C
Distal


(A)
F_(1/3) scores of shape−motifs
Score on Test data
Score on Training data
(B)
Recall on test data
Frequency
Recall
(C)
Fraction (%) of overlapping sites
Mean ICP of shape-motifs
(D)
gkmSVM scores of sequence−only sites
gkmSVM scores of overlapping sites
gkmSVM scores of shape−only sites
Frequency
gkmSVM scores
(E)
% of shape−only or sequence sites having 50% overlap with sequence−sites of other TFs
% of shape−only or sequence sites having 75% overlap with sequence−sites of other TFs
% of shape−only or sequence sites having 90% overlap with sequence−sites of other TFs
sequence−sites
shape−only sites


HelT-motifs
MGW-motifs
ProT-motifs
Roll-motifs
TRANSFAC motifs
Count of motifs
Information content per position
GC-content
Motif length

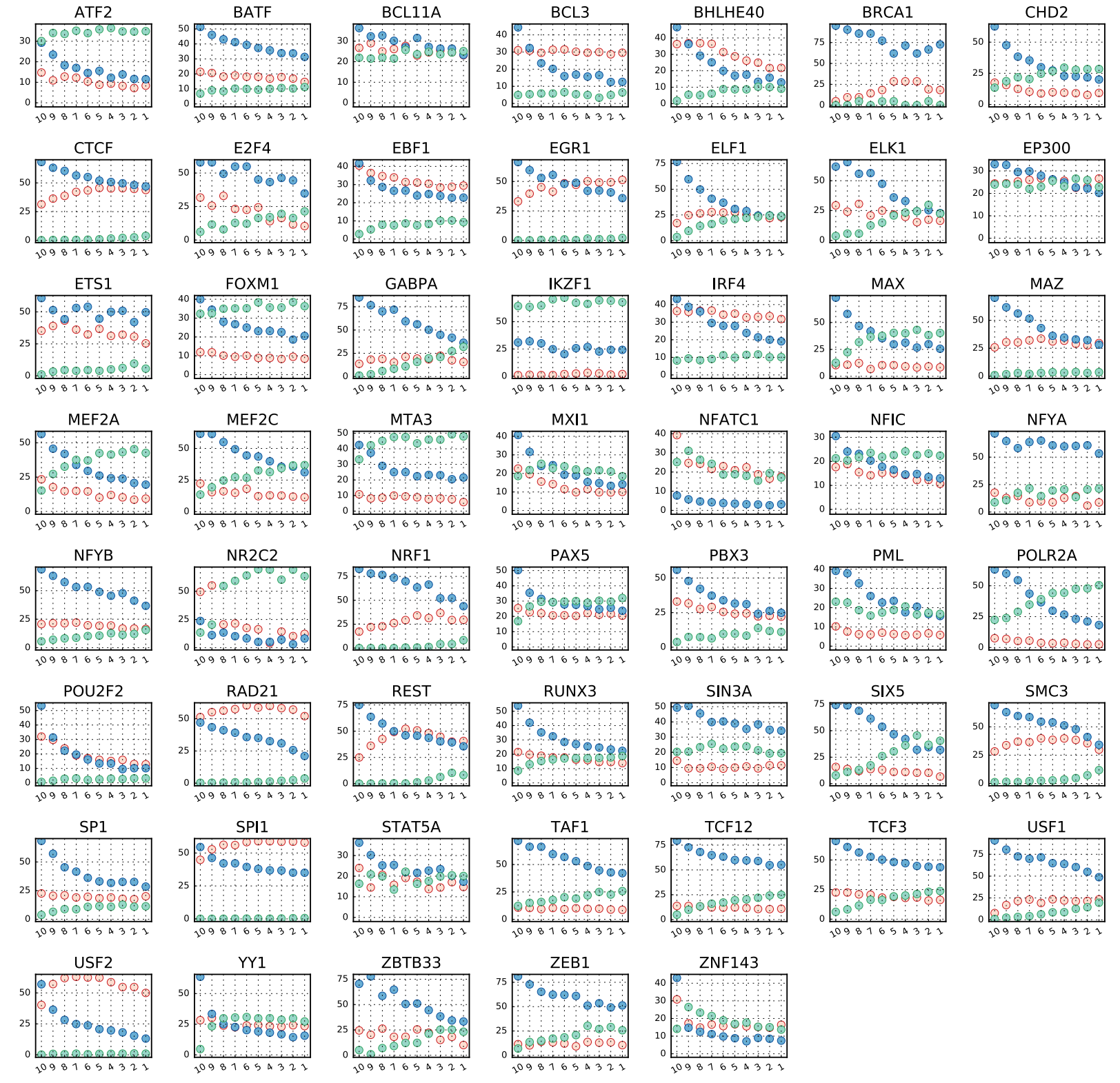

(A)
Seq-motif underlying
ChIP-derived shape-motif
Seq-motif underlying
SELEX-derived shape-motif
ChIP-derived motif
SELEX-derived motif
PAX5, HelT (0.91)
ELK1, Roll (0.87)
MAX, MGW (0.85)
CTCF, MGW (0.85)
POU2F2, HelT (0.82)
MAX, HelT (0.82)
CTCF, HelT (0.82)
MAFK, MGW (0.81)
PAX5, ProT (0.80)
USF1, HelT (0.80)
MAFK, HelT (0.79)
RFX5, HelT (0.78)
MAX, ProT (0.78)
CEBPB, HelT (0.76)
CEBPB, ProT (0.75)
CTCF, ProT (0.74)
NRF1, HelT (0.74)
MAFK, ProT (0.72)
IRF4, ProT (0.72)
USF1, ProT (0.72)
E2F4, Roll (0.71)
E2F4, MGW (0.69)
RUNX3, HelT (0.67)
NRF1, ProT (0.65)
E2F4, HelT (0.65)
CTCF, Roll (0.63)
USF1, MGW (0.62)
TCF3, ProT (0.62)
ELK1, MGW (0.60)
MAX, Roll (0.58)
ELK1, ProT (0.55)
NRF1, Roll (0.46)
USF1, Roll (0.45)
EGR1, Roll (0.39)
MAFK, Roll (0.33)
NRF1, MGW (0.20)
Feature Values
Position
Position
(B)
Shape-motif similarity across ChIP and HT-SELEX
-log10(p-value)
IoU

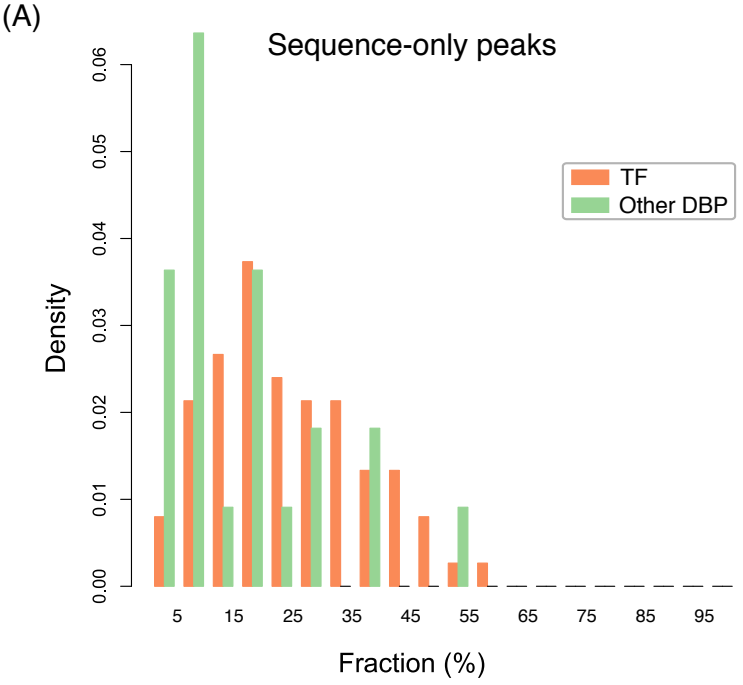
Shape-only
Sequence-only
Co-occurrence

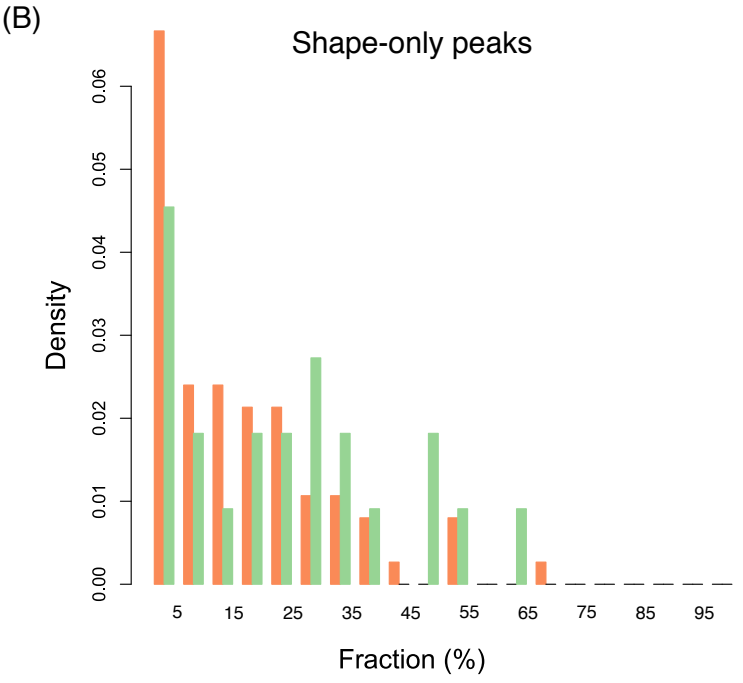
Fraction of shape-only, sequence-only, and co-occurrence peaks in different deciles of ChIP intensity
Cell-type: K562
X-axis: decile of ChIP-peaks computed from ChIP intensity (10 = decile of top 10%, 1 = decile of lowest 10%)
Y-axis: fraction (%)

