Supplementary data

Contents

Supplementary methods

Method Overview

Mocap combines information from motif, chromatin accessibility and a range of sequence-associated genomic features to predict TFBS. We start with a liberal set of candidate motif sites, and create a generic unsupervised method (MocapG) that ranks motif sites based on local accessibility. We then, building on this baseline method, adopt a supervised machine learning scheme (training sparse logistic regression classifiers MocapS) to investigate the contribution of different features to TFBS prediction and probe the diversity in TF binding behaviors. In training MocapS, we model both cell type-specific chromatin accessibility and footprint patterns around motif sites. We investigate (for different types of TFs) the contribution of sequence properties, such as GC/CpG content, surrounding motif sites to TFBS prediction. We use L1-regularized logistic regression to integrate motif-associated genomic features, which we find to be a good balance of performance and interpretability (in comparison to methods such as support vector machines, prior published methods, and methods adopting other regularization schemes). Lastly, we extend the trained classifiers of MocapS to TFCT conditions without ChIP-Seq label in a data-driven manner (MocapX). The goal is to create a TFBS prediction pipeline that balances method sensitivity with selectivity, and be able to generalize the specifically trained models to more factors and cell types via cross-sample TFBS predictions.

Below we detail the curation of motif models extracted from public databases, the computation of motif-associated genomic features, and the implementation of MocapS and MocapX.

Motif models curation

We drew PWMs from two motif collections (Figure S1A). To reduce information redundancy for TFs with multiple motif representations, we filtered motifs based on pairwise PWM comparisons. Briefly, for a given motif pair, we vectorize the PWMs and calculate length-normalized Euclidean distance between the two vectors as follow

$$d_2 = \sqrt{\sum_{i=1}^{v} \frac{(p_{1i} - p_{2i})^2}{v}}$$

where p_1 and p_2 represent the length-normalized PWM vectors, v denotes the length of the vectorized PWMs. For motif pairs of unequal length, we let $v = min(p_1, p_2)$ where we used the alignment that results in the smallest Euclidean distance among all possible ungapped PWM alignments. For motif pairs that are highly similar (with normalized Euclidean distance < 0.14), the PWM of lower entropy was retained (Figure S1B-D).

Detecting motif-associated footprint patterns from chromatin accessibility data

To assess the probability that a footprint profile (F) exists around a motif site, we used a pair of binomial tests to gauge whether the motif region is protected from enzyme digestion as compared to its left and right flanking regions (4).

For DNase-Seq, we scored for the depletion of DNase I cuts at motif sites in a strand-specific manner, as strand imbalance is often conspicuous because of the size selection steps in DNase-Seq protocol (Figure S6). Specifically, on the positive strand, we assessed if a motif region is depleted of cuts as compared to the left flanking region.

$$P(c_m^+; p, n^+) = \sum_{i=0}^{c_m^+} \binom{n^+}{i} p^i (1-p)^{n^+-i}$$

On the negative strand, we assessed if a motif region is depleted of cuts as compared to the right flanking region.

$$P(c_m^-; p, n^-) = \sum_{i=0}^{c_m^-} \binom{n^-}{i} p^i (1-p)^{n^--i}$$

where $p = \frac{l_m}{l_m + l_f}$. p denotes the expected size-normalized cut probability, where l_m and l_f denotes the size of the motif and its flanking region respectively. We empirically set the size of flanking region as 1.75 the size of motif length to optimize ChIP-Seq peak recovery. $n^+ = c_m^+ + c_f^+$ is the total number of cut counts within the motif site c_m^+ and its left flanking region c_f^+ on the positive strand, whereas $n^- = c_m^- + c_f^-$ is the total number of cuts within the motif site c_m^- and its right flanking region c_f^- on the negative strand.

For ATAC-Seq, as the strand imbalance immediately flanking binding motif is less pronounced, we merged strand information to assess if a motif region is depleted of cuts as compared to the left and right flanking region respectively.

$$P(c_m; p, n^L) = \sum_{i=0}^{c_m} \binom{n^L}{i} p^i (1-p)^{n^L-i}$$

$$P(c_m; p, n^R) = \sum_{i=0}^{c_m} \binom{n^R}{i} p^i (1-p)^{n^R-i}$$

where $n^L = c_m + c_f^L$ is the total number of cuts within the motif site c_m and its left flanking region c_f^L on both strands, whereas $n^R = c_m + c_f^R$ is the total number of cuts within the motif site c_m and its right flanking region c_f^R on both strands.

The footprint score was calculated for the DNase I footprint as

$$F = P(c_m^+; p, n^+) \cdot P(c_m^-; p, n^-)$$

and for the ATAC-Seq footprint as

$$F = P(c_m; p, n^L) \cdot P(c_m; p, n^R)$$

We shuffled the per base-pair cut count data around each site and reran the same binomial tests 500 times. A p-value threshold corresponding to a false discovery rate of 0.01 was chosen as the threshold for determining whether a footprint profile exists for each motif site.

Profiling motif-associated sequence features

To profile the sequence environment surrounding TF binding motifs, we calculated GC content and CpG frequency for a region 100bp upstream and downstream of each motif site. We also computed a binary CpG island feature to assess if a motif resides in a larger region of inflated CpG count. A CpG island was assessed with the following criteria (2, 5)

$$CGI(N_{CpG}, N_C, N_G, L) = \begin{cases} 2 & \text{if } \dfrac{N_{CpG} L}{((N_C + N_G)/2)^2} > 0.6 \ \& \ \dfrac{N_C + N_G}{L} > 0.5 \\ 1 & \text{otherwise} \end{cases}$$

where L is the size of the region encompassing 400bp upstream and 400bp downstream of the length of a motif, N_C, N_G and N_{CpG} are the total count of C, G and CpG dinucleotide respectively.

Sparse logistic regression training (MocapS)

We trained sparse logistic regression models to classify candidate motif sites as true (motif overlapping with ChIP-Seq) or false (motif not overlapping with ChIP-Seq peaks) binding sites using a range of motif-associated genomic features, including motif matching quality, chromatin accessibility, TF footprint, TSS proximity, evolutionary conservation, motif region mapability and sequence-associated GC/CpG features. The goal is to build an ensemble of TFCT-specific and yet generalizable predictive models that address the differences in TF binding preferences and facilitate cross-sample binding site prediction. The logistic regression model is specified as below:

$$p(y = 1 | x; \theta) = \sigma(\theta^T x) = \frac{1}{1 + e^{-(\theta^T x + c)}}$$

where x represents motif samples, $y \in [1, -1]$ corresponds to bound and unbound motif labels respectively; θ specifies the 30 motif-associated genomic features that are mostly highly correlated with the ChIP-Seq

feature in the tested motif set and cell type environment (after two rounds of correlation filtering); c specifies the intercept. Features, such as SNP density, TES proximity and exon, have relatively low correlation with ChIP-Seq signals, and a weak contribution to the overall prediction of TFBS, so we excluded them from the model *a priori* for all TFCTs.

To create generalizable models and ensure only a most relevant subset of the genomic features are included in the logistic regression model for each TFCT, we specified a sparsity constraint to minimize the below objective function with L1 regularization

$$\min_{\theta} \sum_{i=1}^{M} -log\, p(y^{(i)}|x^{(i)};\theta) + \lambda||\theta||_1$$

where M is the number of candidate motif sites used for model training; λ is the shrinkage parameter corresponding to the sparsity of the trained logistic regression model where with L1-regularization, all but the most relevant features are shrink to zero. As the number of true binding sites takes up only 0.01%-5% of the total number of candidate sites (imbalanced class labels), we weighed the true and false motif classes to create a weight-balanced training dataset. For each TFCT condition, we tuned the shrinkage parameter λ to select a λ that optimizes the AUPR to achieve more sensible site ranking (also, AUPR scores are less biased than AUROC scores in assessing classification performance in imbalanced datasets). To avoid overfitting and promote model generalization, λ parameter was further tuned such that the resulting classifier is sparser than an optimally performing classifier with near-optimum (within one standard error of the optimum) cross-validation performance. We adopted a final bootstrap aggregation step to reduce estimator bias and help achieve more robust classifications.

Implementation of cross-sample TFBS prediction (MocapX)

To extend the utility of the above sparse logistic regression classifiers ensemble to TFCT conditions where ChIP-Seq data is missing, we used robustly weighted least squares regression to derive a mapping vector β to assign new TFCT prediction problems to trained classifiers based on genomic feature distance. When minimizing the below error function

$$\min \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij}(\gamma log \frac{Y_{ij}+Y_{ji}}{2} - \sum_{k=1}^{p} X_{ij,k}\beta_k)^2 = \min \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij}(e_{ij})^2$$

where n is the number of TFCT conditions; p is the number of features; $X_{ij,k}$ is the binary probability that feature k from TFCT i and TFCT j are derived from the same distribution based on KS test for continuous features and Chi-square test for categorical features; Y_{ij} is the predictive performance (AUPR) of model

trained in TFCT i when applied to classify motifs in TFCT j. We used predictive performance of MocapG as a baseline and used adaptive grid search to assign γ to seperate sample pairs that cross-predict well from those that predicts poorly. For DNase-Seq sample mappings, for example, we set γ as follows

$$\gamma = \begin{cases} 1 & \frac{Y_{ij}+Y_{ji}}{2Y_{MocapG}} > 1.1 \\ 1.31 & \frac{Y_{ij}+Y_{ji}}{2Y_{MocapG}} \leq 1.1 \end{cases}$$

γ was re-optimized with the exclusion of footprint features, when deriving mappings between ATAC-Seq samples and DNase-Seq data trained models. w was fitted via iteratively re-weighted least squares regression with the Tukey's bisquare family psi functions as below (1, 3).

$$w(u) = \begin{cases} ((1-(\frac{u}{c})^2))^2 & |u| \leq c \\ 0 & |u| > c \end{cases}$$

where u is the regression residual.

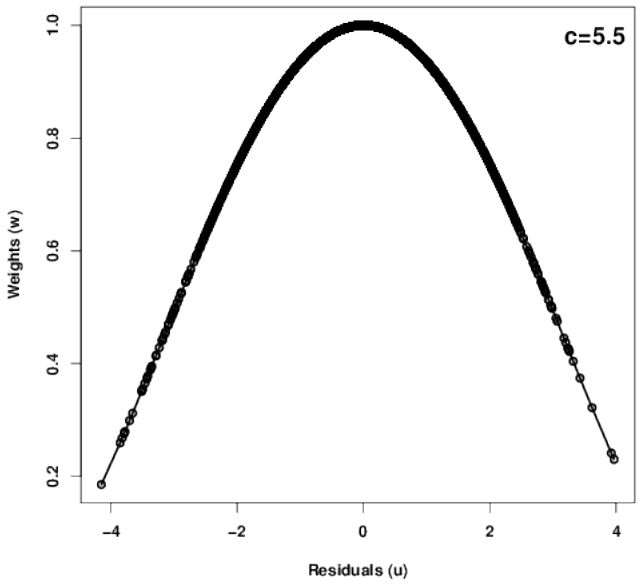


Figure 1: Tukey's bisquare function specifies sample weights based on regression residuals.

We selected the bisquare psi function hyperparameter c based on cross-validation sample-mapping performance. Distances between feature vectors were then used to assign weight w to a hold-out TFCT condition based on the weights of its nearest neighbors. We used $\hat{w}$ and $\hat{\beta}$ to compute a sample mapping $i \rightarrow j$ that max-

imizes $\hat{Y} = \arg\max_{i \to j} \hat{w}_{ij} \sum_{k=1}^{p} X_{ij,k} \hat{\beta}_k$, but to avoid mappings that do not significantly improve performance as compared to the generic method MocapG, we assign the mapping if and only if the estimated cross-prediction performance $\hat{Y}$ compares favorably with Y_{MocapG} in TFCT i as below

$$Y^{(j)}_{MocapX} = \begin{cases} Y^{(j)}_{MocapG} & \hat{Y} < Y^{(i)}_{MocapG} \\ Y^{(j)}_{MocapS_i} & \hat{Y} \geq Y^{(i)}_{MocapG} \end{cases}$$

References

1. Robert Andersen. *Modern methods for robust regression*. Number 152. Sage, 2008.

2. M Gardiner-Garden and M Frommer. Cpg islands in vertebrate genomes. *Journal of molecular biology*, 196(2):261–282, 1987.

3. Peter J Huber. Wiley series in probability and mathematics statistics. *Robust Statistics*, pages 309–312, 1981.

4. Jason Piper, Markus C Elze, Pierre Cauchy, Peter N Cockerill, Constanze Bonifer, and Sascha Ott. Wellington: a novel method for the accurate identification of digital genomic footprints from dnase-seq data. *Nucleic acids research*, page gkt850, 2013.

5. Serge Saxonov, Paul Berg, and Douglas L Brutlag. A genome-wide analysis of cpg dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5):1412–1417, 2006.

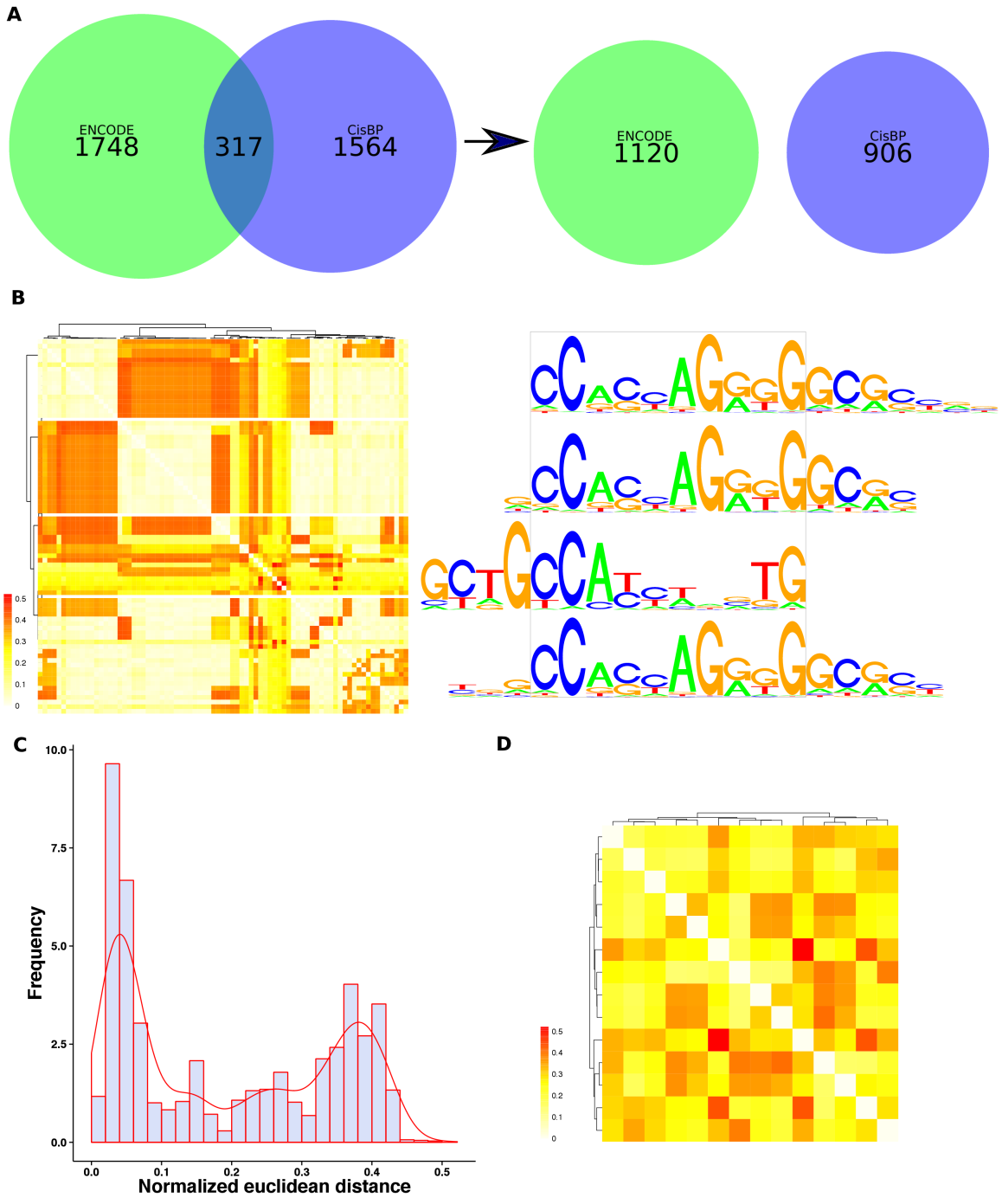
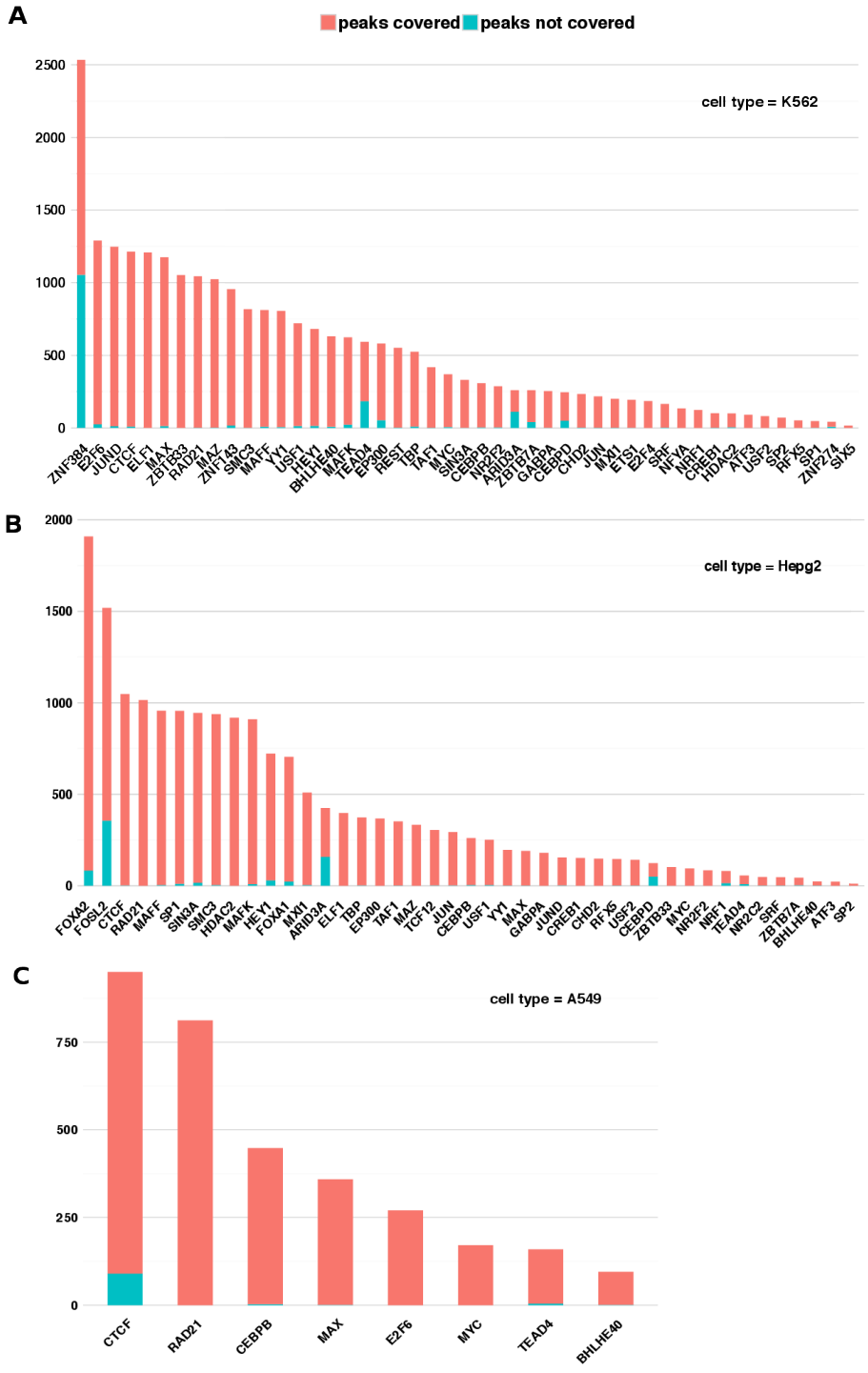


Figure S1: Filtering for highly redundant motifs representing the same TF. (**A**) Venn diagram showing size and overlap of the two motif collection, ENCODE (green) and CisBP (blue) motif set before and after redundancy filtering. ENCODE motif collection harbors motifs predominantly derived from ENCODE ChIP-Seq experiments, whereas CisBP motif collection includes motifs derived from large-scale PBM and SELEX studies. The overlap between the two motif collections are motifs drawn from databases such as TRANSFAC and JASPAR. (**B**) Similarity heatmap showing 79 motifs representing the TF, CTCF. White: highly similar motifs; Red: dissimilar motifs. Motifs are clustered based on normalized Euclidean distances between vectorized motifs and separated by row into four major groups. Representative motif logos for each of the four groups are drawn on the right. Cluster 1, 2, 4 represent motifs containing the canonical binding

site, with additional degenerate sites on the right, left and both side of the core motif respectively. Cluster 3 contains motifs that are more diverged from the canonical motif reflecting the multivalency in CTCF binding. (**C**) Histogram showing the distribution of between motif distances for CTCF. A conservative cutoff was selected at 0.14. (**D**) Similarity heatmap after redundancy filtering leaving 14 motifs that are more dissimilar from each other.



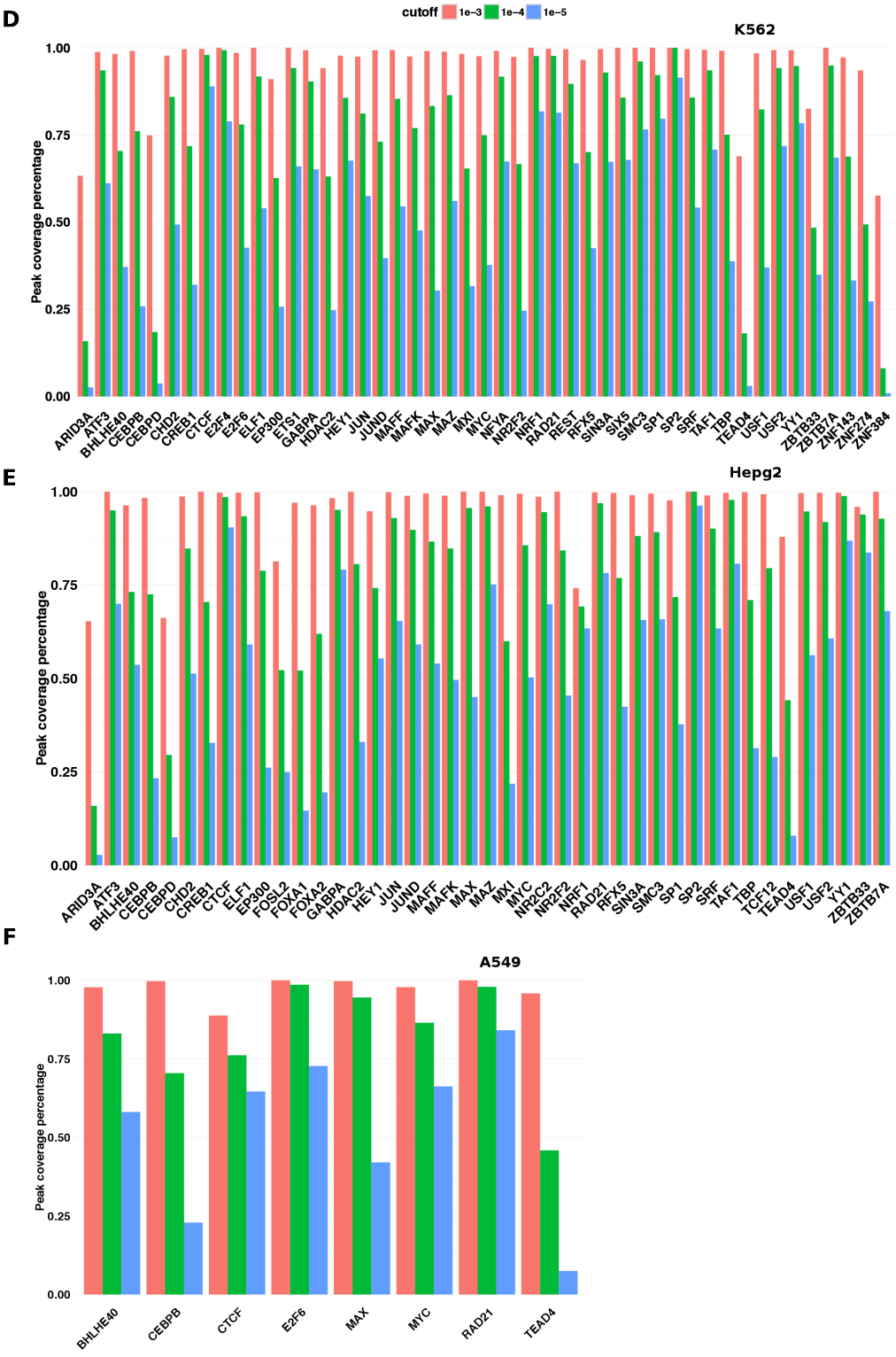


Figure S2: ChIP-Seq peak coverage by motifs. (**A-C**) Bar plot showing the number of ChIP-Seq peaks covered (red) and not covered (blue) by our analyses on testing chromosome 15 in cell type Hepg2, K562 and A549. (**D-F**) Grouped barplot showing the percentage of ChIP-Seq peak coverage using different FIMO motif scan cutoffs in the three cell types on chromosome 3.

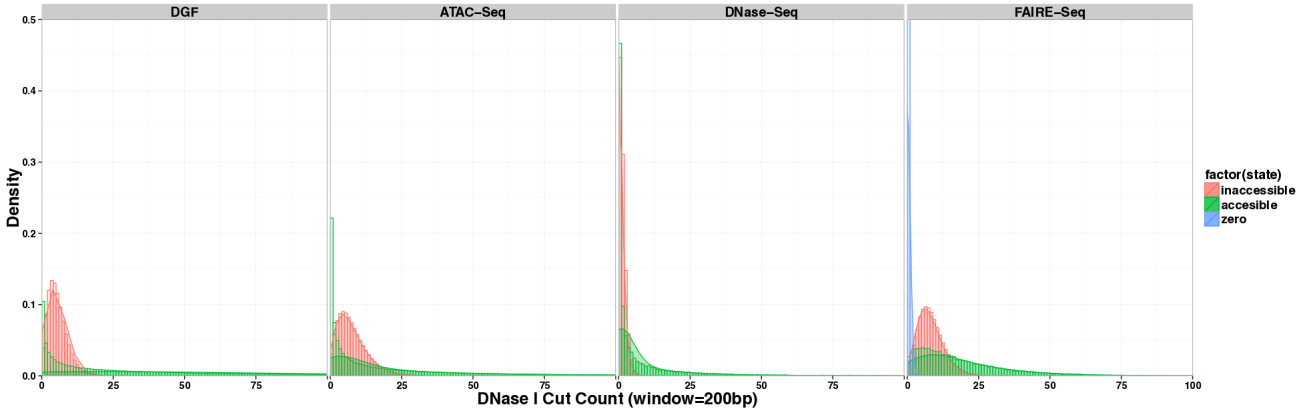


Figure S3: Cut count density distribution across experiment types. Data are simulated using zero-inflated negative binomial mixture model parameters derived from different experiment types. Zero component is insignificant in all but FAIRE-Seq. Red: inaccessible component; Green: accessible component; Blue; zero component.

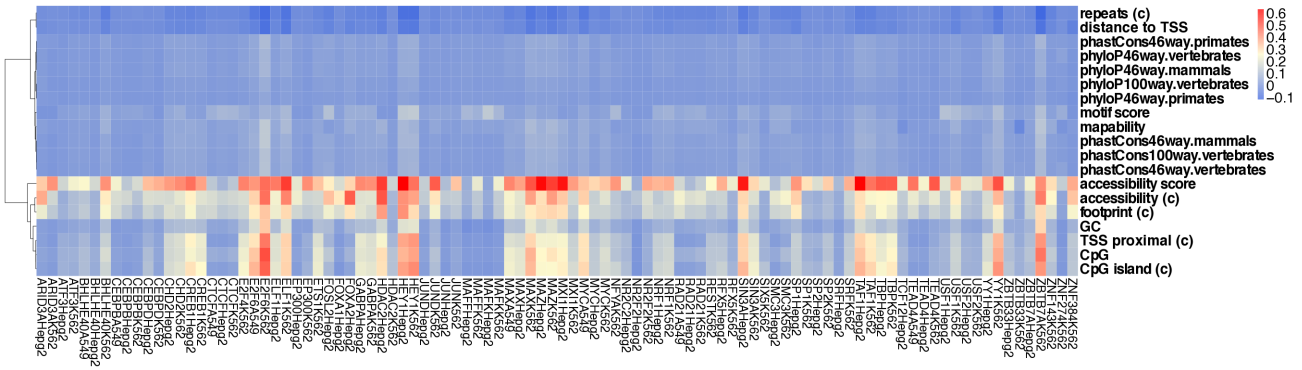
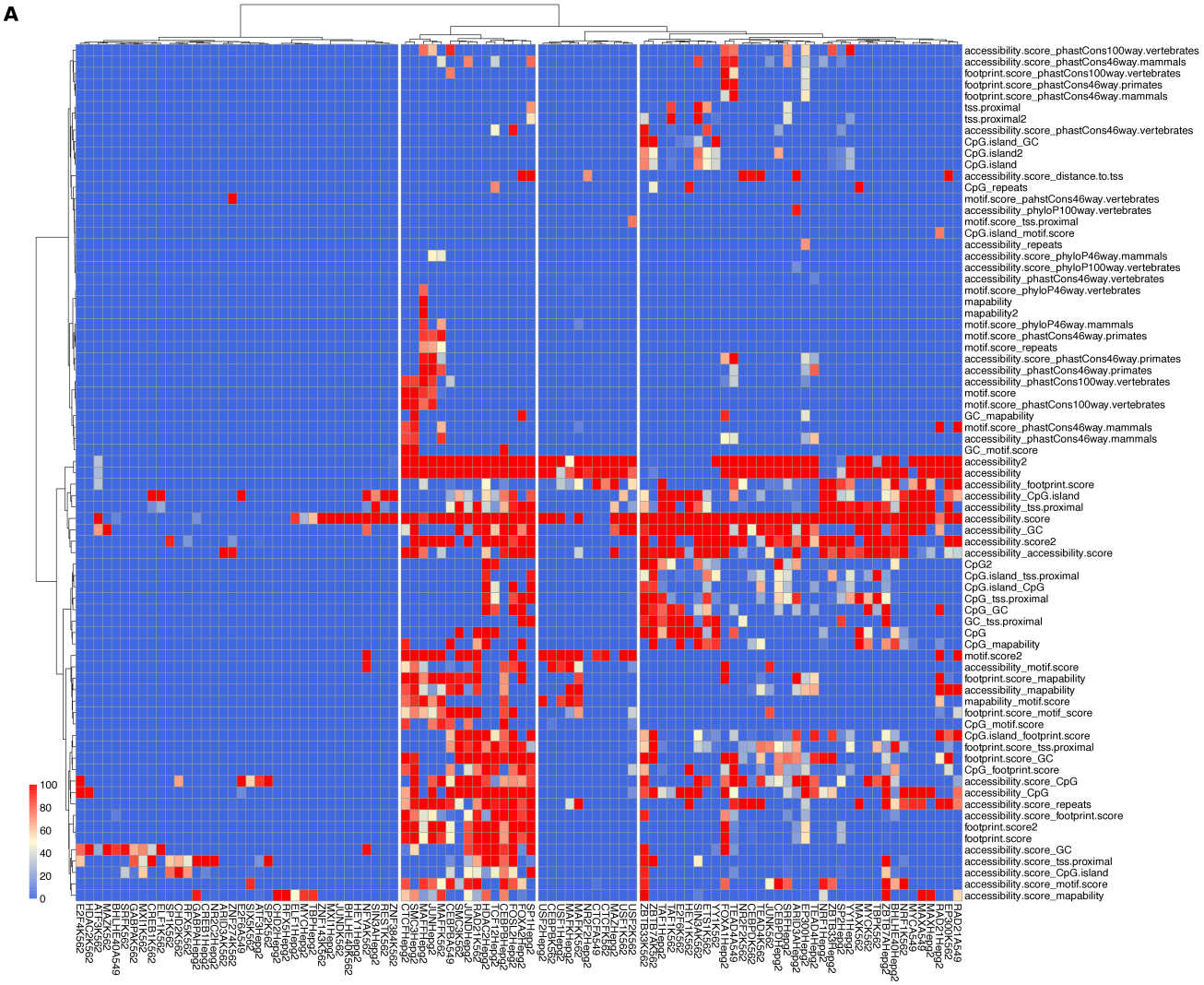


Figure S4: Feature correlations with ChIP-Seq signals. Heatmap showing the PCC between 19 genomic features and ChIP-Seq signals. Red indicates high positive correlation between a genomic feature and its associated ChIP-Seq signal; Blue indicates low positive or negative correlation. Genomic features are clustered based on their correlation value across 98 TF-cell type samples. Categorical features are marked with (c).