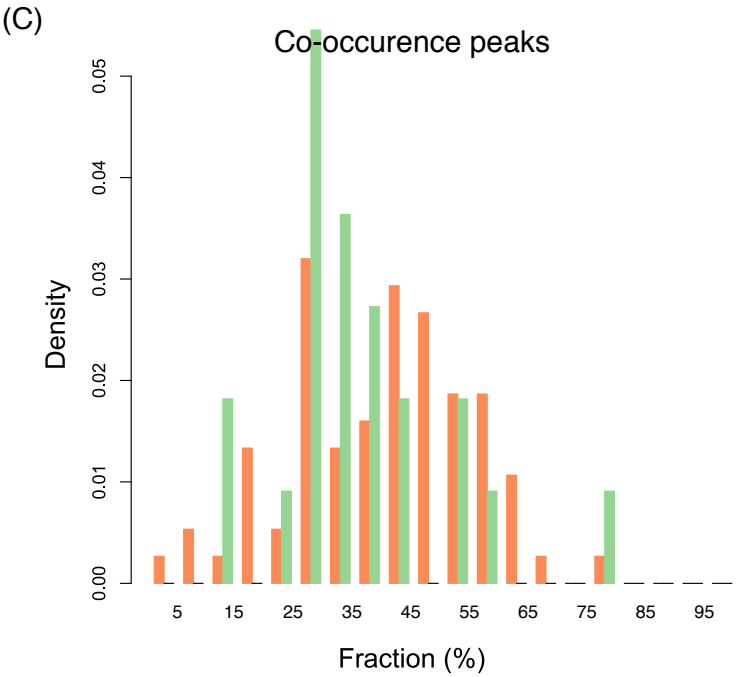
Shape-only
Sequence-only
Co-occurrence

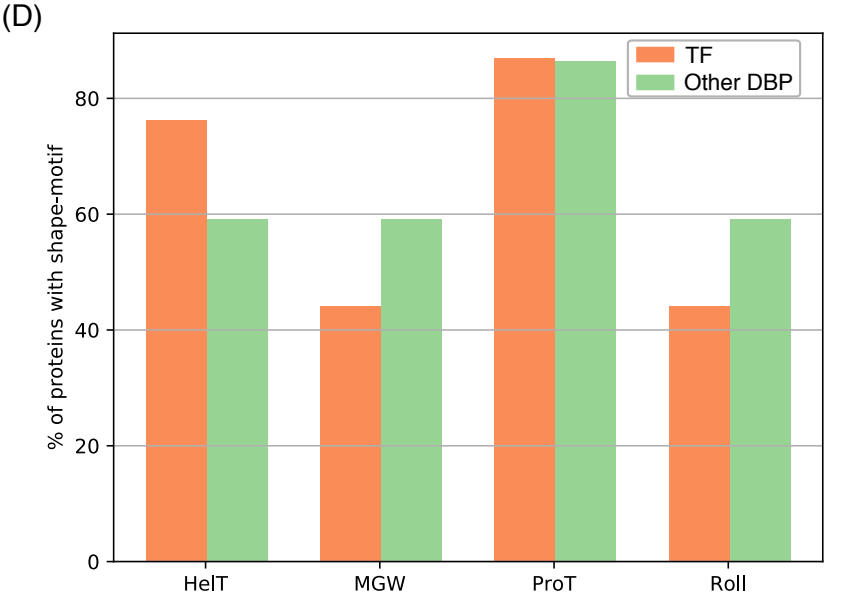
Fraction of shape-only, sequence-only, and co-occurrence peaks in different deciles of ChIP intensity
Cell-type: Gm12878
X-axis: decile of ChIP-peaks computed from ChIP intensity (10 = decile of top 10%, 1 = decile of lowest 10%)
Y-axis: fraction (%)

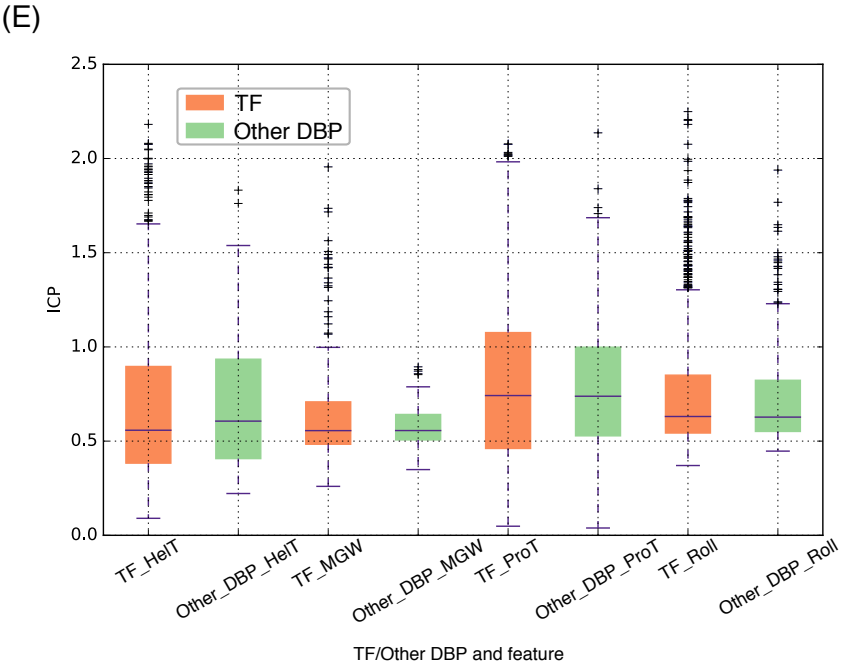



(A)
Sequence-only peaks
TF
Other DBP
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
5
15
25
35
45
55
65
75
85
95
Fraction (%)

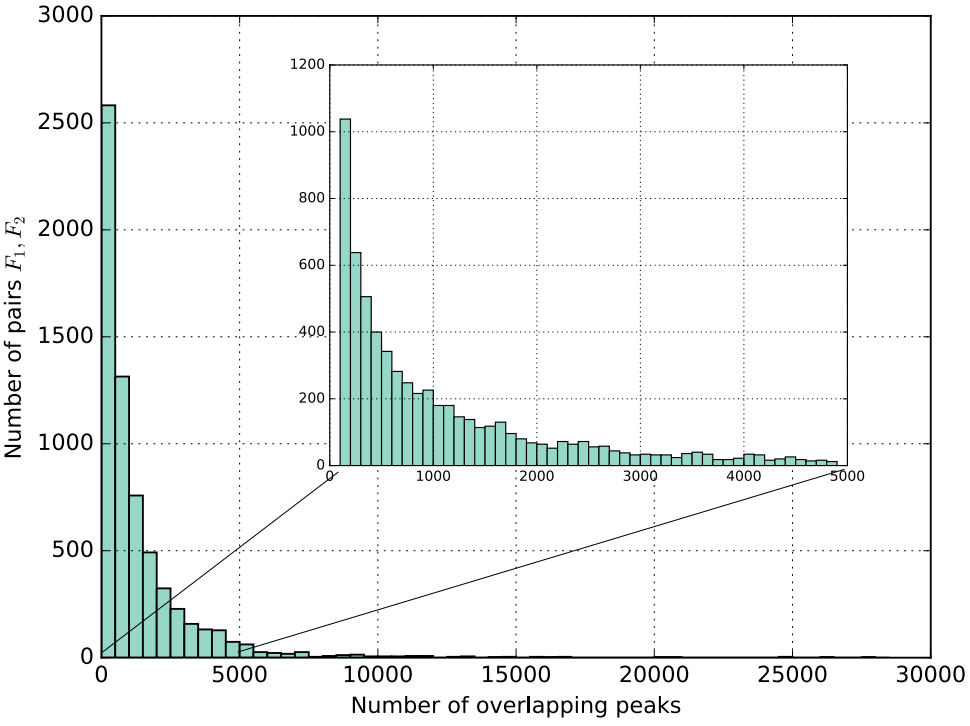

(B)
Shape-only peaks
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
5
15
25
35
45
55
65
75
85
95
Fraction (%)


(C)
Co-occurence peaks
Density
0.00
0.01
0.02
0.03
0.04
0.05
5
15
25
35
45
55
65
75
85
95
Fraction (%)

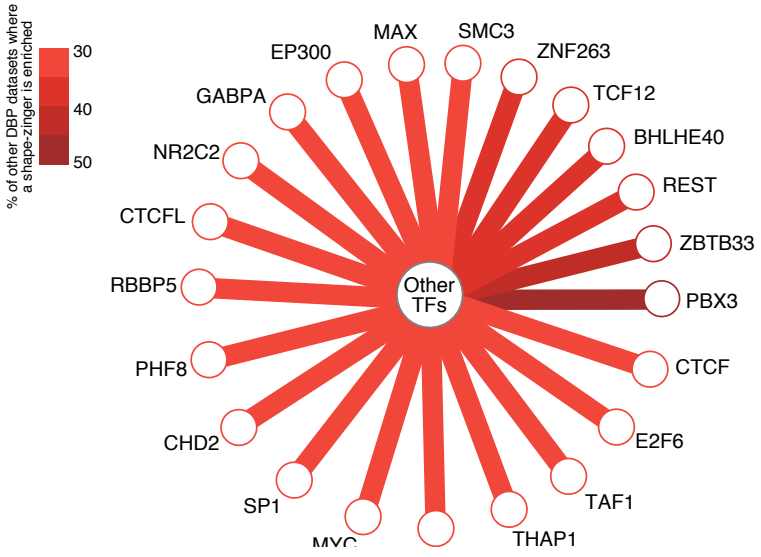

(D)
TF
Other DBP
% of proteins with shape-motif
0
20
40
60
80
HelT
MGW
ProT
Roll


(E)
TF
Other DBP
ICP
0.0
0.5
1.0
1.5
2.0
2.5
TF_HelT
Other_DBP_HelT
TF_MGW
Other_DBP_MGW
TF_ProT
Other_DBP_ProT
TF_Roll
Other_DBP_Roll
TF/Other DBP and feature

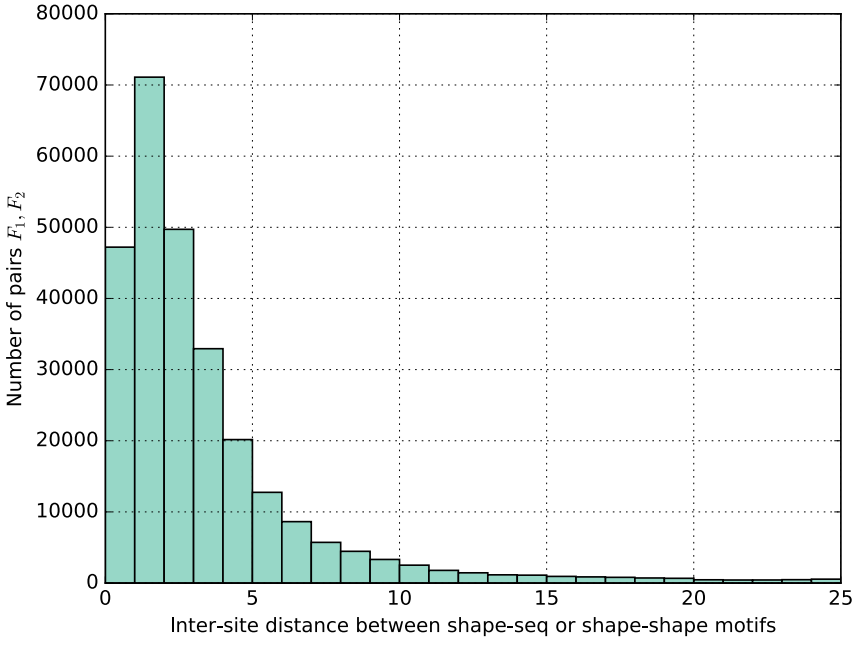
(A)