A
accessibility.score_phastCons100way.vertebrates
accessibility.score_phastCons46way.mammals
footprint.score_phastCons100way.vertebrates
footprint.score_phastCons46way.primates
footprint.score_phastCons46way.mammals
tss.proximal
tss.proximal2
accessibility.score_phastCons46way.vertebrates
CpG.island_GC
CpG.island2
CpG.island
accessibility.score_distance.to.tss
CpG_repeats
motif.score_pahstCons46way.vertebrates
accessibility_phyloP100way.vertebrates
motif.score_tss.proximal
CpG.island_motif.score
accessibility_repeats
accessibility.score_phyloP46way.mammals
accessibility.score_phyloP100way.vertebrates
accessibility_phastCons46way.vertebrates
motif.score_phyloP46way.vertebrates
mapability
mapability2
motif.score_phyloP46way.mammals
motif.score_phastCons46way.primates
motif.score_repeats
accessibility.score_phastCons46way.primates
accessibility_phastCons46way.primates
accessibility_phastCons100way.vertebrates
motif.score
motif.score_phastCons100way.vertebrates
GC_mapability
motif.score_phastCons46way.mammals
accessibility_phastCons46way.mammals
GC_motif.score
accessibility2
accessibility
accessibility_footprint.score
accessibility_CpG.island
accessibility_tss.proximal
accessibility.score
accessibility_GC
accessibility.score2
accessibility_accessibility.score
CpG2
CpG.island_tss.proximal
CpG.island_CpG
CpG_tss.proximal
CpG_GC
GC_tss.proximal
CpG
CpG_mapability
motif.score2
accessibility_motif.score
footprint.score_mapability
accessibility_mapability
mapability_motif.score
footprint.score_motif_score
CpG_motif.score
CpG.island_footprint.score
footprint.score_tss.proximal
footprint.score_GC
CpG_footprint.score
accessibility.score_CpG
accessibility_CpG
accessibility.score_repeats
accessibility.score_footprint.score
footprint.score2
footprint.score
accessibility.score_GC
accessibility.score_tss.proximal
accessibility.score_CpG.island
accessibility.score_motif.score
accessibility.score_mapability
100
80
60
40
20
0
RAD21A549
EP300K562
MAX549
MAXA549
MAXHepg2
NFYAK562
ZBTB7AHepg2
BHLHE40Hepg2
ZBTB7AK562
MAXK562
YY1K562
SP2Hepg2
ZBTB33Hepg2
TEAD4Hepg2
NFYAA549
EP300Hepg2
ARID3AHepg2
CEBPBHepg2
JUNK562
NR2F2K562
FOXA1Hepg2
YY1Hepg2
SIN3AK562
HEY1K562
EZH2K562
TAF1Hepg2
ZBTB7AK562
USF2K562
USF1K562
MAZHepg2
CTCFAA549
NR2F2Hepg2
MAFKHepg2
USF1Hepg2
CEBPAK562
SP1Hepg2
RFX5K562
TCF12Hepg2
HDAC2Hepg2
JUND1Hepg2
SMC3K562
MAFKK562
JUNHepg2
CTCFHepg2
SMC3Hepg2
ZNF384K562
RFX5AH8pg2
NFYAK562
BHLHE40K562
JUNDK562
MXI1Hepg2
ZKSCAN1K562
TBPHepg2
MTCHHepg2
ELF1Hepg2
CHD2Hepg2
RFX5Hepg2
ATF2K562
ZKSCAN1Hepg2
SIX5K562
NR2F2Hepg2
ARID3AK562
GABPAHepg2
RFX5K562
CHD2K562
SP1K562
ELF1K562
MXI1K562
GABPAK562
BHLHE40K562
MAZK562
ATF3K562
E2F4K562
RAD21A549

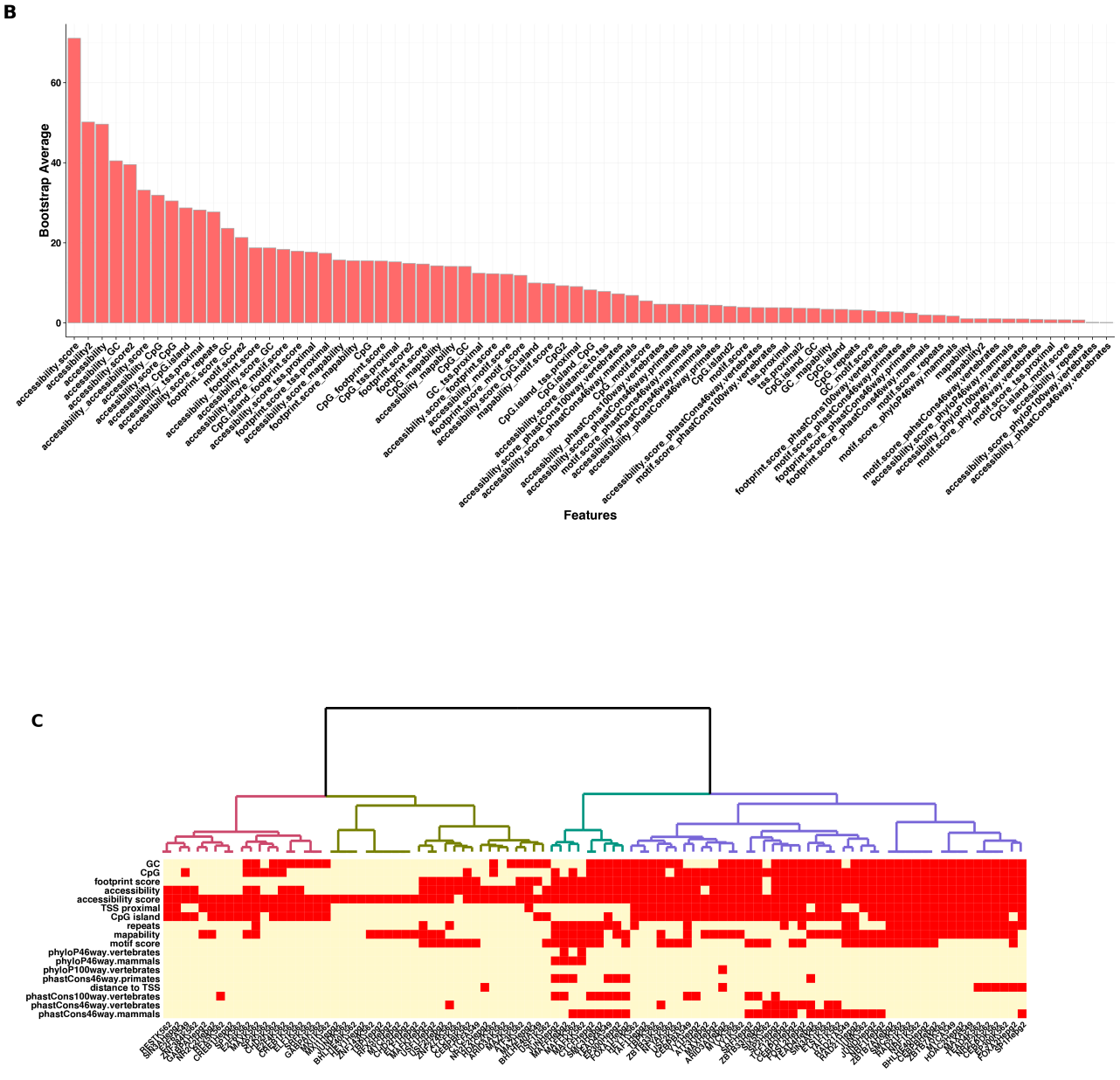


Figure S5: Training sparse logistic regression classifiers by bootstrap aggregation. (**A**) Heatmap showing the number of bootstraps each single or interaction feature was selected during 100 bootstrap aggregation. Red indicates more stable features that are likely to be selected by the majority of bootstraps, whereas blue indicates features less likely to be selected. Features are clustered based on the number of bootstraps that select them. (**B**) A barplot summary of the selected features ranked based on their bootstrap stability across samples. (**C**) Heatmap showing feature usage by different TFCT conditions. TFCT conditions are clustered into four major clusters based on feature usage preference. Group 1 (red) prefers models with local accessibility, TSS proximity and GC/CpG sequence features. Group 2 (khaki) predominantly selected a combination of accessibility and motif scores. Group 3 (green) and Group 4 (purple) both make use of a fuller range of features, with Group 4, additionally, preferring the binary predictor TSS proximity.

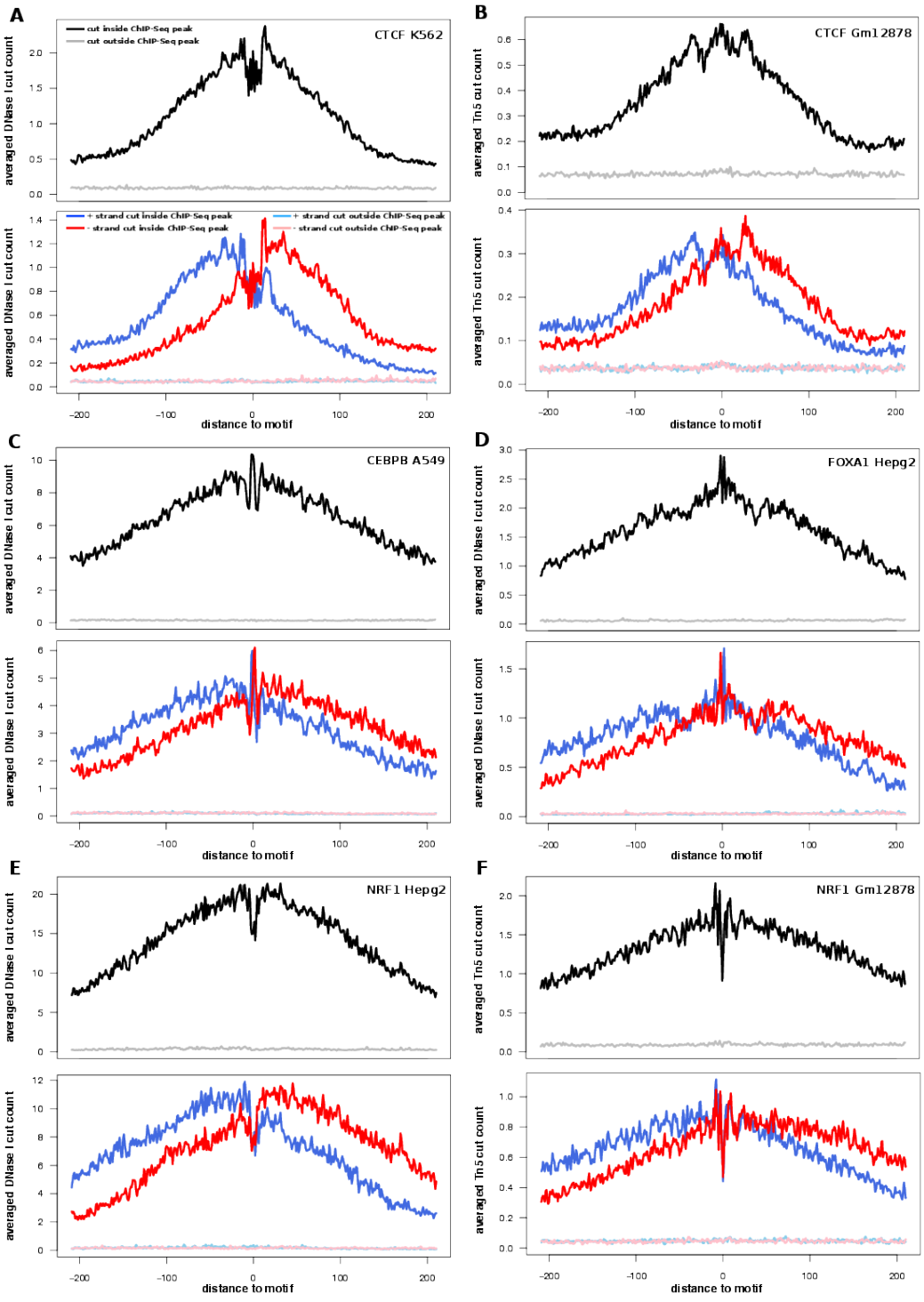


Figure S6: Footprint profiles vary across factors, cell types and experimental protocols. Black: cut count pattern around motifs overlapping ChIP-Seq peaks; grey: cut count pattern around motifs not overlapping ChIP-Seq peaks; blue/red: cut count pattern on the positive/negative strand around motifs overlapping ChIP-Seq peaks; light blue/pink: cut count pattern on the positive/negative strand around motifs not overlapping ChIP-Seq peaks.

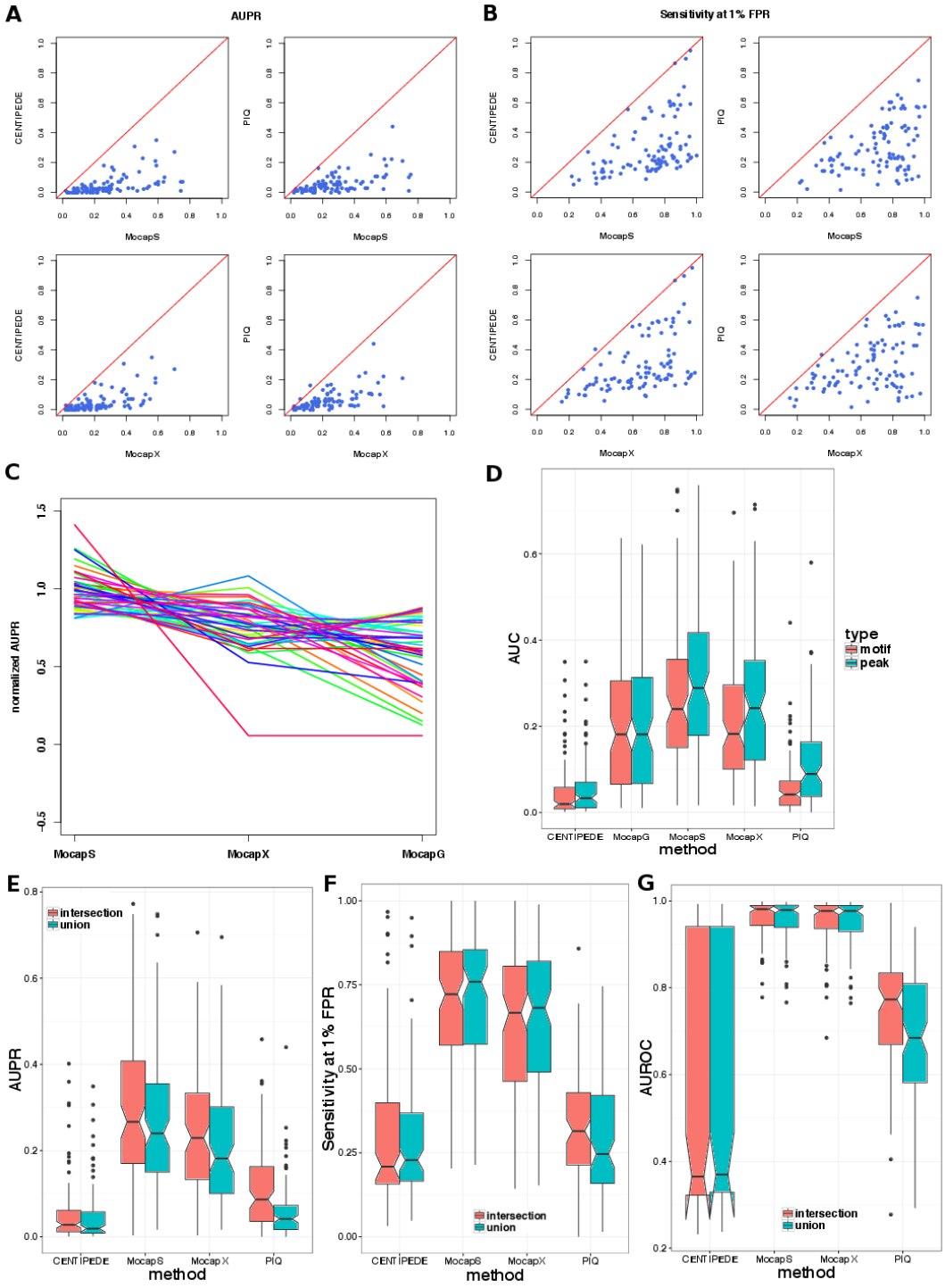


Figure S7: Performance comparison between methods. (**A-B**) Scatterplots showing pairwise comparisons between MocapS/MocapX and CENTIPEDE/PIQ in AUPRC and Sensitivity at 1% FPR. (**C**) Line plot contrasting differences in performance between MocapS, MocapX and MocapG for each of the 35 TF-cell type conditions with mapping to another TF-cell type trained sparse logistic regression model using MocapX. (**D**) Grouped boxplots showing similar results comparing the areas under precision over motif site recall curves (red) or precision over ChIP-Seq peak recall curves (green) across methods. (**E-G**) Grouped boxplots showing similar performance comparison results using the intersection (red) or union (green) of all method-predicted motif sets as the scope of comparison for AUPR, Sensitivity at 1% FPR and AUROC respectively.