Number of pairs F_1, F_2
Number of overlapping peaks

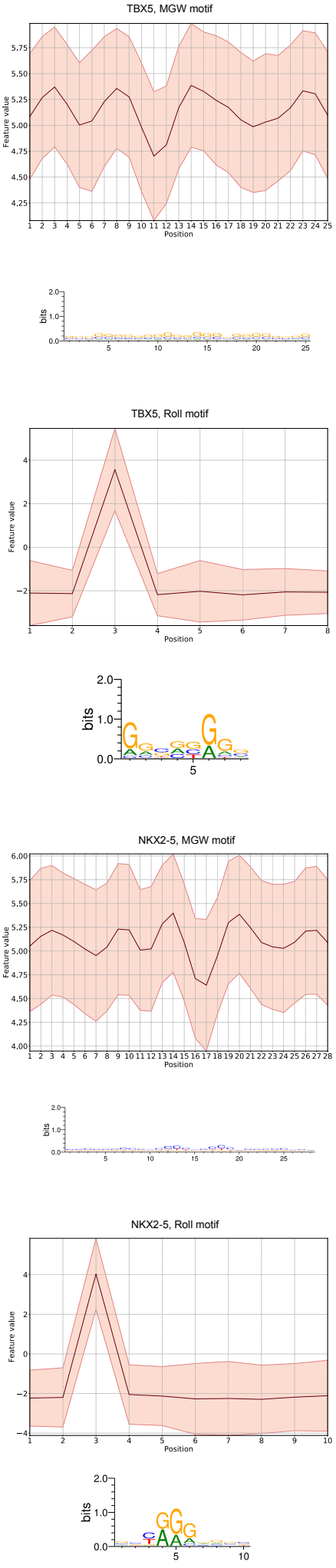
(B)

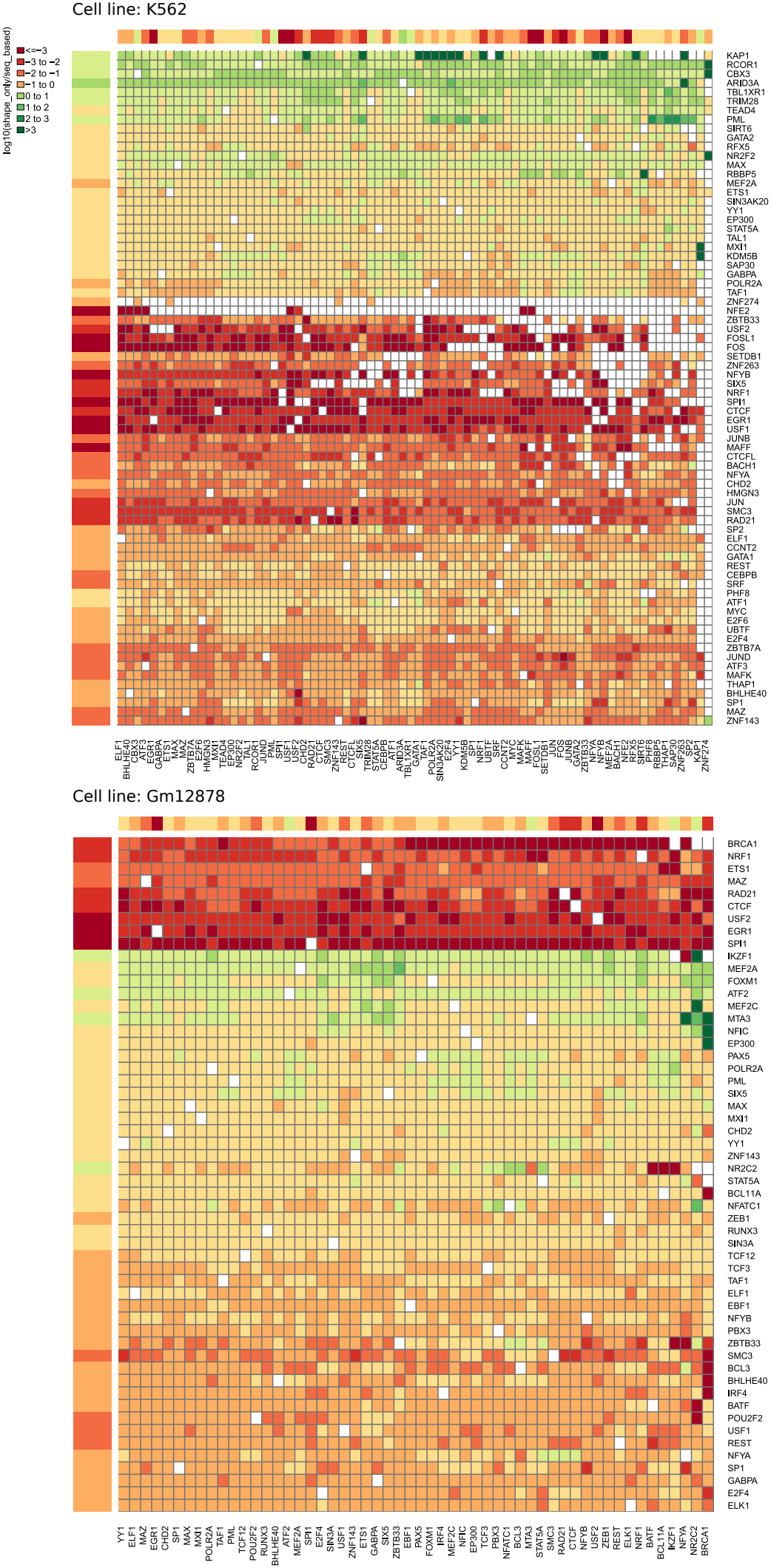

% of other DBP datasets where a shape-zinger is enriched
30
40
50
Other TFs
MAX
SMC3
EP300
ZNF263
GABPA
TCF12
NR2C2
BHLHE40
CTCFL
REST
RBBP5
ZBTB33
PBX3
PHF8
CTCF
CHD2
E2F6
SP1
TAF1
MYC
THAP1

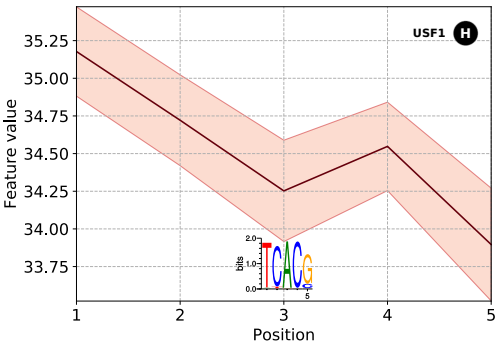
(C)

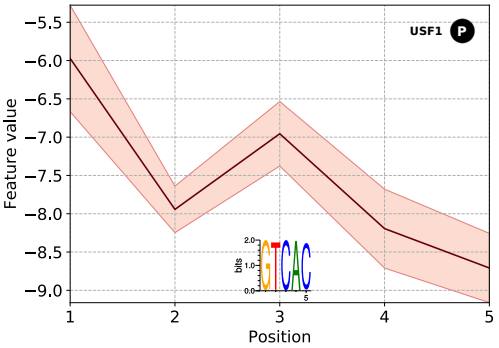

Number of pairs F_1, F_2
Inter-site distance between shape-seq or shape-shape motifs

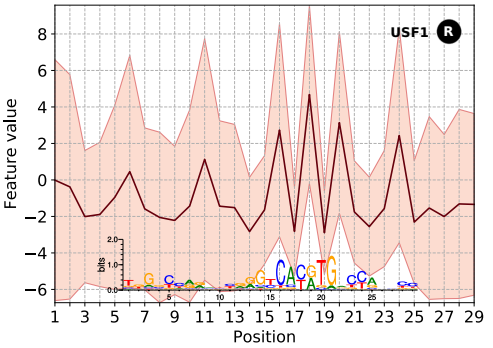
(D)

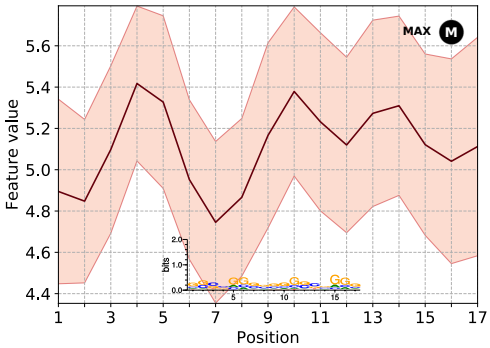

TBX5, MGW motif
Feature value
Position
bits
TBX5, Roll motif
Feature value
Position
bits
NKX2-5, MGW motif
Feature value
Position
bits
NKX2-5, Roll motif
Feature value
Position
bits

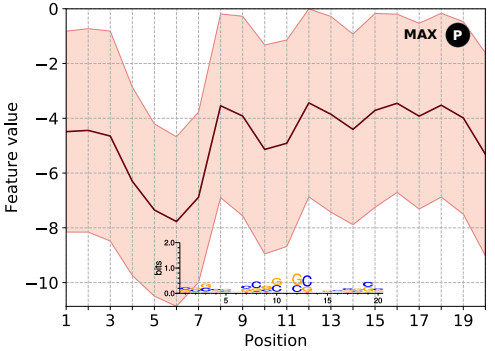

Cell line: K562
log10(shape_only/seq_based)
<=−3
−3 to −2
−2 to −1
−1 to 0
0 to 1
1 to 2
2 to 3
>3
KAP1
RCOR1
CBX3
ARID3A
TBL1XR1
TRIM28
TEAD4
PML
SIRT6
GATA2
RFX5
NR2F2
MAX
RBBP5
MEF2A
ETS1
SIN3AK20
YY1
EP300
STAT5A
TAL1
MXI1
KDM5B
SAP30
GABPA
POLR2A
TAF1
ZNF274
NFE2
ZBTB33
USF2
FOSL1
FOS
SETDB1
ZNF263
NFYB
SIX5
NRF1
SPI1
CTCF
EGR1
USF1
JUNB
MAFF
CTCFL
BACH1
NFYA
CHD2
HMGN3
JUN
SMC3
RAD21
SP2
ELF1
CCNT2
GATA1
REST
CEBPB
SRF
PHF8
ATF1
MYC
E2F6
UBTF
E2F4
ZBTB7A
JUND
ATF3
MAFK
THAP1
BHLHE40
SP1
MAZ
ZNF143
ELF1
BHLHE40
ATF3
EGR1
GABPA
ETS1
MAX
ZBTB7A
E2F6
HMGN3
MXI1
TEAD4
EP300
NR2F2
JUNB
RCOR1
JUND
PML
SPI1
USF2
CHD2
RAD21
CTCF
SMC3
ZNF143
REST
CTCFL
SIX5
TRIM28
STAT5A
CEBPB
SIRT6
ARID3A
TBL1XR1
GATA1
MAFF
POLR2A
SIN3AK20
E2F4
YY1
KDM5B
SP1
NRF1
UBTF
SRF
CCNT2
MYC
MAFK
THAP1
FOSL1
SETDB1
JUN
FOS
JUNB
GATA2
ZBTB33
NFYA
NFYB
MEF2A
BACH1
NFE2
RFX5
SIRT6
PHF8
RBBP5
THAP1
STAT5A
ZNF263
SP2
KAP1
ZNF274
Cell line: Gm12878
BRCA1
NRF1
ETS1
MAZ
RAD21
CTCF
USF2
EGR1
SPI1
IKZF1
MEF2A
FOXM1
ATF2
MEF2C
MTA3
NFIC
EP300
PAX5
POLR2A
PML
SIX5
MAX
MXI1
CHD2
YY1
ZNF143
NR2C2
STAT5A
BCL11A
NFATC1
ZEB1
RUNX3
SIN3A
TCF12
TCF3
TAF1
ELF1
EBF1
NFYB
PBX3
ZBTB33
SMC3
BCL3
BHLHE40
IRF4
BATF
POU2F2
USF1
REST
NFYA
SP1
GABPA
E2F4
ELK1
YY1
ELF1
MAZ
EGR1
CHD2
SPI1
MAX
MXI1
POLR2A
TAF1
PML
TCF12
POU2F2
RUNX3
BHLHE40
ATF2
MEF2A
SPI1
E2F4
SIN3A
USF1
ZNF143
ETS1
GABPA
SIX5
ZBTB33
EBF1
PAX5
FOXM1
IRF4
MEF2C
NFIC
EP300
TCF3
PBX3
NFATC1
BCL3
MTA3
STAT5A
SMC3
RAD21
CTCF
NFYB
USF2
ZEB1
REST
ELK1
NRF1
BATF
BCL11A
IKZF1
NFYA
NR2C2
BRCA1