Table S1: Performance comparison for DGF samples

TF-Cell type	Sample mapping	MocapG			MocapS			MocapX			CENTIPEDE			PIQ		
		AUPR	AUROC	Sensitivity*	AUPR	AUROC	Sensitivity	AUPR	AUROC	Sensitivity	AUPR	AUROC	Sensitivity	AUPR	AUROC	Sensitivity
ARID3AHepg2	default	0.25	0.93	0.84	0.29	0.96	0.82	0.25	0.93	0.84	0.04	0.38	0.32	0.01	0.54	0.08
ARID3AK562	default	0.25	0.97	0.83	0.25	0.97	0.83	0.24	0.96	0.83	0.01	0.37	0.32	0	0.42	0.05
ATF3Hepg2	default	0.04	1	0.82	0.03	0.99	0.83	0.04	1	0.82	0	0.39	0.3	0.01	0.91	0.65
ATF3K562	default	0.07	0.99	0.74	0.06	0.99	0.74	0.07	0.99	0.74	0	0.24	0.15	0.03	0.8	0.39
BHLHE40A549	default	0.1	0.99	0.68	0.12	0.99	0.68	0.1	0.99	0.68	0	0.34	0.09	0.02	0.82	0.36
BHLHE40Hepg2	default	0.03	0.99	0.68	0.03	0.98	0.68	0.03	0.99	0.68	0	0.46	0.28	0.01	0.78	0.4
BHLHE40K562	default	0.3	0.94	0.58	0.3	0.94	0.58	0.3	0.94	0.58	0.03	0.34	0.18	0.1	0.77	0.27
CEBPBA549	default	0.13	0.93	0.59	0.21	0.94	0.61	0.13	0.93	0.59	0.01	0.34	0.15	0.01	0.62	0.15
CEBPBHepg2	JUNDHepg2	0.05	0.94	0.64	0.14	0.95	0.72	0.12	0.95	0.7	0.04	0.95	0.54	0	0.48	0.14
CEBPBK562	default	0.06	0.89	0.44	0.08	0.92	0.52	0.06	0.89	0.44	0	0.27	0.12	0.01	0.58	0.16
CEBPDHepg2	default	0.16	0.99	0.92	0.18	0.99	0.93	0.16	0.99	0.92	0.05	0.99	0.89	0.04	0.55	0.24
CEBPDK562	CEBPDHepg2	0.35	0.99	0.85	0.37	0.99	0.86	0.24	0.99	0.7	0.09	0.99	0.58	0.03	0.6	0.23
CHD2Hepg2	default	0.25	0.98	0.77	0.26	0.99	0.78	0.25	0.98	0.77	0.02	0.31	0.19	0.12	0.88	0.6
CHD2K562	default	0.26	0.95	0.64	0.3	0.95	0.69	0.26	0.95	0.64	0.02	0.34	0.14	0.07	0.8	0.32
CREB1Hepg2	default	0.31	0.99	0.81	0.32	0.98	0.83	0.31	0.99	0.81	0.15	0.99	0.6	0.03	0.48	0.16
CREB1K562	default	0.21	0.98	0.68	0.29	0.99	0.85	0.21	0.98	0.68	0.02	0.33	0.24	0.03	0.48	0.17
CTCFA549	RAD21A549	0.08	0.99	0.56	0.17	0.98	0.41	0.17	0.99	0.68	0	0.26	0.09	0.07	0.8	0.37
CTCFHepg2	CTCFK562	0.04	0.91	0.19	0.23	0.94	0.41	0.16	0.87	0.31	0.03	0.92	0.16	0.03	0.76	0.23
CTCFK562	CTCFHepg2	0.06	0.95	0.33	0.22	0.94	0.37	0.21	0.96	0.42	0	0.24	0.06	0.06	0.78	0.33
E2F4K562	SIN3AK562	0.31	0.99	0.86	0.51	1	0.96	0.39	1	0.96	0.02	0.32	0.22	0.07	0.88	0.57
E2F6A549	default	0.3	0.98	0.54	0.44	0.99	0.71	0.3	0.98	0.54	0.03	0.34	0.14	0.06	0.6	0.18
E2F6K562	default	0.44	0.91	0.44	0.75	0.98	0.76	0.44	0.91	0.44	0.07	0.32	0.16	0.12	0.6	0.17
ELF1Hepg2	ELF1K562	0.29	0.98	0.74	0.3	0.98	0.74	0.25	0.98	0.63	0.02	0.32	0.24	0.17	0.88	0.56
ELF1K562	default	0.55	0.94	0.59	0.6	0.94	0.64	0.55	0.94	0.59	0.07	0.33	0.2	0.18	0.8	0.28
EP300Hepg2	default	0.1	0.99	0.76	0.36	0.98	0.9	0.1	0.99	0.76	0.07	0.99	0.59	0.01	0.58	0.15
EP300K562	default	0.26	0.94	0.78	0.35	0.98	0.79	0.26	0.94	0.78	0.07	0.94	0.51	0.02	0.66	0.22
ETS1K562	YY1K562	0.17	0.99	0.76	0.27	1	0.97	0.27	1	0.97	0.01	0.33	0.23	0.05	0.83	0.51
FOSL2Hepg2	default	0.2	0.86	0.29	0.26	0.91	0.31	0.2	0.86	0.29	0.18	0.89	0.27	0.13	0.59	0.18
FOXA1Hepg2	default	0.14	0.93	0.64	0.3	0.95	0.7	0.14	0.93	0.64	0.1	0.95	0.57	0	0.5	0.05
FOXA2Hepg2	default	0.38	0.84	0.6	0.45	0.91	0.58	0.38	0.84	0.6	0.31	0.9	0.56	0.07	0.65	0.16
GABPAHepg2	MYCHepg2	0.25	0.99	0.84	0.28	0.99	0.85	0.27	0.99	0.84	0.02	0.32	0.23	0.14	0.85	0.56
GABPAK562	MYCK562	0.28	0.98	0.77	0.3	0.97	0.8	0.29	0.99	0.1	0.02	0.33	0.2	0.05	0.81	0.25
HDAC2Hepg2	default	0.56	0.99	0.86	0.59	0.98	0.86	0.56	0.99	0.86	0.35	0.99	0.86	0.1	0.51	0.22
HDAC2K562	default	0.04	0.98	0.68	0.07	1	0.93	0.04	0.98	0.68	0.01	0.99	0.32	0.01	0.58	0.35
HEY1Hepg2	HEY1K562	0.64	0.97	0.79	0.64	0.97	0.79	0.52	0.98	0.63	0.1	0.38	0.29	0.44	0.85	0.6
HEY1K562	HEY1Hepg2	0.43	0.94	0.52	0.63	0.97	0.75	0.43	0.94	0.52	0.07	0.38	0.19	0.22	0.81	0.38
JUNDHepg2	default	0.04	0.98	0.72	0.15	0.98	0.84	0.04	0.98	0.72	0.02	0.98	0.58	0.01	0.69	0.24
JUNDK562	default	0.41	0.97	0.7	0.41	0.97	0.7	0.41	0.97	0.7	0.03	0.33	0.24	0.08	0.66	0.26
JUNHepg2	JUNDHepg2	0.04	0.8	0.44	0.09	0.88	0.48	0.1	0.82	0.46	0.03	0.8	0.38	0.01	0.64	0.18
JUNK562	CEBPBK562	0.18	0.99	0.83	0.22	0.99	0.86	0.21	0.99	0.82	0.01	0.3	0.23	0.03	0.76	0.43
MAFFHepg2	MAFFK562	0.01	0.55	0.11	0.08	0.77	0.21	0.05	0.76	0.19	0.01	0.55	0.1	0.01	0.59	0.06
MAFFK562	MAFFHepg2	0.08	0.71	0.29	0.15	0.85	0.37	0.11	0.85	0.3	0.01	0.35	0.09	0.03	0.61	0.14
MAFKHepg2	MAFKK562	0.01	0.52	0.09	0.1	0.8	0.26	0.05	0.8	0.18	0.01	0.57	0.08	0.01	0.6	0.02
MAFKK562	MAFKHepg2	0.09	0.74	0.33	0.15	0.88	0.42	0.08	0.85	0.26	0.01	0.35	0.1	0.02	0.64	0.14
MAXA549	MYCA549	0.32	0.99	0.78	0.34	0.99	0.74	0.24	0.99	0.57	0.02	0.34	0.14	0.07	0.84	0.36
MAXHepg2	default	0.23	0.99	0.84	0.23	0.99	0.81	0.23	0.99	0.84	0.02	0.33	0.28	0.11	0.88	0.57
MAXK562	ZBTB33K562	0.48	0.97	0.64	0.6	0.99	0.78	0.39	0.96	0.48	0.05	0.31	0.21	0.12	0.79	0.25
MAZHepg2	MAZK562	0.53	1	0.9	0.54	1	0.9	0.5	0.99	0.82	0.04	0.37	0.31	0.11	0.64	0.32
MAZK562	default	0.51	0.98	0.65	0.58	0.99	0.72	0.51	0.98	0.65	0.06	0.34	0.22	0.1	0.59	0.21
MXI1Hepg2	default	0.58	0.99	0.88	0.58	0.99	0.88	0.58	0.99	0.88	0.21	0.99	0.65	0.02	0.5	0.08
MXI1K562	default	0.26	0.99	0.82	0.28	0.99	0.87	0.26	0.99	0.82	0.07	0.99	0.39	0.01	0.51	0.15
MYCA549	default	0.22	0.97	0.45	0.25	0.97	0.47	0.22	0.97	0.45	0.03	0.33	0.14	0.07	0.81	0.26
MYCHepg2	default	0.1	0.99	0.78	0.1	0.99	0.78	0.1	0.99	0.78	0.01	0.4	0.3	0.04	0.88	0.55
MYCK562	default	0.18	0.98	0.68	0.22	0.99	0.77	0.18	0.98	0.68	0.01	0.3	0.17	0.05	0.84	0.33
NFYAK562	default	0.23	0.87	0.58	0.36	0.96	0.76	0.23	0.87	0.58	0.01	0.36	0.19	0.16	0.76	0.43
NR2C2Hepg2	SP2Hepg2	0.06	0.99	0.84	0.07	0.99	0.83	0.08	0.97	0.82	0.02	0.98	0.36	0.02	0.53	0.22
NR2F2Hepg2	default	0.02	0.95	0.36	0.07	0.94	0.61	0.02	0.95	0.36	0.03	0.97	0.4	0	0.65	0.12
NR2F2K562	NR2F2Hepg2	0.22	0.98	0.8	0.26	0.99	0.78	0.17	0.98	0.57	0.06	0.98	0.38	0.02	0.64	0.23
NRF1Hepg2	NRF1K562	0.13	0.99	0.83	0.15	0.98	0.88	0.16	0.98	0.81	0.01	0.32	0.21	0.05	0.85	0.56
NRF1K562	NRF1Hepg2	0.18	0.99	0.81	0.28	0.99	0.87	0.22	0.99	0.86	0.01	0.31	0.16	0.05	0.83	0.44
RAD21A549	CTCFA549	0.14	0.99	0.61	0.21	0.99	0.69	0.13	0.98	0.33	0.02	0.36	0.16	0.05	0.78	0.29
RAD21Hepg2	CTCFHepg2	0.06	0.92	0.22	0.12	0.91	0.35	0.16	0.93	0.37	0.05	0.91	0.19	0.02	0.73	0.21
RAD21K562	CTCFK562	0.08	0.95	0.31	0.15	0.96	0.46	0.14	0.93	0.29	0.01	0.3	0.1	0.03	0.79	0.24
RESTK562	SIN3AK562	0.18	0.85	0.52	0.2	0.86	0.57	0.17	0.92	0.57	0.02	0.38	0.18	0.05	0.75	0.26
RFX5Hepg2	default	0.17	0.91	0.7	0.18	0.9	0.68	0.17	0.91	0.7	0.04	0.94	0.49	0.05	0.55	0.2
RFX5K562	default	0.08	0.94	0.76	0.09	0.94	0.76	0.08	0.94	0.76	0	0.27	0.22	0.02	0.56	0.2
SIN3AHepg2	default	0.47	0.95	0.66	0.5	0.96	0.66	0.47	0.95	0.66	0.23	0.95	0.49	0.25	0.77	0.45
SIN3AK562	RESTK562	0.17	0.98	0.68	0.29	0.99	0.88	0.22	0.99	0.8	0.01	0.3	0.2	0.05	0.88	0.41
SIX5K562	default	0.06	1	0.86	0.29	0.97	0.95	0.06	1	0.86	0	0.98	0.37	0.07	0.6	0.38
SMC3Hepg2	SMC3K562	0.06	0.91	0.13	0.19	0.93	0.36	0.08	0.89	0.19	0.02	0.36	0.13	0.03	0.68	0.14
SMC3K562	SMC3Hepg2	0.11	0.96	0.25	0.19	0.97	0.4	0.16	0.93	0.3	0.02	0.32	0.13	0.04	0.71	0.21
SP1Hepg2	default	0.24	0.92	0.4	0.31	0.93	0.5	0.24	0.92	0.4	0.17	0.93	0.3	0.06	0.58	0.16
SP1K562	default	0.13	1	0.95	0.15	1	0.94	0.13	1	0.95	0	0.34	0.21	0.01	0.76	0.42
SP2Hepg2	YY1Hepg2	0.02	1	1	0.03	1	1	0.02	1	0.99	0	0.99	0.24	0.02	0.71	0.57
SP2K562	ETS1K562	0.13	0.98	0.9	0.19	0.96	0.9	0.14	1	0.93	0	0.31	0.17	0.04	0.73	0.43
SRFHepg2	default	0.02	0.97	0.68	0.04	0.96	0.76	0.02	0.97	0.68	0	0.37	0.21	0.01	0.82	0.43
SRFK562	SRFHepg2	0.18	0.98	0.75	0.19	0.98	0.76	0.13	0.97	0.7	0.02	0.4	0.27	0.05	0.85	0.4
TAF1Hepg2	TAF1K562	0.55	0.99	0.96	0.59	1	0.96	0.58	1	0.95	0.18	0.99	0.59	0.22	0.94	0.74
TAF1K562	SIN3AK562	0.34	0.99	0.82	0.53	1	0.95	0.49	1	0.94	0.03	0.34	0.25	0.11	0.89	0.57
TBPHepg2	default	0.37	0.99	0.92	0.38	0.99	0.92	0.37	0.99	0.92	0.12	0.99	0.7	0.03	0.29	0.08
TBPK562	default	0.38	0.98	0.84	0.48	0.98	0.88	0.38	0.98	0.84	0.14	0.98	0.55	0.03	0.36	0.12
TCF12Hepg2	default	0.05	0.91	0.37	0.12	0.91	0.41	0.05	0.91	0.37	0.03	0.92	0.21	0.01	0.58	0.13
TEAD4A549	TEAD4Hepg2	0.36	0.99	0.92	0.47	0.99	0.9	0.42	0.99	0.8	0.01	0.36	0.28	0.03	0.63	0.17
TEAD4Hepg2	default	0.08	0.99	0.97	0.15	0.99	0.96	0.08	0.99	0.97	0.02	0.99	0.95	0	0.53	0.1
TEAD4K562	default	0.4	0.97	0.85	0.4	0.97	0.83	0.4	0.97	0.85	0.1	0.97	0.61	0.02	0.52	0.15
USF1Hepg2	USF1K562	0.06	0.85	0.36	0.17	0.92	0.54	0.12	0.93	0.35	0.01	0.39	0.16	0.16	0.68	0.37
USF1K562	default	0.23	0.85	0.4	0.27	0.91	0.42	0.23	0.85	0.4	0.03	0.35	0.13	0.1	0.58	0.18
USF2Hepg2	USF2K562	0.06	0.88	0.53	0.15	0.91	0.6	0.15	0.92	0.55	0.03	0.88	0.35	0.05	0.7	0.32
USF2K562	USF2Hepg2	0.06	0.98	0.6	0.21	0.98	0.72	0.17	0.97	0.66	0	0.26	0.15	0.04	0.67	0.4
YY1Hepg2	YY1K562	0.19	0.99	0.87	0.2	0.99	0.85	0.15	0.99	0.76	0.07	0.99	0.59	0.08	0.84	0.62
YY1K562	ETS1K562	0.45	0.97	0.65	0.7	0.99	0.86	0.7	0.99	0.86	0.27	0.98	0.5	0.21	0.86	0.58
ZBTB33Hepg2	default	0.08	0.99	0.73	0.13	0.99	0.79	0.08	0.99	0.73	0.01	0.33	0.24	0.03	0.84	0.46
ZBTB33K562	MAXK562	0.03	0.83	0.19	0.05	0.89	0.23	0.03	0.8	0.15	0.01	0.6	0.05	0.01	0.61	0.07
ZBTB7AHepg2	default	0.07	0.99	0.75	0.07	0.99	0.74	0.07	0.99	0.75	0	0.3	0.21	0.02	0.81	0.43
ZBTB7AK562	ZBTB7AHepg2	0.41	0.93	0.46	0.74	0.99	0.81	0.41	0.98	0.51	0.07	0.32	0.19	0.1	0.7	0.22
ZNF143K562	default	0.18	0.95	0.43	0.18	0.95	0.43	0.18	0.95	0.43	0.03	0.35	0.19	0.05	0.67	0.21
ZNF274K562	default	0.03	0.78	0.54	0.02	0.81	0.47	0.03	0.78	0.54	0.01	0.81	0.32	0	0.57	0.01
ZNF384K562	default	0.32	0.87	0.46	0.31	0.87	0.46	0.32	0.87	0.46	0.05	0.35	0.2	0.05	0.5	0.11