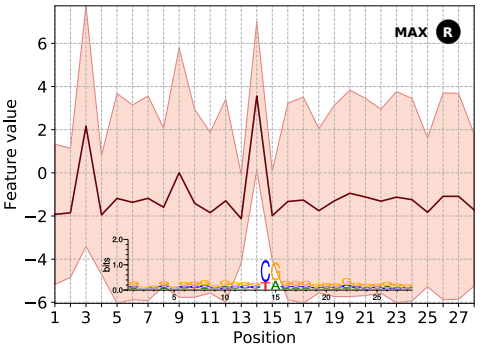

USF1 H
Feature value
Position

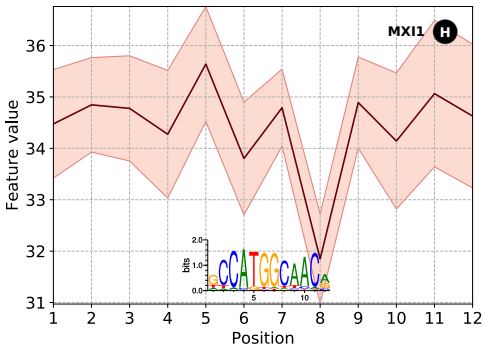

USF1 P
Feature value
Position

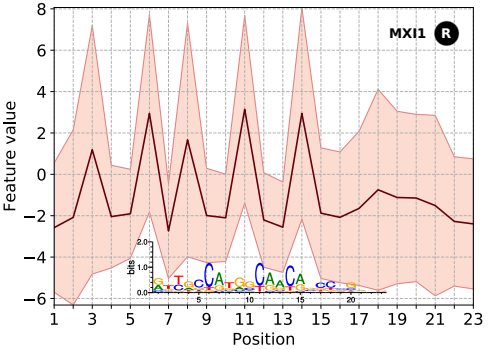

USF1 R
Feature value
Position

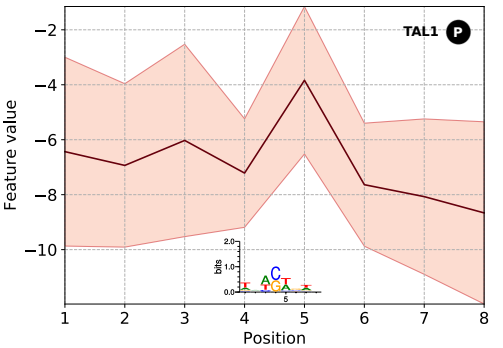

MAX M
Feature value
Position

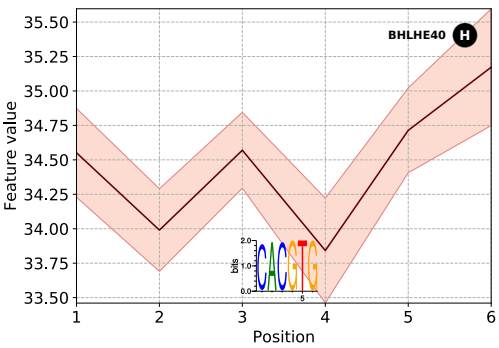

MAX P
Feature value
Position

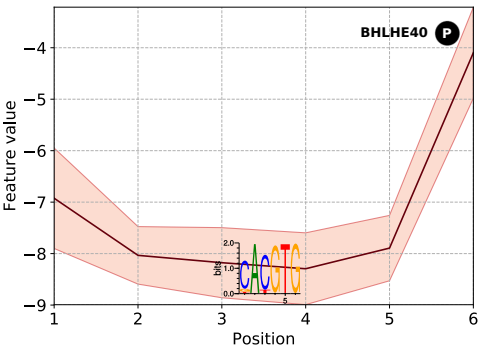

MAX R
Feature value
Position

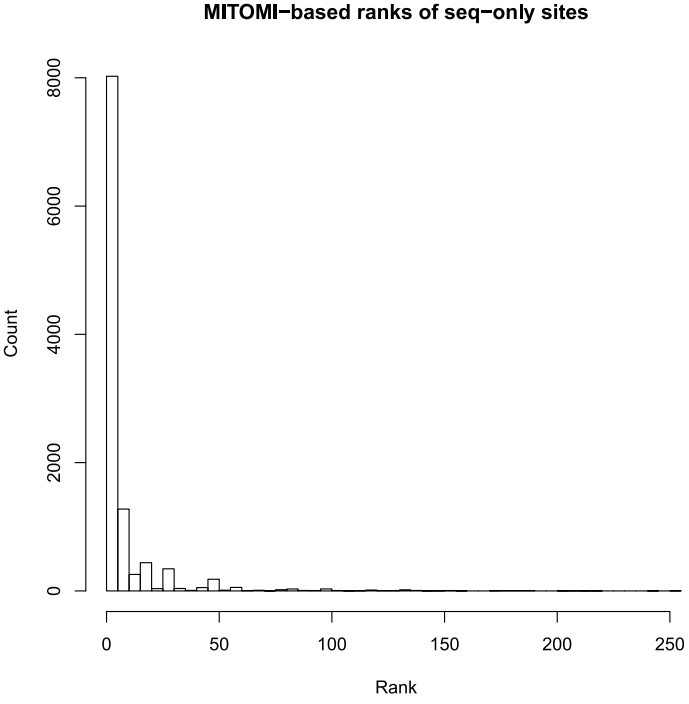

MXI1 H
Feature value
Position

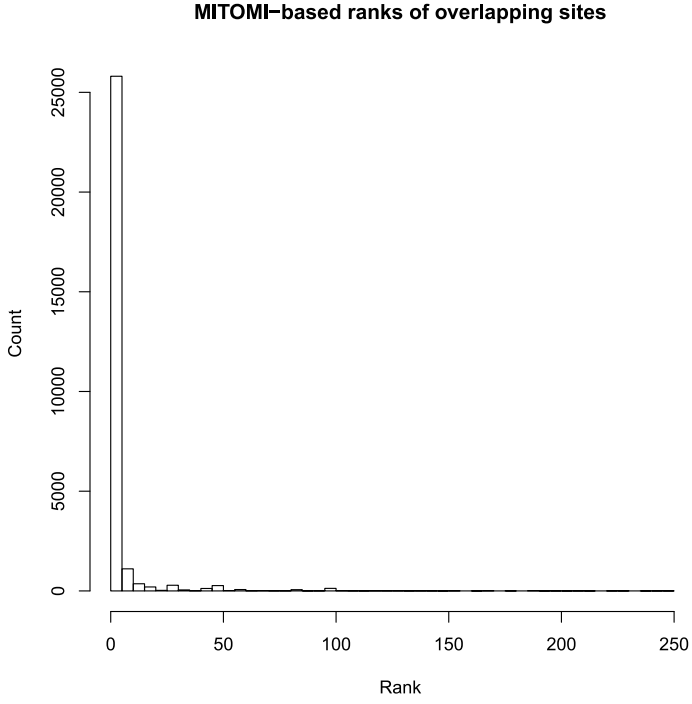

MXI1 R
Feature value
Position

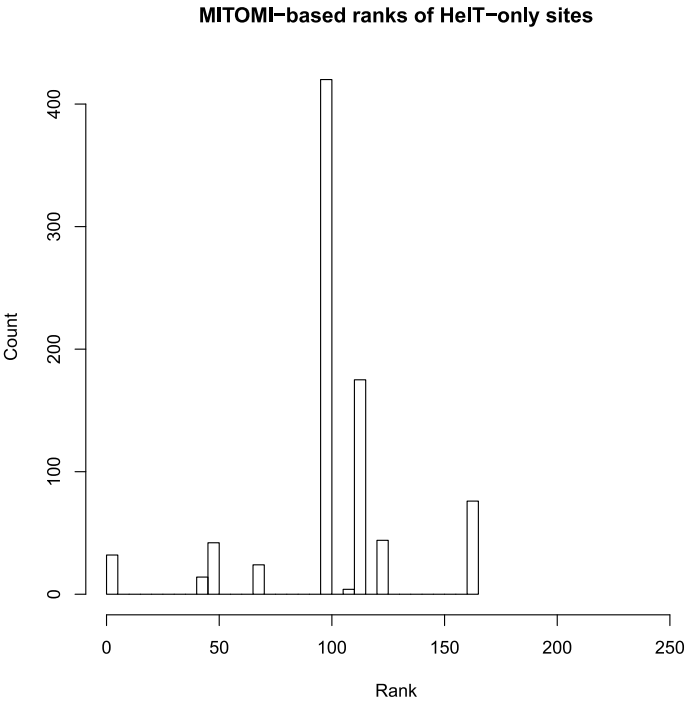

TAL1 P
Feature value
Position

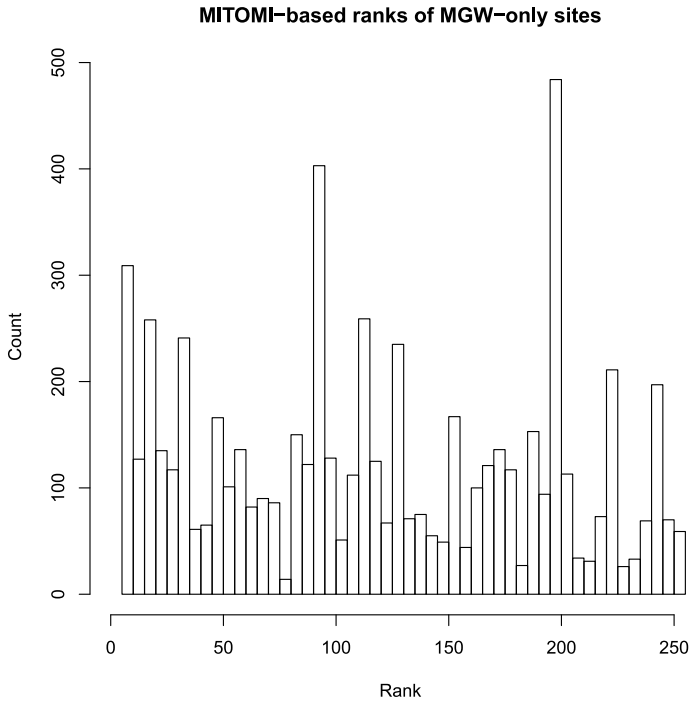

BHLHE40 H
Feature value
Position

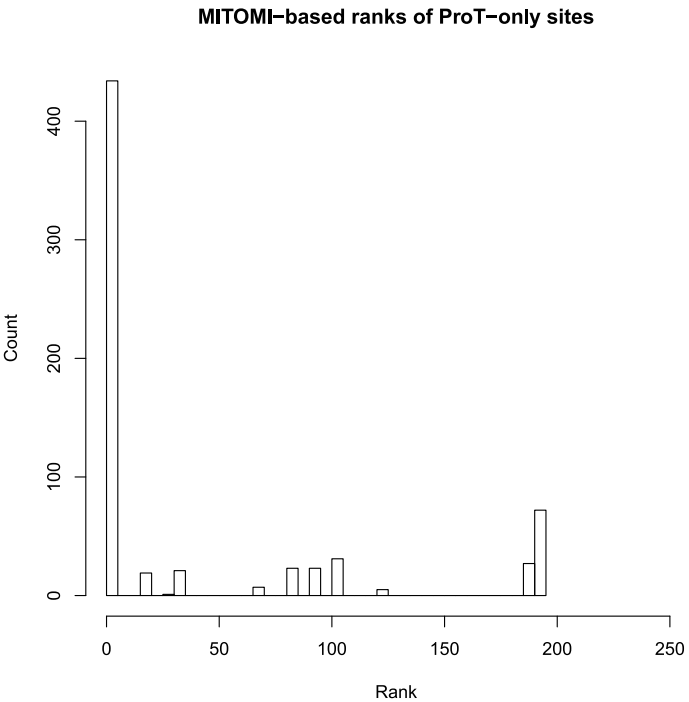

BHLHE40 P
Feature value
Position

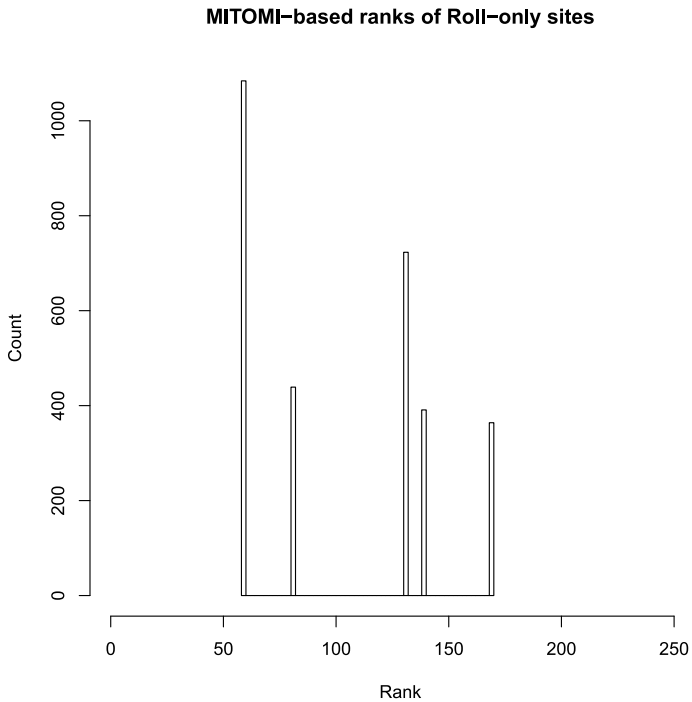

MITOMI−based ranks of seq−only sites
Count
Rank
0
2000
4000
6000
8000
0
50
100
150
200
250


MITOMI−based ranks of overlapping sites
Count
Rank
0
5000
10000
15000
20000
25000
0
50
100
150
200
250


MITOMI−based ranks of HelT−only sites
Count
Rank
0
100
200
300
400
0
50
100
150
200
250


MITOMI−based ranks of MGW−only sites
Count
Rank
0
100
200
300
400
500
0
50
100
150
200
250


MITOMI−based ranks of ProT−only sites
Count
Rank
0
100
200
300
400
0
50
100
150
200
250


MITOMI−based ranks of Roll−only sites
Count
Rank
0
200
400
600
800
1000
0
50
100
150
200
250

Supplementary Table 1, related to Figure 4.
Average number of sequence-only, shape-only, and overlapping sites per peak

DBP	Cell line	Sequence-only sites	Shape-only sites	Overlapping-sites
ARID3A	K562	0.26	1.78	0.04
ATF1	K562	0.59	0.23	1.07
ATF2	Gm12878	0.52	0.71	0.37
ATF3	K562	1.61	0.23	1.16
BACH1	K562	1.45	0.24	4.23
BATF	Gm12878	0.47	0.20	1.53
BCL11A	Gm12878	1.08	0.88	0.29
BCL3	Gm12878	0.96	0.23	1.36
BHLHE40	Gm12878	1.12	0.16	1.27
BHLHE40	K562	1.10	0.14	1.76
BRCA1	Gm12878	1.84	0.20	6.12
CBX3	K562	0.32	0.99	0.08
CCNT2	K562	1.33	0.37	0.88
CEBPB	K562	2.81	0.55	1.73
CHD2	Gm12878	0.63	1.25	0.63
CHD2	K562	1.88	0.37	1.12
CTCF	Gm12878	4.83	0.29	3.74
CTCF	K562	4.74	0.19	3.37
CTCFL	K562	1.86	0.40	2.23
E2F4	Gm12878	1.78	1.17	1.88
E2F4	K562	1.66	0.79	2.17
E2F6	K562	0.92	1.61	1.31
EBF1	Gm12878	2.27	0.23	1.68
EGR1	Gm12878	5.85	0.14	5.74
EGR1	K562	5.72	0.13	4.43
ELF1	Gm12878	1.79	0.81	1.58
ELF1	K562	2.53	1.02	2.65
ELK1	Gm12878	2.27	1.05	1.66
EP300	Gm12878	0.78	1.02	0.15
EP300	K562	0.91	1.10	0.52
ETS1	Gm12878	2.12	0.63	1.57
ETS1	K562	0.75	1.69	0.48
FOS	K562	2.68	0.05	3.64
FOSL1	K562	2.33	0.03	3.24
FOXM1	Gm12878	0.40	1.04	0.29
GABPA	Gm12878	1.48	1.18	2.51
GABPA	K562	0.76	1.59	1.63
GATA1	K562	3.14	1.54	1.38
GATA2	K562	1.27	1.37	0.55
HMGN3	K562	1.54	0.35	2.93
IKZF1	Gm12878	0.31	3.58	0.09
IRF4	Gm12878	1.39	0.30	0.71
JUN	K562	3.01	0.15	2.74
JUNB	K562	2.83	0.17	2.75
JUND	K562	2.41	0.09	1.55

DBP	Cell line	Sequence-only sites	Shape-only sites	Overlapping-sites
KAP1	K562	0.58	3.89	0.15
KDM5B	K562	0.66	1.57	1.17
MAFF	K562	2.64	0.03	1.67
MAFK	K562	2.82	0.21	3.36
MAX	Gm12878	0.85	1.71	1.34
MAX	K562	0.53	3.33	1.61
MAZ	Gm12878	2.30	0.19	3.11
MAZ	K562	1.91	0.33	2.12
MEF2A	Gm12878	1.48	1.54	0.30
MEF2A	K562	3.23	1.08	0.19
MEF2C	Gm12878	1.52	1.73	0.34
MTA3	Gm12878	0.47	1.57	0.33
MXI1	Gm12878	0.45	1.01	0.31
MXI1	K562	1.11	0.63	0.82
MYC	K562	0.78	0.80	1.64
NFATC1	Gm12878	0.69	1.11	0.03
NFE2	K562	1.78	0.02	4.75
NFIC	Gm12878	0.42	0.57	0.42
NFYA	Gm12878	1.78	1.32	3.53
NFYA	K562	3.33	0.39	4.01
NFYB	Gm12878	1.28	0.77	2.72
NFYB	K562	4.10	0.02	1.67
NR2C2	Gm12878	1.85	4.17	0.17
NR2F2	K562	1.05	1.34	0.72
NRF1	Gm12878	4.35	0.36	6.99
NRF1	K562	4.57	0.29	7.21
PAX5	Gm12878	1.13	0.84	0.54
PBX3	Gm12878	1.00	0.25	1.07
PHF8	K562	0.54	0.77	0.97
PML	Gm12878	0.34	0.82	0.52
PML	K562	0.48	2.19	0.63
POLR2A	Gm12878	0.49	1.56	1.01
POLR2A	K562	0.62	0.69	1.36
POU2F2	Gm12878	1.06	0.04	1.61
RAD21	Gm12878	2.59	0.10	1.50
RAD21	K562	2.37	0.17	1.75
RBBP5	K562	0.36	1.49	0.69
RCOR1	K562	0.36	2.55	0.20
REST	Gm12878	3.24	0.23	2.77
REST	K562	2.03	0.27	1.69
RFX5	K562	1.35	0.62	0.36
RUNX3	Gm12878	0.63	0.42	1.23
SAP30	K562	0.43	1.30	0.85
SETDB1	K562	0.79	0.13	0.23
SIN3A	Gm12878	0.50	1.06	0.74
SIN3AK20	K562	0.49	1.51	0.42
SIRT6	K562	0.78	1.33	0.53
SIX5	Gm12878	1.45	2.03	2.20

DBP	Cell line	Sequence-only sites	Shape-only sites	Overlapping-sites
SIX5	K562	1.66	0.05	3.15
SMC3	Gm12878	1.95	0.33	2.06
SMC3	K562	2.31	0.20	2.16
SP1	Gm12878	1.75	0.52	2.83
SP1	K562	2.99	1.00	4.33
SP2	K562	1.37	1.70	4.66
SPI1	Gm12878	3.71	0.06	2.82
SPI1	K562	3.78	0.03	4.07
SRF	K562	7.67	0.12	1.93
STAT5A	Gm12878	1.13	0.50	1.59
STAT5A	K562	1.99	1.75	0.64
TAF1	Gm12878	1.08	1.18	1.36
TAF1	K562	0.89	1.46	1.64
TAL1	K562	1.81	0.76	0.21
TBL1XR1	K562	0.82	2.62	0.36
TCF12	Gm12878	1.51	1.49	2.07
TCF3	Gm12878	1.10	1.38	1.39
TEAD4	K562	0.85	1.08	0.46
THAP1	K562	1.26	0.49	1.25
TRIM28	K562	0.57	2.28	0.32
UBTF	K562	1.06	0.83	3.13
USF1	Gm12878	1.83	0.99	3.89
USF1	K562	2.47	0.03	2.90
USF2	Gm12878	3.07	0.02	2.34
USF2	K562	3.29	0.09	4.17
YY1	Gm12878	1.85	1.05	0.74
YY1	K562	2.68	2.52	1.73
ZBTB33	Gm12878	2.02	0.71	3.27
ZBTB33	K562	2.31	0.19	1.62
ZBTB7A	K562	1.39	0.60	2.35
ZEB1	Gm12878	1.84	2.06	2.60
ZNF143	Gm12878	0.58	0.55	0.14
ZNF143	K562	1.34	0.12	0.18
ZNF263	K562	4.79	0.33	5.80
ZNF274	K562	1.29	0.27	1.42

Supplementary Table 2, related to Figures 2 and 4.
Mean information content per position of shape motifs

DBP	Cell line	HelT	MGW	ProT	Roll
ATF1	K562	0.38		0.44	
ATF2	Gm12878	0.45		0.46	
ATF3	K562	0.47		0.58	
BACH1	K562	0.70	1.27	0.79	0.58
BATF	Gm12878	0.72		0.74	
BCL11A	Gm12878	0.41		0.61	
BCL3	Gm12878	1.02		0.84	0.76
BHLHE40	Gm12878	0.64		0.64	
BHLHE40	K562	0.60		0.56	
BRCA1	Gm12878	0.75	0.55	0.76	
CBX3	K562	0.58			
CCNT2	K562		0.73	0.73	0.92
CEBPB	K562	0.32		0.10	
CHD2	Gm12878		0.55	0.66	0.76
CHD2	K562	0.56	0.45	0.79	0.66
CTCF	Gm12878	0.86	0.56	0.86	0.68
CTCF	K562	1.01	0.59	0.88	0.71
CTCFL	K562	0.91	0.78	0.94	0.87
E2F4	Gm12878		0.53	0.72	
E2F4	K562	0.89	0.75	0.98	0.77
E2F6	K562	0.49	0.57	0.95	0.81
EBF1	Gm12878		0.74		0.68
EGR1	Gm12878	1.05	0.77	1.14	0.84
EGR1	K562	0.98	0.73	0.90	0.90
ELF1	Gm12878	0.92	0.58	0.89	0.64
ELF1	K562	0.55	0.51	0.96	0.52
ELK1	Gm12878		0.50	0.68	0.47
EP300	Gm12878	0.37		0.43	
EP300	K562	0.68		0.34	
ETS1	Gm12878		0.52	0.77	
ETS1	K562	1.22	0.56	0.90	0.72
FOS	K562	0.57		1.06	
FOSL1	K562	0.61		0.74	
FOXM1	Gm12878	0.38		0.18	
GABPA	Gm12878	0.57	0.52	1.17	0.61
GABPA	K562	0.27	0.47	0.82	0.52
GATA1	K562			0.61	
GATA2	K562	1.06		0.35	
HMGN3	K562		0.68	0.96	0.79
IKZF1	Gm12878	0.48		0.34	
IRF4	Gm12878	0.66		0.71	
JUN	K562	0.51		0.54	
JUNB	K562	0.65		0.65	
JUND	K562	0.49		0.85	
KAP1	K562	0.50		0.17	

DBP	Cell line	HelT	MGW	ProT	Roll
KDM5B	K562		0.49	0.77	0.75
MAFF	K562	0.91	1.47	0.42	
MAFK	K562	0.90	1.31	0.45	0.64
MAX	Gm12878	0.67	0.52	0.89	1.13
MAX	K562	0.60	0.44	0.74	0.82
MAZ	Gm12878	1.03	0.69	1.05	0.81
MAZ	K562	1.06	0.62	0.96	0.74
MEF2A	Gm12878	0.51		0.32	
MEF2A	K562	0.74			
MEF2C	Gm12878	0.41		0.14	
MTA3	Gm12878	0.34		0.43	
MXI1	Gm12878		0.50	0.80	
MXI1	K562	0.16			
MYC	K562	0.89	0.50	0.66	
NFATC1	Gm12878	0.66		0.68	
NFE2	K562	0.67		0.89	
NFIC	Gm12878	0.56		0.43	
NFYA	Gm12878	1.14	0.40		
NFYA	K562	0.95	0.57	0.77	0.56
NFYB	Gm12878	1.24	0.51	0.58	1.16
NFYB	K562	0.77	0.78	1.29	1.11
NR2C2	Gm12878	0.93	0.29	0.77	0.87
NR2F2	K562	0.47		0.43	
NRF1	Gm12878	0.90	0.60	1.22	1.12
NRF1	K562	0.67	0.53	1.07	
PAX5	Gm12878	0.34		0.49	
PBX3	Gm12878	1.04		1.13	0.54
PHF8	K562		0.58	0.94	0.62
PML	Gm12878		0.44	0.57	
PML	K562			0.26	
POLR2A	Gm12878	0.99	0.51	0.87	0.65
POLR2A	K562		0.55	0.94	0.70
POU2F2	Gm12878	1.01			
RAD21	Gm12878			0.69	0.64
RAD21	K562	1.08	0.66	0.71	0.70
RBBP5	K562	1.16	0.54	0.89	0.93
RCOR1	K562	0.49		0.78	
REST	Gm12878	0.80	0.41	0.91	0.74
REST	K562	0.73	0.37	1.03	0.61
RFX5	K562	0.50			
RUNX3	Gm12878	0.63		0.34	
SAP30	K562		0.53	0.83	0.70
SETDB1	K562				0.71
SIN3A	Gm12878		0.52	0.79	0.70
SIN3AK20	K562		0.55	0.84	0.84
SIRT6	K562			0.18	
SIX5	Gm12878	0.59	0.47	1.00	0.64
SIX5	K562	0.60		0.84	0.77

DBP	Cell line	HelT	MGW	ProT	Roll
SMC3	Gm12878	1.02	0.73	0.67	0.76
SMC3	K562	1.09	0.58	0.71	0.77
SP1	Gm12878	0.83	0.48	0.76	0.93
SP1	K562		0.47		1.19
SP2	K562		0.50	0.60	1.14
SPI1	Gm12878	1.30		1.12	0.91
SPI1	K562	1.38		1.02	0.86
SRF	K562			1.08	
STAT5A	Gm12878	0.53		0.44	
STAT5A	K562			0.21	
TAF1	Gm12878	1.12	0.57	1.03	0.70
TAF1	K562	1.22	0.57	1.02	0.72
TAL1	K562			0.70	
TBL1XR1	K562	0.53		0.31	
TCF12	Gm12878			0.29	0.57
TCF3	Gm12878			0.26	
TEAD4	K562	0.54		0.30	
THAP1	K562		0.55	0.94	0.89
TRIM28	K562	0.51		0.56	
UBTF	K562	1.25	0.64	1.04	0.71
USF1	Gm12878	0.81	0.48	0.94	0.93
USF1	K562	0.49		0.87	
USF2	Gm12878	0.68		1.08	
USF2	K562	0.35		0.63	
YY1	Gm12878		0.52	1.03	0.97
YY1	K562	1.24	0.59	1.11	1.24
ZBTB33	Gm12878	0.34	0.43	0.53	0.57
ZBTB33	K562	0.70	0.46	0.68	0.61
ZBTB7A	K562	0.98	0.60	1.02	0.68
ZEB1	Gm12878	0.92	0.49	0.77	0.66
ZNF143	Gm12878		0.54	0.85	0.69
ZNF143	K562			1.54	0.60
ZNF263	K562	0.99	0.50	0.79	0.89
ZNF274	K562	0.64	1.40	0.76	1.55

Supplementary Table 3, related to Figure 4.
Fraction (%) of top peaks that are sequence-only, shape-only, and co-occurrence.

Every cell shows three values in the format $f_s/f_{sh}/f_{ov}$, where f_s, f_{sh}, and f_{ov} denote the fraction of sequence-only, shape-only, and co-occurrence peaks, respectively.