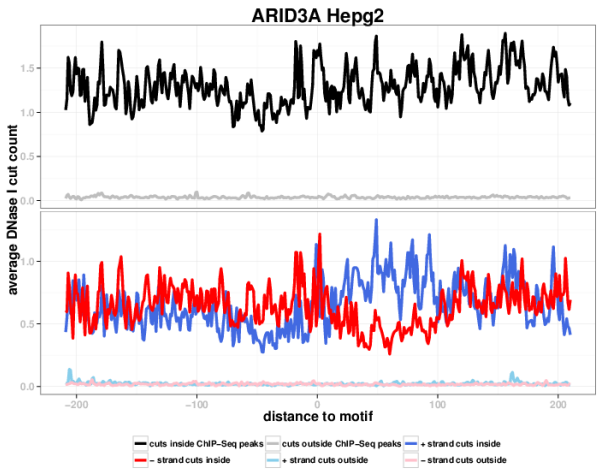
*Sensitivity stands for sensitivity at 1% false positive rate.

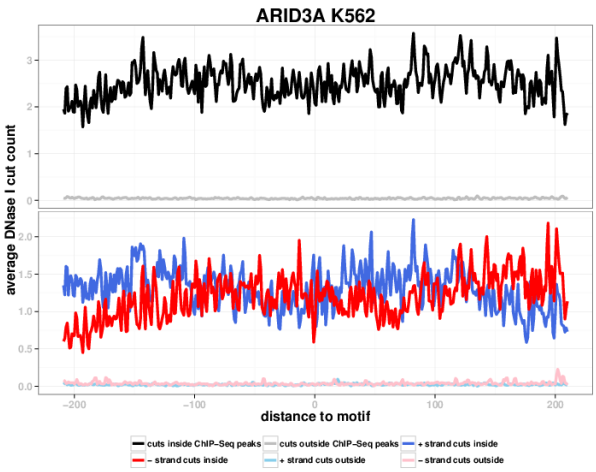
Table S2: Performance comparison for ATAC-Seq samples

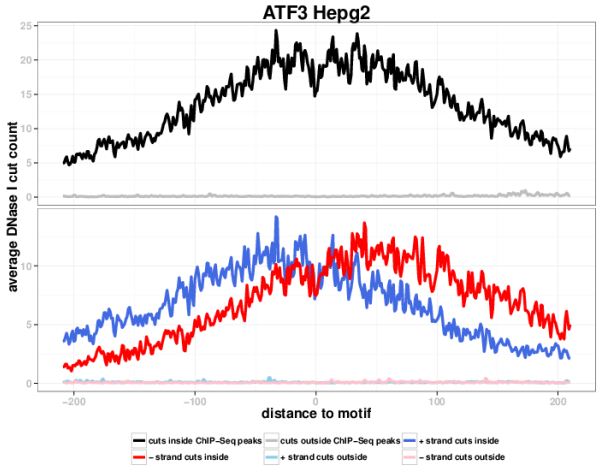
TF*	Sample mapping	MocapG			MocapX			CENTIPEDE		
		AUPR	AUROC	Sensitivity	AUPR	AUROC	Sensitivity	AUPR	AUROC	Sensitivity
BHLHE40	default	0.24	0.95	0.59	0.24	0.95	0.59	0.05	0.39	0.26
CREB1	default	0.08	0.91	0.54	0.08	0.91	0.54	0.03	0.4	0.26
CTCF	CTCFK562	0.05	0.81	0.31	0.12	0.80	0.32	0.01	0.44	0.12
E2F4	default	0.14	0.99	0.87	0.14	0.99	0.87	0.02	0.34	0.26
ELF1	NR2C2Hepg2	0.44	0.98	0.79	0.46	0.97	0.8	0.06	0.34	0.25
EP300	default	0.17	0.95	0.8	0.17	0.95	0.8	0.03	0.35	0.25
ETS1	default	0.15	0.98	0.82	0.15	0.98	0.82	0.02	0.37	0.29
GABPA	default	0.13	0.98	0.84	0.13	0.98	0.84	0.02	0.39	0.3
MAFK	MAFKK562	0.002	0.56	0.07	0.02	0.74	0.17	0.001	0.46	0.04
MAZ	MAZK562	0.52	0.97	0.76	0.51	0.97	0.77	0.1	0.37	0.26
MXI1	default	0.41	0.96	0.79	0.41	0.96	0.79	0.04	0.26	0.15
NFYA	default	0.32	0.98	0.9	0.32	0.98	0.9	0.02	0.39	0.3
NRF1	NRF1K562	0.25	0.99	0.8	0.31	0.99	0.84	0.03	0.35	0.25
RAD21	RAD21Hepg2	0.05	0.83	0.32	0.12	0.86	0.4	0.01	0.44	0.13
SIN3A	SIN3AK562	0.28	0.98	0.8	0.28	0.99	0.85	0.06	0.36	0.28
SIX5	default	0.06	0.98	0.93	0.06	0.98	0.93	0.004	0.34	0.25
SP1	default	0.36	0.96	0.81	0.36	0.96	0.81	0.04	0.37	0.28
SRF	SRFHepg2	0.19	0.95	0.76	0.17	0.93	0.72	0.03	0.4	0.3
TCF12	TCF12Hepg2	0.40	0.91	0.58	0.36	0.95	0.51	0.1	0.47	0.2
USF1	USF1Hepg2	0.07	0.88	0.5	0.18	0.92	0.59	0.01	0.37	0.16
YY1	YY1K562	0.3	0.94	0.73	0.28	0.95	0.71	0.05	0.38	0.25
ZBTB33	default	0.06	0.99	0.73	0.06	0.99	0.73	0.005	0.32	0.23
ZNF143	ZNF143K562	0.15	0.9	0.5	0.15	0.9	0.5	0.04	0.42	0.24
ZNF384	default	0.15	0.68	0.33	0.15	0.68	0.33	0.02	0.36	0.08

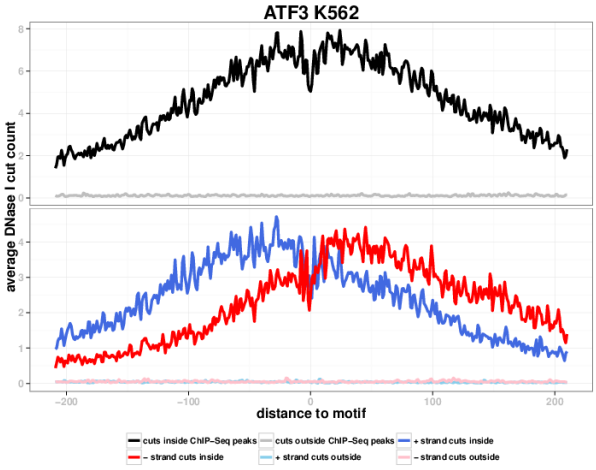
*Cell type: Gm12878.