DBP	Cell line	Top 500 peaks	Top 1000 peaks	Top 2000 peaks	Top 5000 peaks	All peaks
ARID3A	K562	8.0/58.0/13.4	7.1/56.8/12.5	6.5/57.4/11.2	5.8/55.8/10.1	5.4/55.7/10.0
ATF1	K562	27.8/2.8/54.0	28.2/5.6/46.2	25.4/9.0/37.0	20.3/11.6/25.4	16.8/13.0/18.5
ATF2	Gm12878	15.2/31.8/34.0	15.9/29.0/31.3	13.9/31.2/28.8	12.8/32.5/23.8	10.3/34.3/16.7
ATF3	K562	32.0/6.4/45.0	35.2/5.7/46.4	37.6/4.2/47.3	39.0/4.7/43.3	38.8/5.7/38.0
BACH1	K562	10.2/1.2/87.6	14.0/1.2/82.8	17.6/2.6/75.5	-/-/-	20.0/4.0/67.4
BATF	Gm12878	21.8/7.0/58.8	21.6/6.6/57.6	21.9/6.8/52.7	21.2/7.7/49.5	18.2/9.6/39.3
BCL11A	Gm12878	28.0/22.4/35.8	26.7/21.5/37.2	27.2/21.6/35.7	26.9/21.5/33.2	25.5/23.4/29.4
BCL3	Gm12878	28.2/4.0/52.2	30.9/4.8/44.9	30.7/5.1/39.8	30.9/5.5/28.7	30.2/5.4/21.0
BHLHE40	Gm12878	36.4/1.0/49.2	36.6/1.8/45.9	36.7/3.8/40.4	34.7/5.8/29.8	30.1/7.4/23.4
BHLHE40	K562	37.2/1.6/51.2	38.5/2.7/47.2	38.0/3.5/42.2	35.7/4.3/35.7	28.4/5.5/26.7
BRCA1	Gm12878	-/-/-	-/-/-	-/-/-	-/-/-	17.9/1.9/76.9
CBX3	K562	13.0/37.2/33.6	12.2/40.9/28.6	10.6/43.6/24.3	8.6/47.5/17.8	7.2/50.1/12.9
CCNT2	K562	39.8/4.8/33.4	37.0/5.3/35.6	35.8/6.2/36.9	33.4/8.0/34.4	30.9/9.5/29.2
CEBPB	K562	37.6/1.0/60.8	36.9/1.3/60.7	38.3/1.5/58.8	39.9/3.2/54.1	41.4/5.4/47.0
CHD2	Gm12878	17.8/11.8/65.8	16.5/14.0/61.8	16.2/16.6/53.5	12.6/20.5/41.2	11.0/24.0/32.8
CHD2	K562	40.8/4.0/46.4	41.2/5.2/42.9	40.0/6.8/39.0	-/-/-	38.3/7.5/33.7
CTCF	Gm12878	27.8/0.2/71.6	28.8/0.1/70.8	29.8/0.1/69.8	32.2/0.2/67.3	41.7/1.3/55.1
CTCF	K562	37.6/0.0/62.4	36.4/0.0/63.4	35.5/0.0/64.2	37.8/0.1/61.7	47.5/1.2/48.7
CTCFL	K562	24.6/2.6/71.2	26.2/2.1/69.7	28.6/2.0/66.8	32.1/2.6/61.6	34.8/3.1/57.6
E2F4	Gm12878	29.8/8.4/55.2	26.4/11.0/53.8	-/-/-	-/-/-	21.3/14.1/48.9
E2F4	K562	24.8/3.8/62.0	22.2/5.3/62.3	19.2/8.6/59.2	-/-/-	18.8/9.3/57.7
E2F6	K562	12.0/7.6/77.4	11.3/9.2/75.1	11.8/12.6/70.0	11.6/17.1/62.3	11.6/19.8/57.1
EBF1	Gm12878	43.2/0.6/49.8	43.1/1.3/47.8	41.4/2.3/44.3	38.8/3.9/38.4	32.5/7.7/27.4
EGR1	Gm12878	31.8/0.0/67.6	32.8/0.0/66.8	36.1/0.0/63.4	40.7/0.3/57.5	45.5/0.8/49.3
EGR1	K562	36.8/0.0/62.6	38.9/0.0/60.4	40.3/0.1/59.0	44.6/0.1/54.7	52.3/0.5/44.0
ELF1	Gm12878	13.0/2.0/83.8	15.0/2.7/79.6	17.9/4.3/74.8	23.0/9.3/61.3	24.5/17.7/39.6
ELF1	K562	13.8/1.0/85.2	17.0/1.7/81.1	20.1/2.5/76.2	25.1/3.8/68.6	29.9/9.0/54.9
ELK1	Gm12878	26.6/4.6/63.8	26.3/6.8/59.8	23.3/13.8/47.5	-/-/-	22.0/16.3/42.8
EP300	Gm12878	25.0/24.6/32.0	25.1/24.5/32.8	24.4/24.4/32.8	25.1/23.6/31.3	25.1/24.1/26.9
EP300	K562	16.8/19.8/51.2	16.1/22.7/50.9	15.9/23.9/49.8	15.6/26.3/45.8	14.0/30.7/37.5
ETS1	Gm12878	39.6/2.8/51.6	37.3/3.5/51.7	-/-/-	-/-/-	34.2/4.8/50.1
ETS1	K562	21.4/21.0/48.4	18.8/22.6/48.8	16.3/26.2/46.2	-/-/-	15.8/28.0/42.9
FOS	K562	27.2/0.0/72.8	28.4/0.0/71.4	31.9/0.0/67.5	36.2/0.3/61.0	36.9/0.6/55.7
FOSL1	K562	25.2/0.0/70.8	28.0/0.2/68.2	32.4/0.1/63.4	36.6/0.4/57.0	37.6/0.5/51.7
FOXM1	Gm12878	11.0/30.4/43.2	12.8/30.9/41.1	12.2/31.8/39.6	11.3/33.2/34.1	9.8/35.5/26.2
GABPA	Gm12878	13.8/1.4/84.6	16.5/2.8/78.5	18.1/6.7/70.7	-/-/-	18.1/14.5/59.3
GABPA	K562	10.8/3.0/84.6	11.3/6.3/78.1	12.3/11.4/64.5	12.1/19.3/46.2	11.5/22.7/40.8
GATA1	K562	16.2/14.4/66.4	16.1/17.4/62.6	15.1/20.1/59.3	-/-/-	15.3/21.2/57.7
GATA2	K562	14.4/23.2/53.4	14.4/25.2/50.9	15.4/27.1/46.4	14.9/30.9/41.1	14.4/33.3/37.1
HMGN3	K562	19.0/2.2/61.0	17.6/2.9/62.3	17.6/3.4/61.1	18.6/4.6/57.0	19.1/4.8/55.5
IKZF1	Gm12878	1.0/64.2/30.6	1.2/63.6/31.7	1.1/64.3/31.0	1.8/67.4/26.6	1.8/67.5/26.2
IRF4	Gm12878	31.8/9.4/47.6	35.6/8.7/43.4	36.5/8.9/41.2	36.4/8.7/37.4	34.5/9.9/28.8

DBP	Cell line	Top 500 peaks	Top 1000 peaks	Top 2000 peaks	Top 5000 peaks	All peaks
JUN	K562	30.8/0.2/67.4	34.6/0.7/62.7	37.1/0.9/59.2	42.3/1.7/51.1	42.9/2.0/48.1
JUNB	K562	33.2/0.8/64.2	34.7/1.3/61.1	36.5/1.6/57.1	37.6/2.6/49.5	35.8/4.2/42.6
JUND	K562	38.2/0.4/60.6	41.9/0.4/55.9	44.8/0.6/51.8	48.3/1.3/43.3	41.0/3.3/27.4
KAP1	K562	3.6/45.6/49.6	3.1/49.2/45.8	3.5/51.6/42.5	3.0/55.4/38.2	3.1/55.7/37.8
KDM5B	K562	7.0/31.0/49.8	5.5/30.8/50.9	6.3/30.4/49.2	6.2/33.5/43.6	6.2/33.7/43.0
MAFF	K562	36.0/0.2/63.8	42.9/0.1/56.2	46.5/0.2/51.2	52.0/0.3/42.9	58.2/0.8/28.3
MAFK	K562	23.4/0.4/76.2	28.5/0.2/70.8	33.7/0.6/64.8	37.3/1.2/59.0	42.3/2.9/47.1
MAX	Gm12878	10.4/10.6/76.2	11.7/15.1/68.1	10.9/21.2/60.8	10.0/31.8/44.9	9.6/34.3/39.8
MAX	K562	2.4/9.6/87.2	3.1/12.0/83.8	3.3/18.8/76.2	4.1/28.2/65.1	3.8/44.8/46.2
MAZ	Gm12878	24.2/0.8/72.4	24.5/0.9/71.7	26.9/1.1/68.5	29.8/2.0/59.5	30.2/2.8/44.7
MAZ	K562	19.8/0.8/78.2	23.3/1.8/72.9	26.4/1.9/68.5	31.3/3.3/59.0	34.0/5.5/42.9
MEF2A	Gm12878	25.8/11.8/57.6	24.5/13.0/57.8	22.5/19.1/52.6	17.7/27.8/45.1	13.6/36.5/32.2
MEF2A	K562	31.6/9.2/55.8	32.1/11.6/52.0	30.1/15.4/48.3	-/-/-	27.3/21.2/41.0
MEF2C	Gm12878	23.2/13.4/60.6	20.0/15.7/61.8	17.4/19.8/58.9	15.5/26.8/48.2	14.8/28.3/45.8
MTA3	Gm12878	10.8/33.4/43.4	9.2/36.3/41.2	9.5/39.6/37.2	9.1/43.5/29.4	8.5/44.8/27.0
MXI1	Gm12878	23.2/17.8/45.0	22.9/18.1/41.1	21.1/20.1/36.9	17.1/22.5/28.6	13.5/21.6/21.7
MXI1	K562	31.2/14.0/44.2	29.5/17.0/34.7	24.1/21.2/26.8	-/-/-	23.9/21.4/26.1
MYC	K562	22.0/6.8/63.2	20.9/9.2/56.7	19.1/12.8/49.0	-/-/-	17.4/16.4/41.2
NFATC1	Gm12878	43.4/21.4/9.0	35.3/27.0/7.4	29.8/27.8/6.2	25.1/23.2/4.7	22.9/21.4/4.2
NFE2	K562	26.6/0.0/67.8	28.6/0.0/64.6	-/-/-	-/-/-	31.7/0.1/60.3
NFIC	Gm12878	17.4/20.8/38.4	18.4/21.3/33.7	17.9/21.6/30.6	18.0/21.1/27.1	14.6/22.4/18.8
NFYA	Gm12878	12.8/14.4/65.2	11.6/16.6/62.5	-/-/-	-/-/-	11.4/16.9/61.8
NFYA	K562	33.2/3.8/57.4	29.0/5.4/60.2	26.2/5.8/59.9	-/-/-	25.5/6.3/59.3
NFYB	Gm12878	19.0/4.2/72.4	20.8/5.5/68.6	21.1/6.2/66.0	21.0/7.8/59.6	19.4/10.2/51.6
NFYB	K562	74.0/0.4/19.4	67.9/0.2/26.2	62.3/0.2/31.1	57.7/0.4/33.1	56.5/0.6/32.3
NR2C2	Gm12878	32.4/43.0/13.0	-/-/-	-/-/-	-/-/-	22.2/54.3/9.5
NR2F2	K562	9.6/27.8/52.0	9.9/30.9/47.4	10.2/33.7/44.1	10.2/37.2/38.0	10.3/38.6/33.6
NRF1	Gm12878	17.0/0.0/83.0	20.5/0.0/79.3	25.1/0.2/74.2	-/-/-	27.9/1.9/65.9
NRF1	K562	28.8/0.0/70.4	32.8/0.7/65.5	-/-/-	-/-/-	32.7/1.6/58.1
PAX5	Gm12878	28.6/10.6/55.4	26.9/14.2/51.3	25.1/19.3/46.5	22.5/25.2/37.9	21.7/28.3/30.2
PBX3	Gm12878	31.8/2.8/60.2	33.0/4.6/53.8	31.4/5.9/49.4	28.0/7.5/40.0	26.3/8.9/35.5
PHF8	K562	14.0/20.2/43.0	13.3/18.5/45.4	12.7/18.6/47.0	11.5/17.5/46.4	10.6/17.2/41.3
PML	Gm12878	11.0/24.2/36.4	9.7/23.2/40.0	8.8/22.4/38.3	7.4/19.5/30.6	6.8/18.8/25.2
PML	K562	5.4/42.6/46.0	5.0/42.3/45.5	4.8/44.0/41.8	4.1/48.3/34.2	3.6/52.2/29.4
POLR2A	Gm12878	7.4/21.0/63.6	7.4/21.6/64.1	7.5/22.6/63.0	6.8/24.3/60.1	4.4/38.4/37.7
POLR2A	K562	15.2/12.6/49.0	14.1/13.5/48.3	13.6/14.6/48.0	12.3/15.2/41.2	11.3/15.0/36.8
POU2F2	Gm12878	30.4/0.4/61.8	29.5/0.7/59.1	31.8/0.8/50.4	28.0/1.8/34.8	19.4/2.4/19.9
RAD21	Gm12878	48.4/0.4/49.6	48.0/0.4/49.9	49.1/0.3/48.8	52.4/0.3/46.0	56.7/1.3/35.2
RAD21	K562	31.8/0.4/66.8	33.5/0.5/65.3	34.9/0.3/64.0	38.1/0.5/60.5	48.7/1.6/45.1
RBBP5	K562	5.4/33.4/47.6	4.9/34.9/45.6	5.0/35.0/42.4	-/-/-	4.0/36.1/34.5
RCOR1	K562	5.2/40.6/48.2	4.6/42.9/47.0	3.6/46.4/44.5	3.6/51.7/38.6	2.6/61.2/26.3
REST	Gm12878	25.2/0.0/74.8	29.9/0.0/70.0	37.0/0.1/62.8	43.3/2.6/50.6	42.9/3.0/49.7
REST	K562	18.4/0.0/81.6	22.8/0.0/77.2	32.0/0.0/67.8	37.9/3.2/48.8	34.3/7.3/34.7
RFX5	K562	28.8/17.6/37.8	23.4/22.9/27.0	-/-/-	-/-/-	23.4/22.9/26.7
RUNX3	Gm12878	20.0/4.6/67.2	20.7/6.0/64.9	21.6/6.5/61.0	21.3/8.2/55.3	17.1/15.9/31.5
SAP30	K562	10.8/26.2/46.8	9.8/26.3/47.2	8.6/24.9/45.4	-/-/-	8.1/25.8/43.0
SETDB1	K562	38.6/7.0/20.0	38.4/6.9/17.9	36.0/6.7/15.8	-/-/-	36.1/6.8/15.4
SIN3A	Gm12878	15.4/19.0/49.4	12.0/21.2/49.5	10.8/22.2/47.0	10.5/22.2/41.5	10.6/22.0/40.7