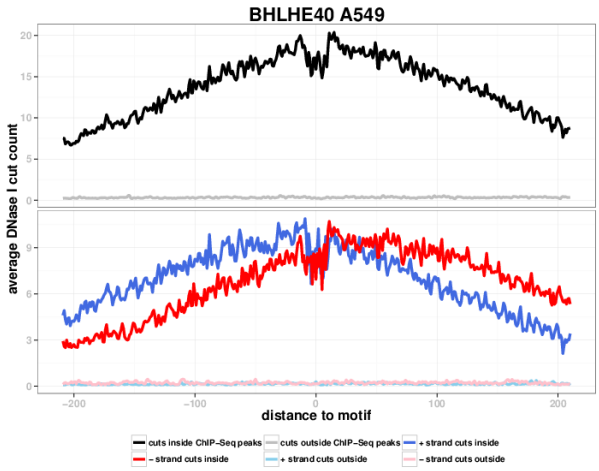
Appendix A Footprint profile plots

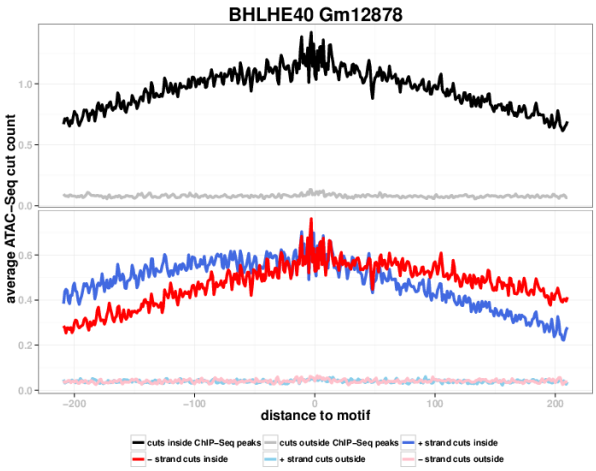


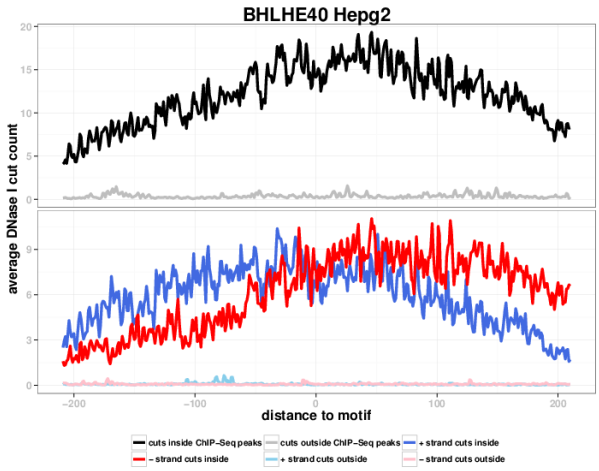


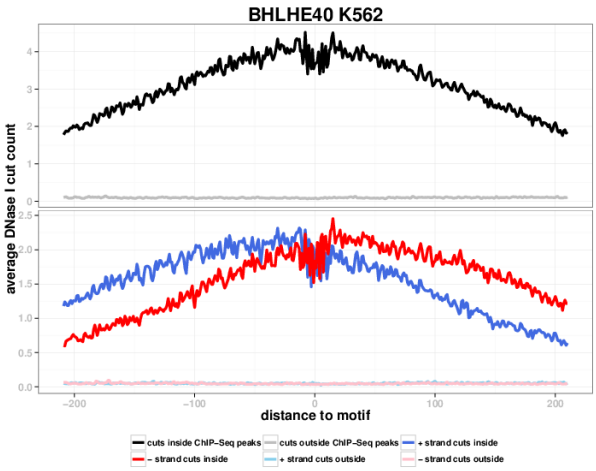


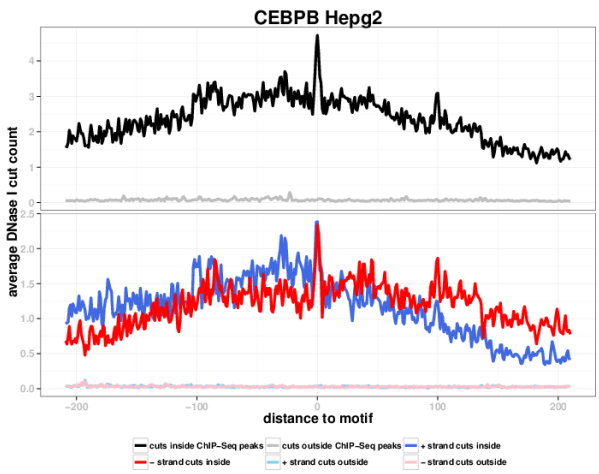


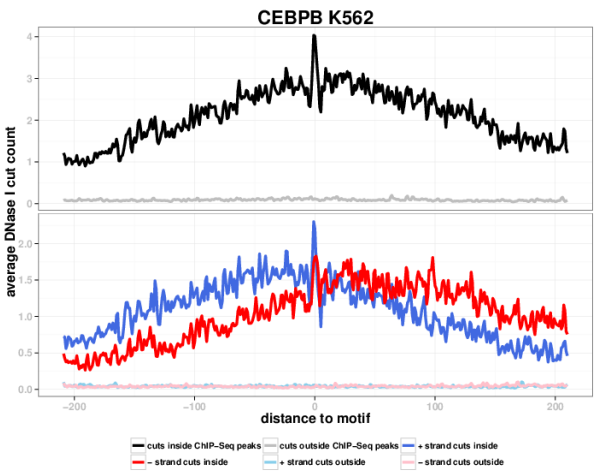

BHLHE40 A549
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

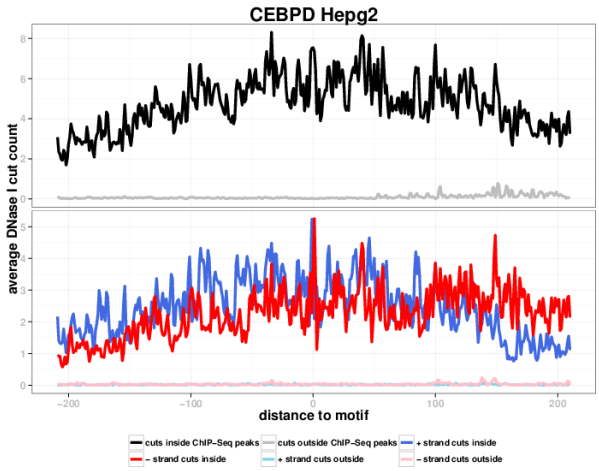

BHLHE40 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

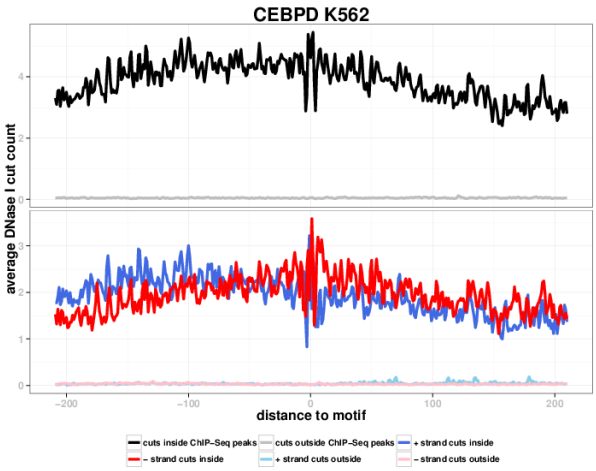

BHLHE40 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

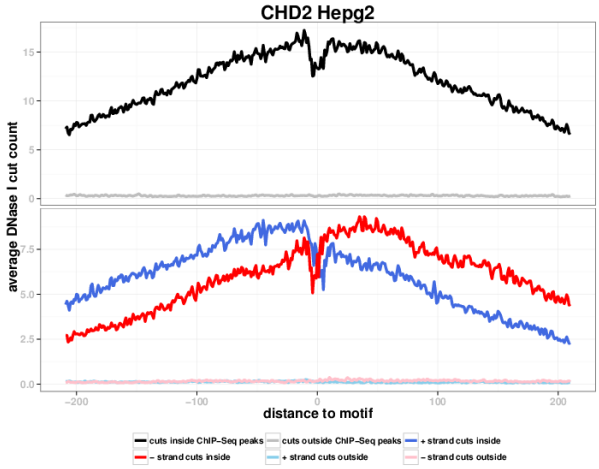

BHLHE40 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

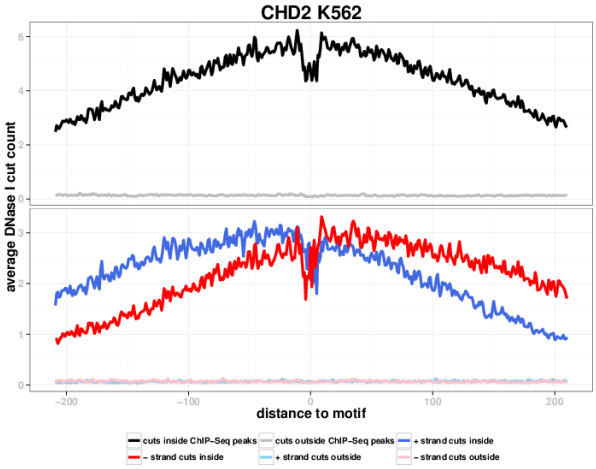

CEBPB Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

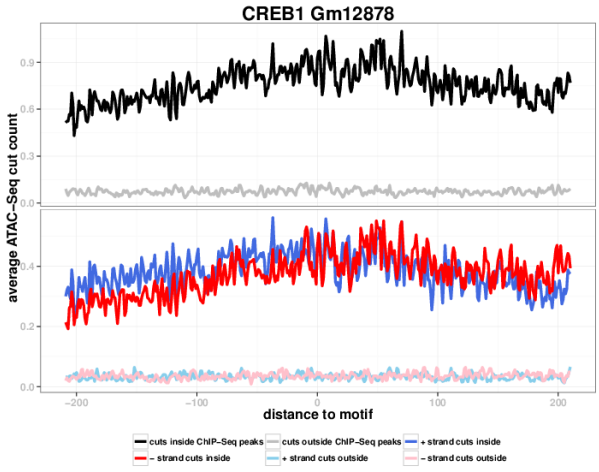

CEBPB K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

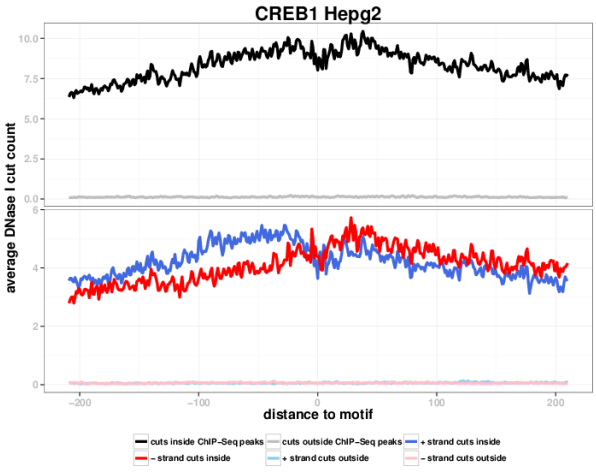

CEBPD Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

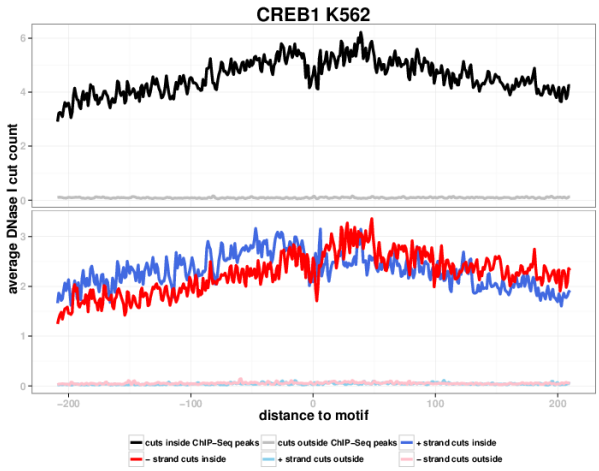

CEBPD K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

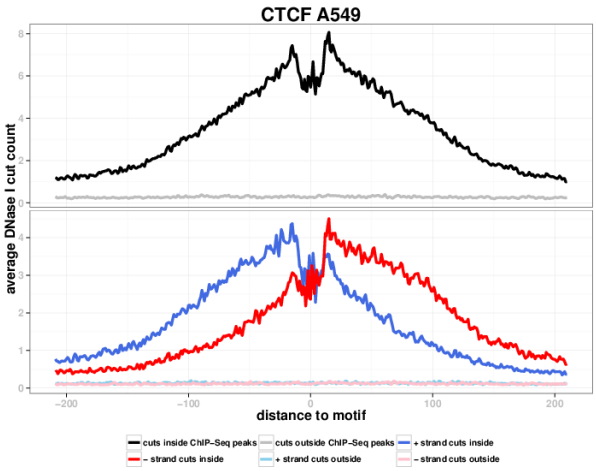

CHD2 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

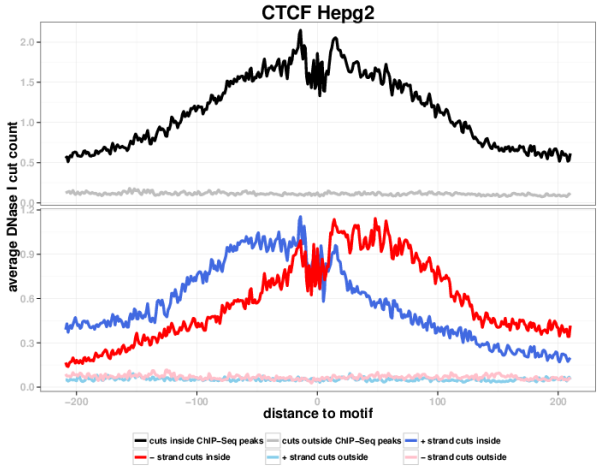

CHD2 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

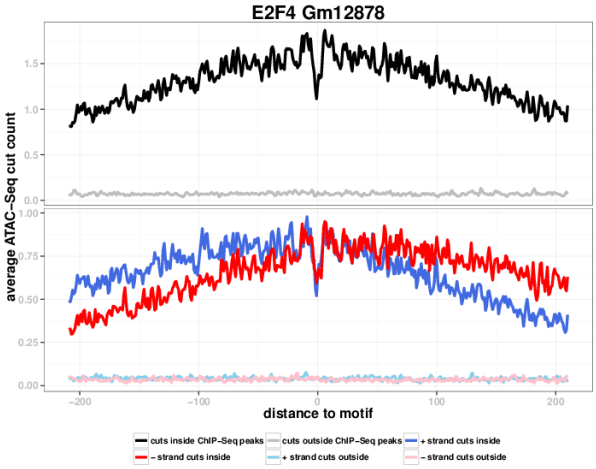

CREB1 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

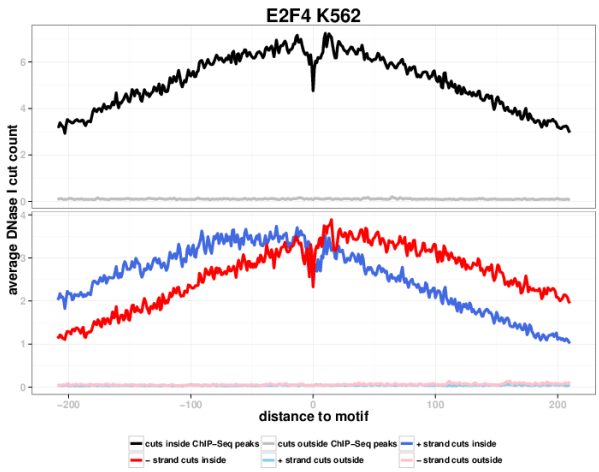

CREB1 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

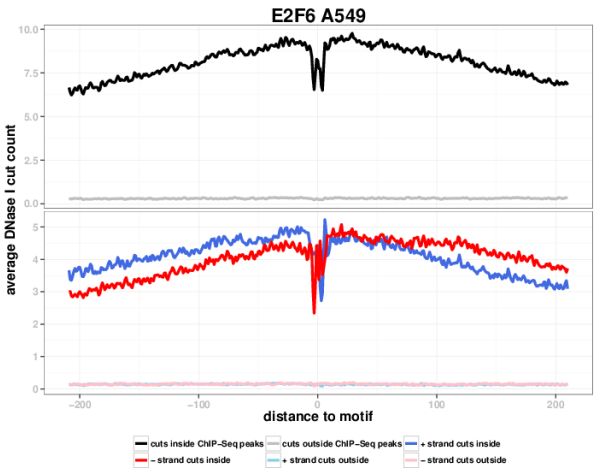

CREB1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

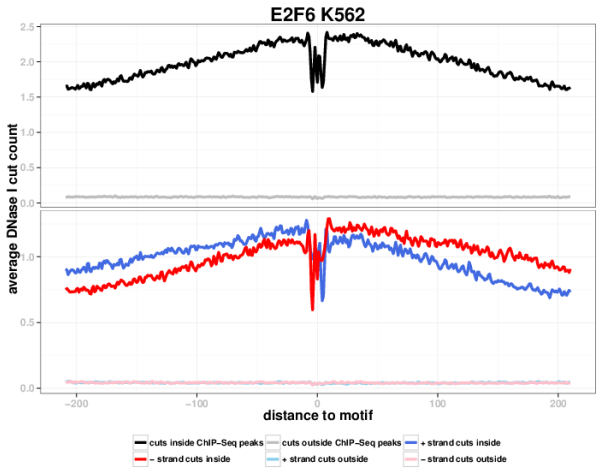

CTCF A549
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

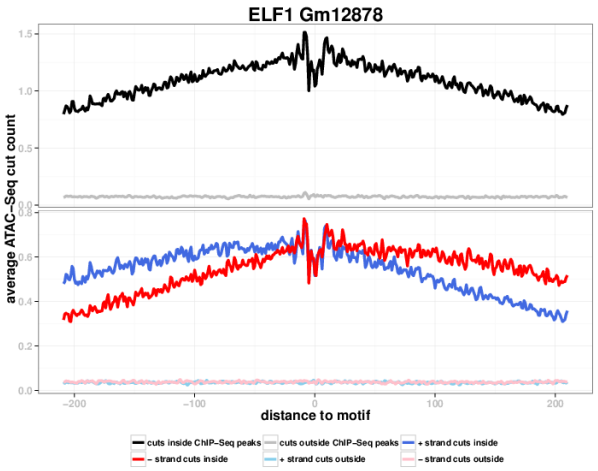

CTCF Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

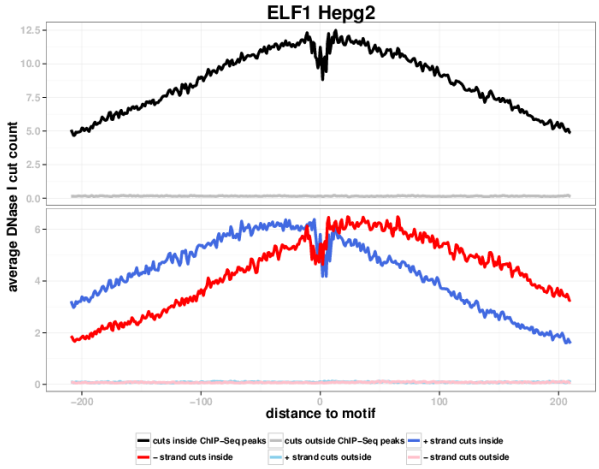

E2F4 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

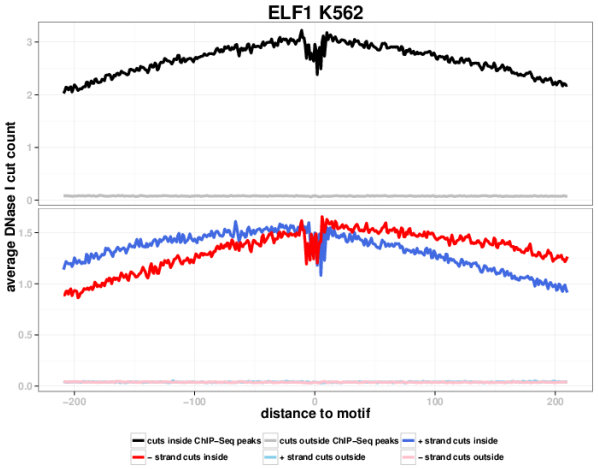

E2F4 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

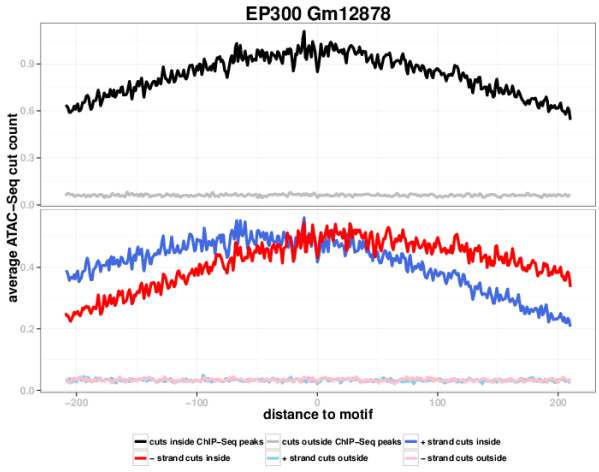

E2F6 A549
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

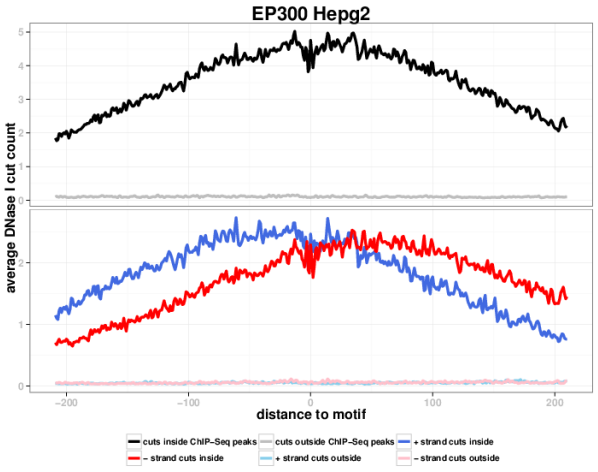

E2F6 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

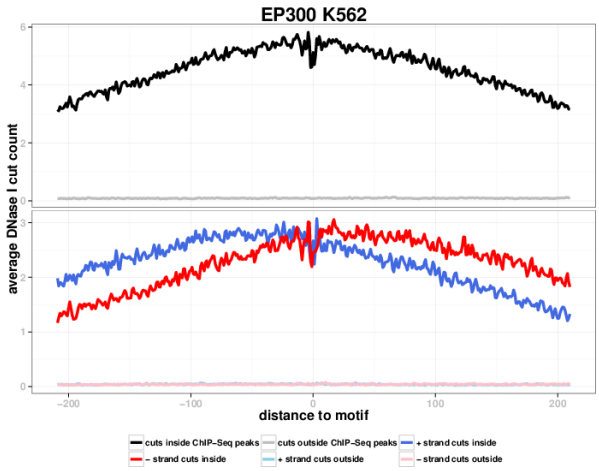

ELF1 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

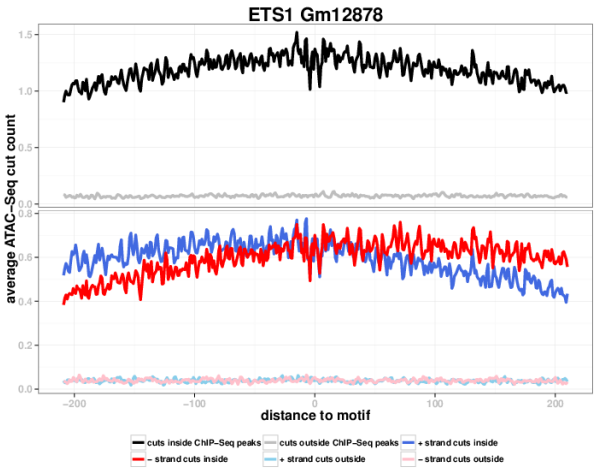

ELF1 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

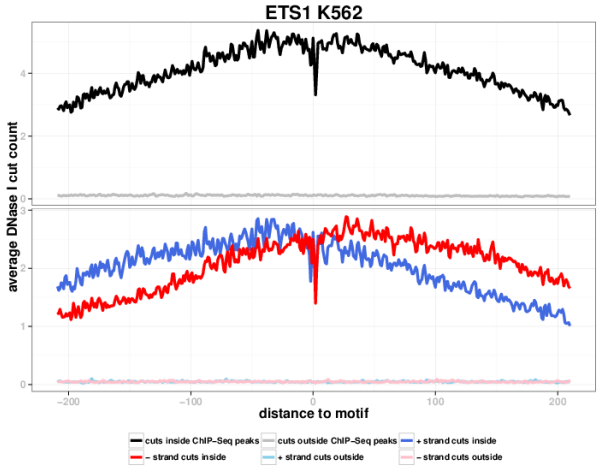

ELF1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

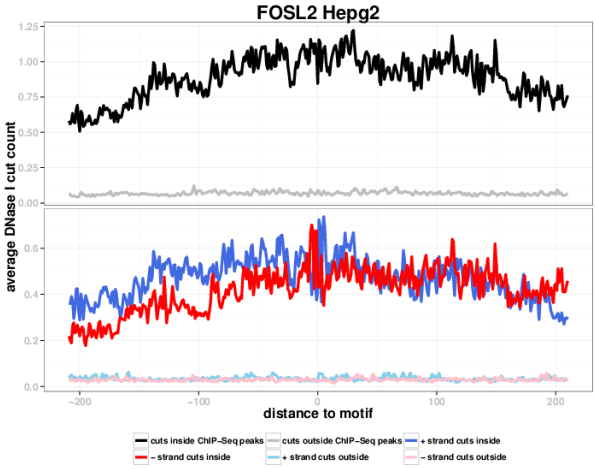

EP300 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

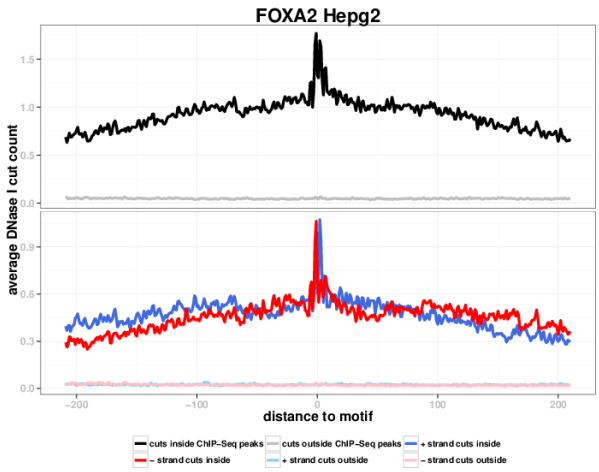

EP300 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

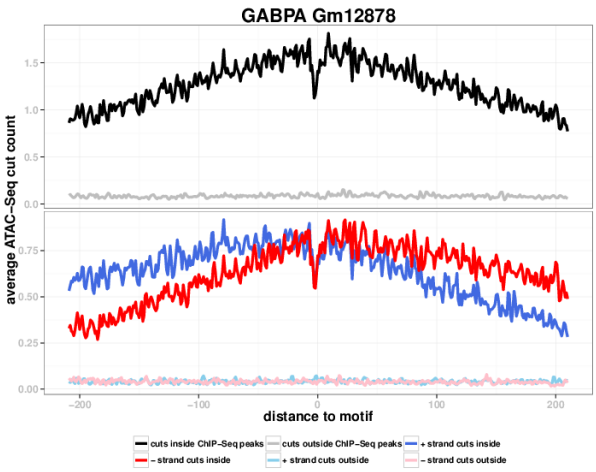

EP300 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

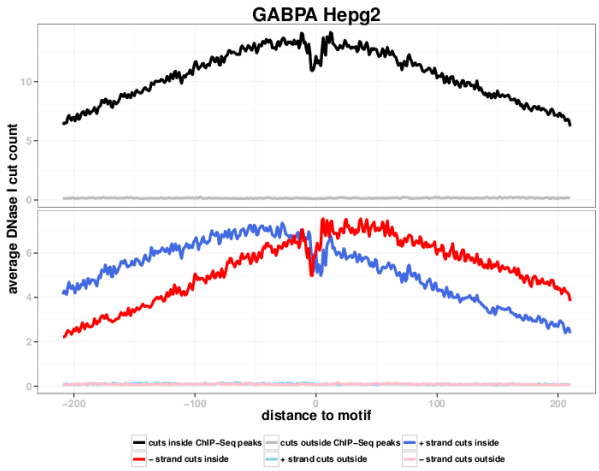

ETS1 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

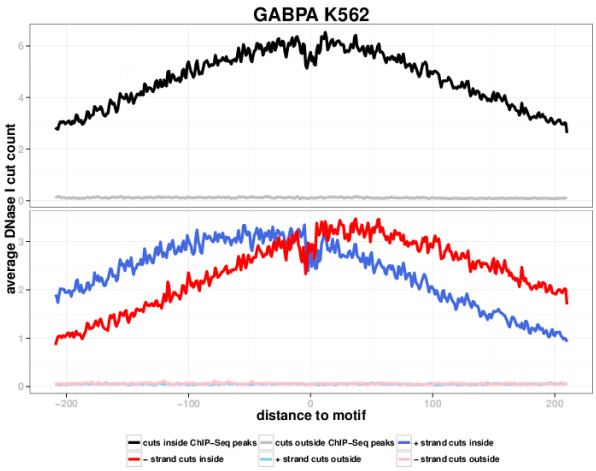

ETS1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

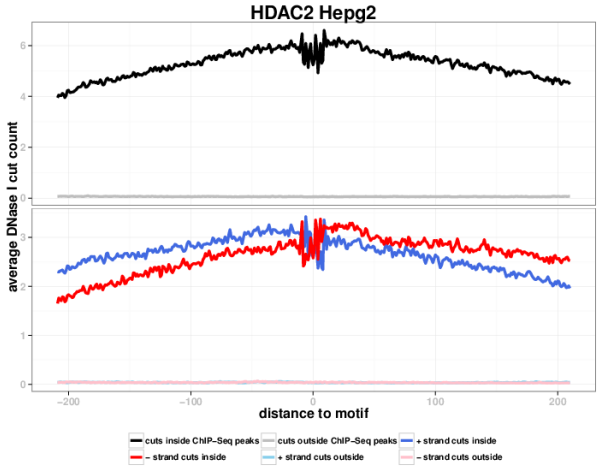

FOSL2 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

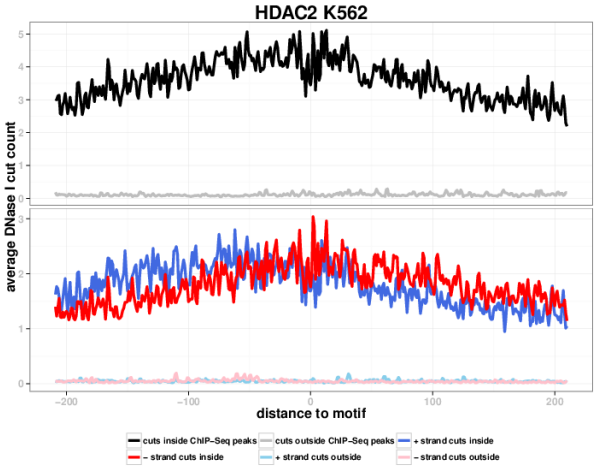

FOXA2 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

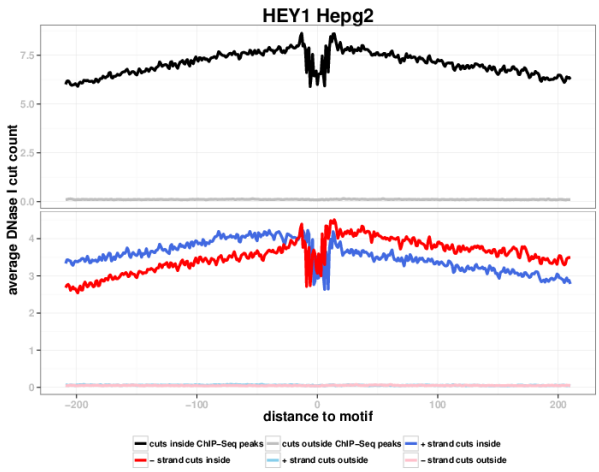

GABPA Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

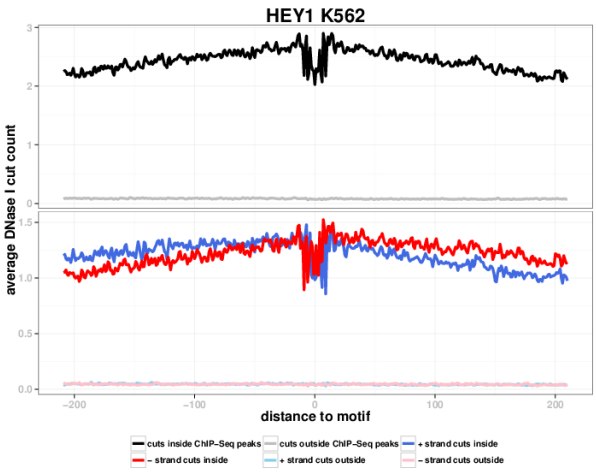

GABPA Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

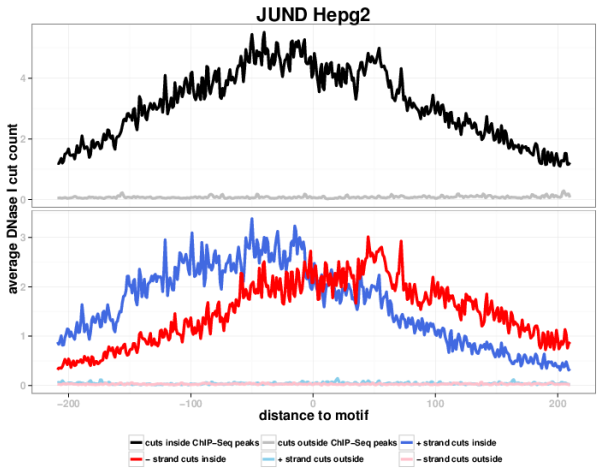

GABPA K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

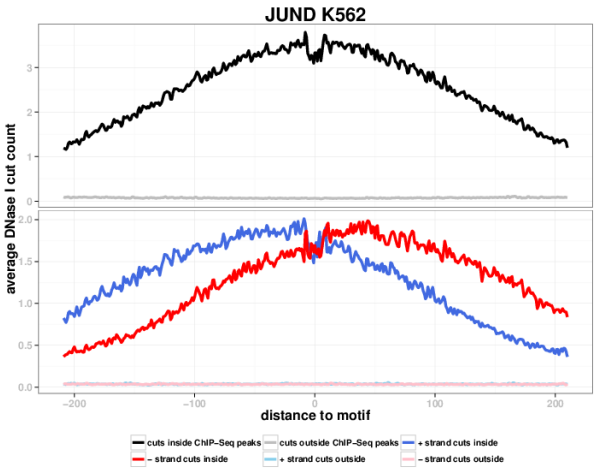

HDAC2 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

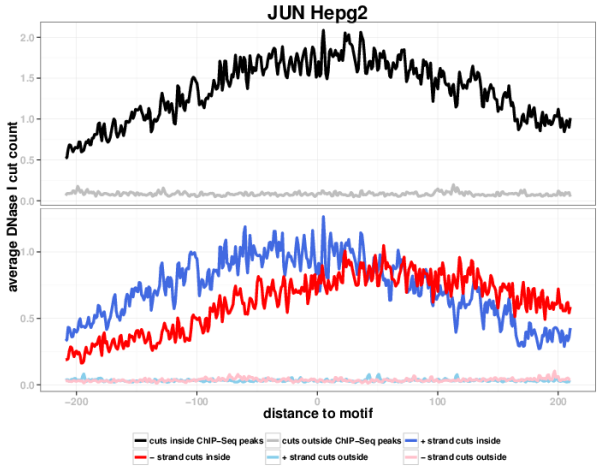

HDAC2 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

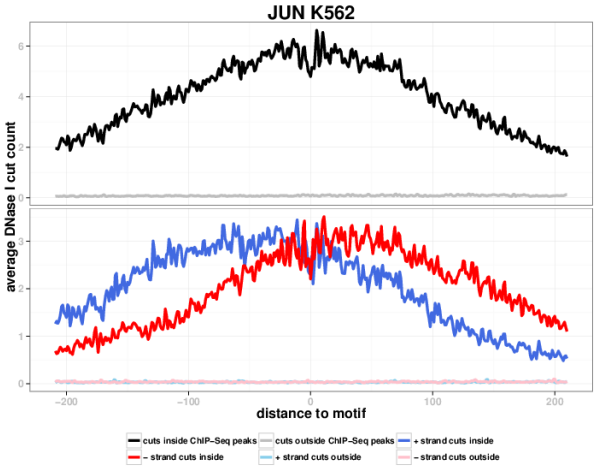

HEY1 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

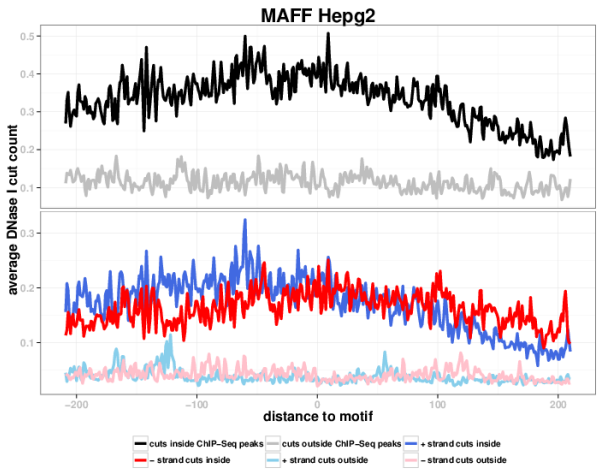

HEY1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