DBP	Cell line	Top 500 peaks	Top 1000 peaks	Top 2000 peaks	Top 5000 peaks	All peaks
SIN3AK20	K562	16.6/21.2/44.6	16.4/24.1/40.1	14.6/26.9/36.3	-/-/-	13.1/28.0/32.3
SIRT6	K562	18.0/27.4/39.4	17.0/33.4/32.1	-/-/-	-/-/-	16.9/33.4/30.8
SIX5	Gm12878	14.8/9.0/74.6	13.7/12.3/69.9	12.7/23.1/57.0	-/-/-	11.8/26.5/51.9
SIX5	K562	32.4/1.4/58.0	26.1/1.8/56.2	-/-/-	-/-/-	23.3/2.0/46.4
SMC3	Gm12878	20.4/0.2/77.8	22.8/0.7/75.1	26.2/1.2/71.3	30.9/1.3/66.4	35.7/3.9/53.3
SMC3	K562	30.2/0.8/68.2	30.7/0.9/67.7	30.4/0.8/67.7	34.4/0.8/62.8	42.7/1.8/50.7
SP1	Gm12878	25.0/3.2/68.4	23.1/3.5/68.7	21.4/4.7/64.4	20.8/7.3/51.3	19.5/9.6/40.7
SP1	K562	15.0/5.2/75.4	16.7/6.4/70.3	17.2/9.7/63.2	-/-/-	17.6/10.0/61.3
SP2	K562	5.4/9.4/83.2	6.4/11.8/78.7	-/-/-	-/-/-	6.9/13.0/76.2
SPI1	Gm12878	33.4/0.0/66.2	37.9/0.0/61.7	41.0/0.1/58.5	46.4/0.1/52.9	56.1/0.3/40.6
SPI1	K562	28.2/0.0/71.6	32.6/0.0/67.3	36.0/0.0/63.8	41.4/0.0/58.2	49.2/0.2/48.6
SRF	K562	58.8/2.6/24.0	55.6/3.1/22.5	49.9/4.5/19.4	-/-/-	49.2/4.6/18.9
STAT5A	Gm12878	22.8/17.2/34.4	19.0/19.1/31.9	18.8/18.7/27.5	-/-/-	17.1/18.7/24.4
STAT5A	K562	14.0/15.2/67.4	14.5/19.4/61.6	14.0/24.8/54.9	13.3/30.0/48.0	12.8/32.1/45.1
TAF1	Gm12878	11.4/12.2/70.6	11.1/12.0/71.4	10.4/14.2/67.8	9.9/17.2/61.3	9.7/19.5/55.1
TAF1	K562	10.4/12.0/74.0	8.5/13.5/72.7	7.8/16.4/69.9	-/-/-	6.6/21.7/61.6
TAL1	K562	28.6/17.4/39.6	28.4/18.4/37.4	29.9/18.9/34.9	29.9/21.5/31.8	26.7/24.4/27.9
TBL1XR1	K562	9.0/42.2/42.2	8.1/44.3/39.7	8.1/45.4/36.5	-/-/-	7.7/49.3/30.7
TCF12	Gm12878	12.0/4.2/82.4	14.0/4.0/80.5	13.2/6.0/78.5	13.1/10.3/72.2	12.1/17.2/63.7
TCF3	Gm12878	22.0/4.4/70.2	23.0/6.2/66.1	22.8/7.1/63.5	21.1/11.5/57.5	19.2/16.6/51.5
TEAD4	K562	15.4/28.8/44.0	15.2/30.2/42.7	15.4/30.2/42.0	15.4/31.9/37.5	12.4/37.3/27.1
THAP1	K562	33.0/8.0/44.4	-/-/-	-/-/-	-/-/-	28.1/9.9/43.3
TRIM28	K562	5.8/52.4/36.4	5.0/50.7/37.2	4.8/51.6/36.1	5.0/53.6/32.3	4.5/55.2/29.1
UBTF	K562	7.2/8.4/80.6	7.4/10.8/74.8	8.1/12.7/66.0	-/-/-	8.3/12.8/64.9
USF1	Gm12878	5.6/0.6/93.4	9.3/1.5/89.0	14.9/2.1/82.2	19.3/5.1/73.2	20.0/8.1/67.8
USF1	K562	30.4/0.4/68.0	35.2/0.3/62.9	40.8/0.2/56.1	46.4/0.3/48.9	49.1/0.5/41.0
USF2	Gm12878	38.4/0.2/59.2	46.7/0.5/49.5	53.9/0.5/39.5	57.6/0.7/28.5	56.5/0.8/25.7
USF2	K562	44.0/0.4/53.2	46.3/0.3/50.3	-/-/-	-/-/-	43.5/1.1/44.6
YY1	Gm12878	24.2/1.2/74.2	25.5/0.9/72.5	27.9/3.2/65.9	28.7/14.7/47.7	24.7/26.5/24.9
YY1	K562	11.0/1.6/87.2	11.3/5.3/83.1	11.1/13.7/74.0	10.5/29.1/57.2	10.7/30.7/55.1
ZBTB33	Gm12878	21.4/7.2/64.2	-/-/-	-/-/-	-/-/-	19.8/14.2/52.3
ZBTB33	K562	36.8/2.8/51.0	33.8/3.2/46.9	-/-/-	-/-/-	32.7/3.5/42.0
ZBTB7A	K562	19.8/1.6/77.4	22.0/2.0/73.9	20.9/2.0/73.3	22.4/3.6/68.4	26.1/5.3/60.4
ZEB1	Gm12878	11.4/9.2/78.4	12.3/12.4/72.2	11.7/16.3/66.7	-/-/-	12.1/20.5/61.0
ZNF143	Gm12878	39.6/1.0/58.4	36.4/5.1/55.3	29.0/16.4/40.1	21.2/21.3/23.6	17.4/18.3/13.2
ZNF143	K562	39.6/0.2/58.6	41.4/1.4/50.6	49.9/2.2/34.8	57.6/2.9/21.7	53.2/4.5/12.3
ZNF263	K562	18.8/1.6/78.2	22.3/2.6/72.0	25.2/4.8/63.9	-/-/-	24.9/5.8/62.5
ZNF274	K562	-/-/-	-/-/-	-/-/-	-/-/-	24.0/10.5/46.4

Table S4, related to Figure 4. Number of datasets where the highest fraction of peaks is sequence-only, co-occurrence, or shape-only across different affinity bins

Affinity bin	Sequence-only	Co-occurrence	Shape-only
10 (top-most 10%)	7	113	5
9	8	105	7
8	11	97	9
7	17	92	9
6	16	84	13
5	17	82	16
4	21	76	17
3	21	73	20
2	23	70	20
1 (lowest 10%)	24	65	23