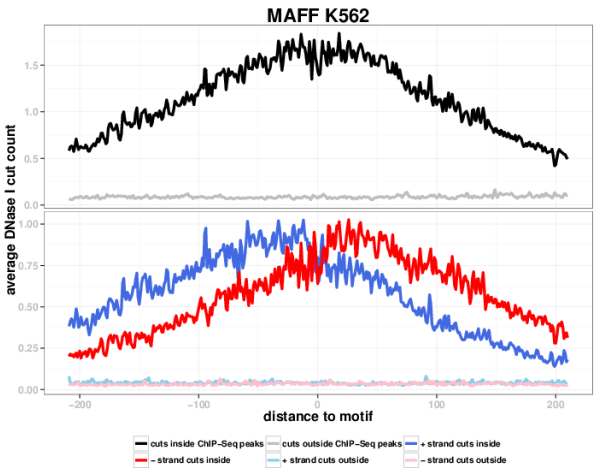

JUND Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

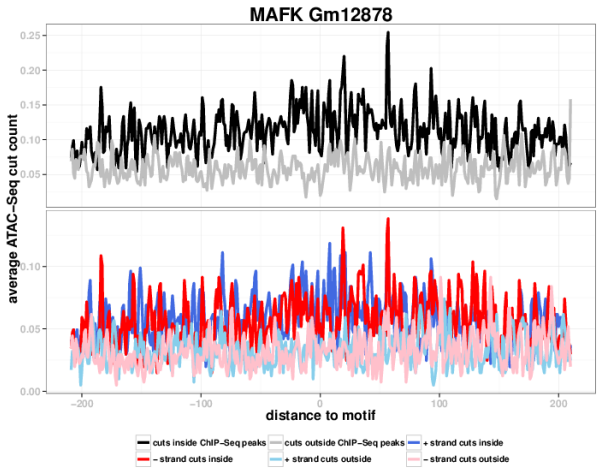

JUND K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

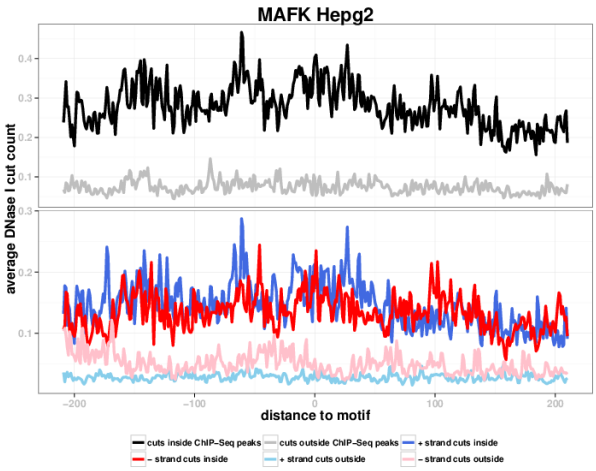

JUN Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

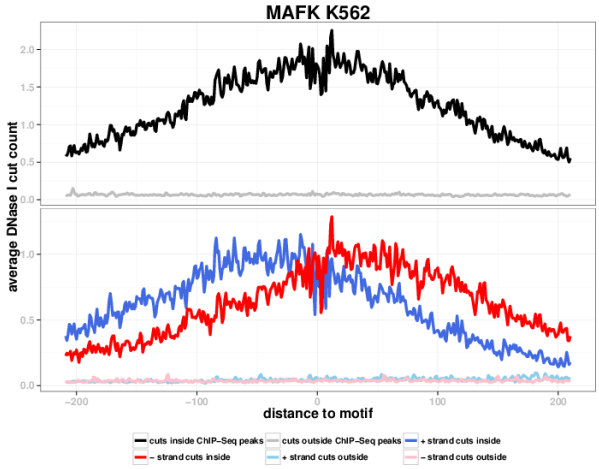

JUN K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

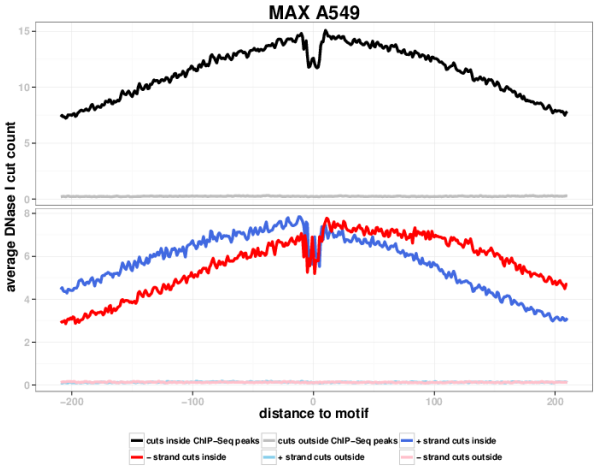

MAFF Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

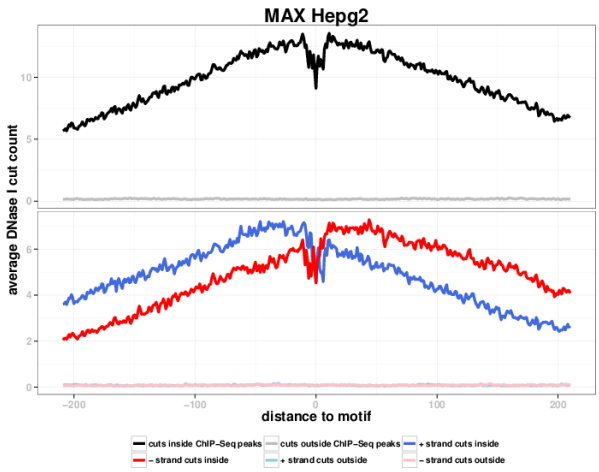

MAFF K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

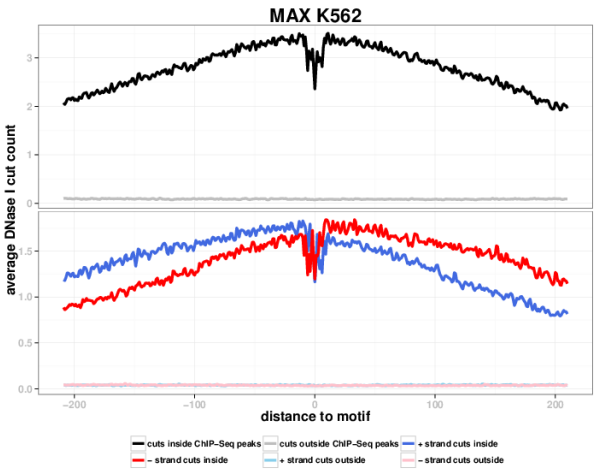

MAFK Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

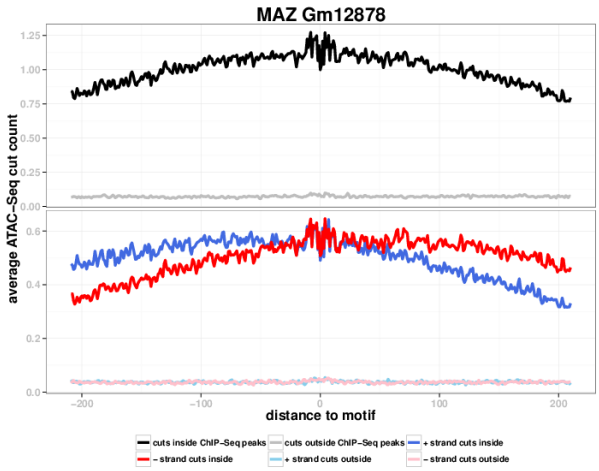

MAFK Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

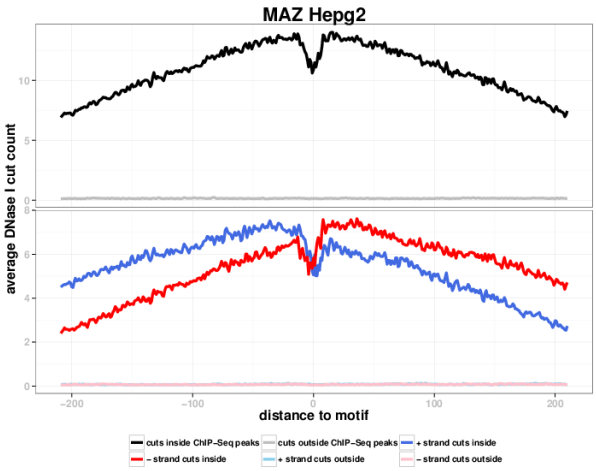

MAFK K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

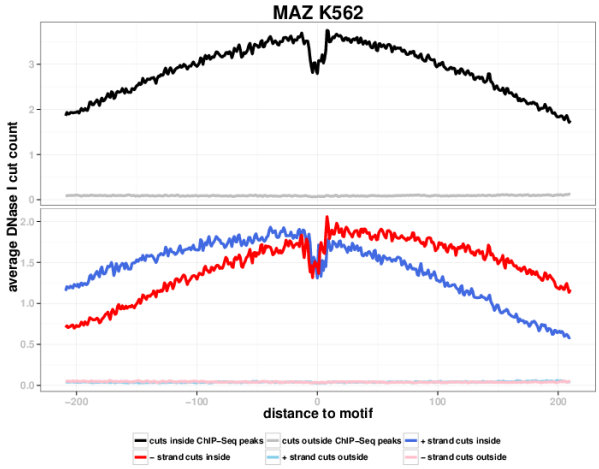

MAX A549
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

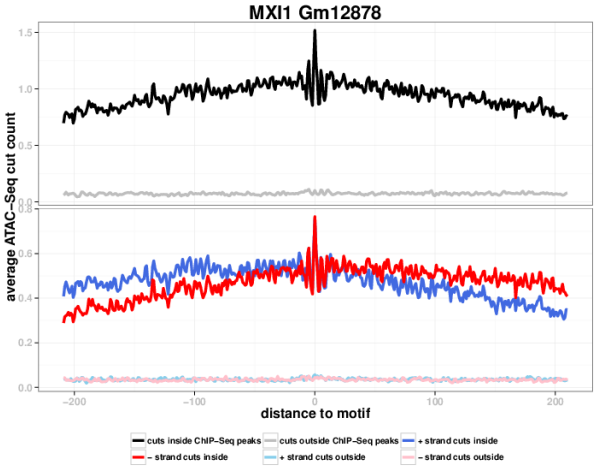

MAX Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

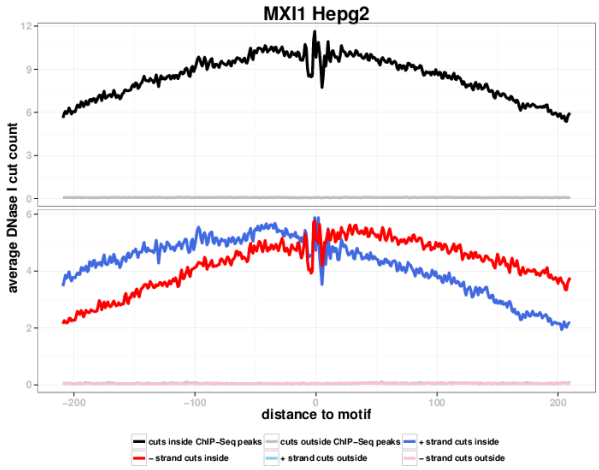

MAX K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

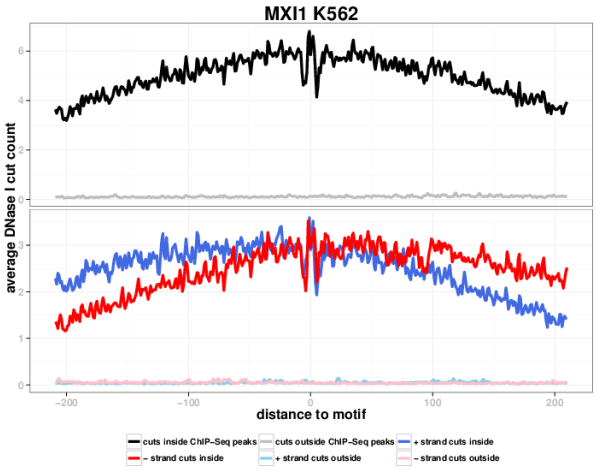

MAZ Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

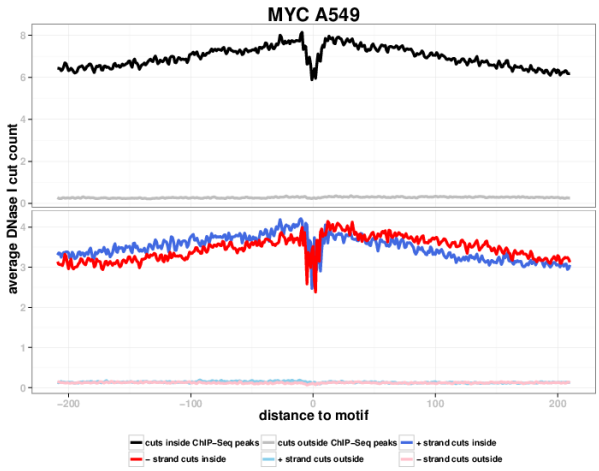

MAZ Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

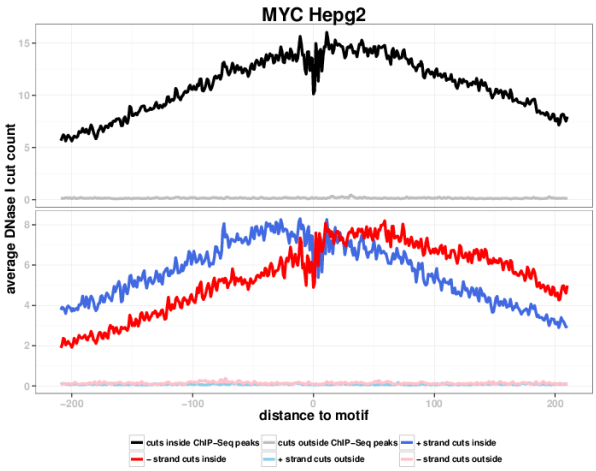

MAZ K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

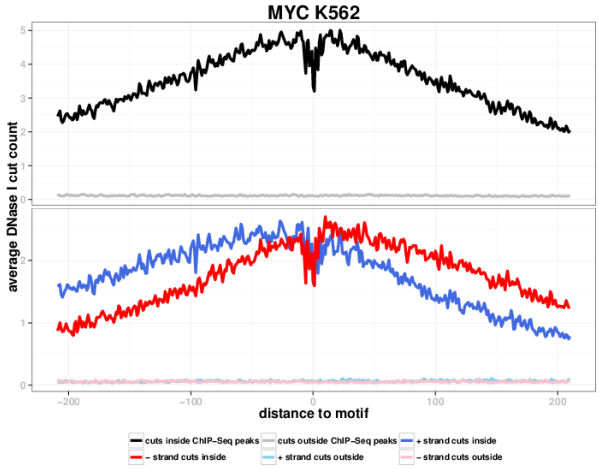

MXI1 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

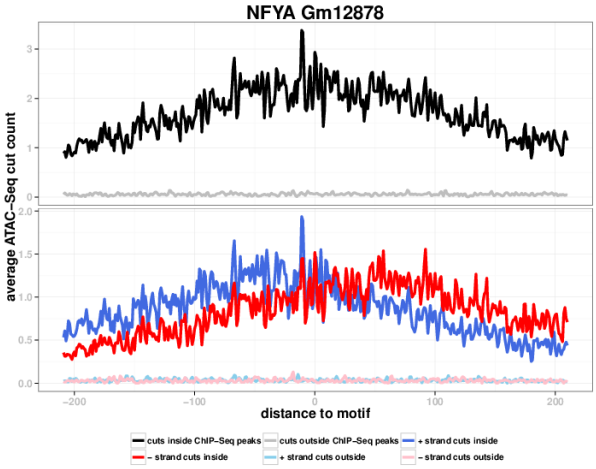

MXI1 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

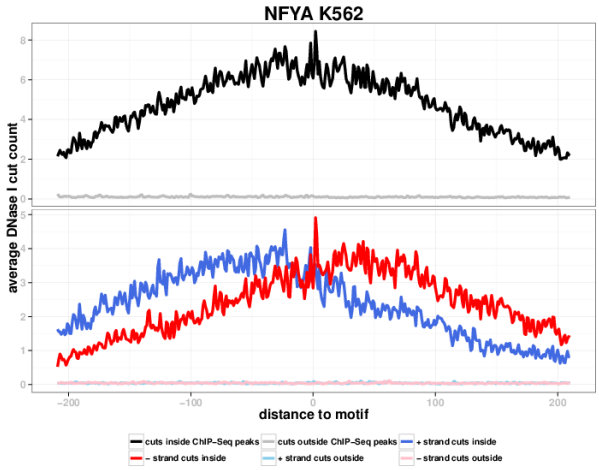

MXI1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

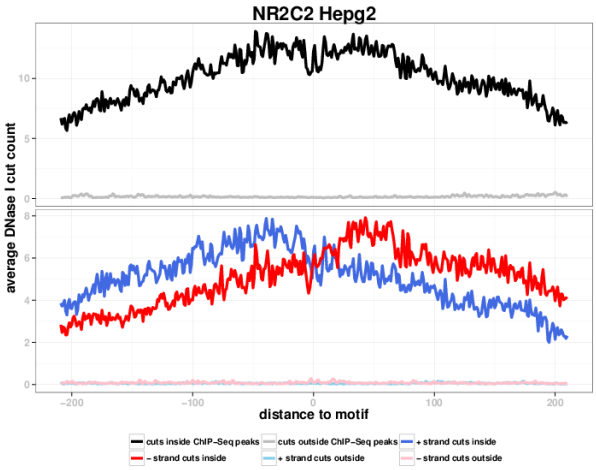

MYC A549
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

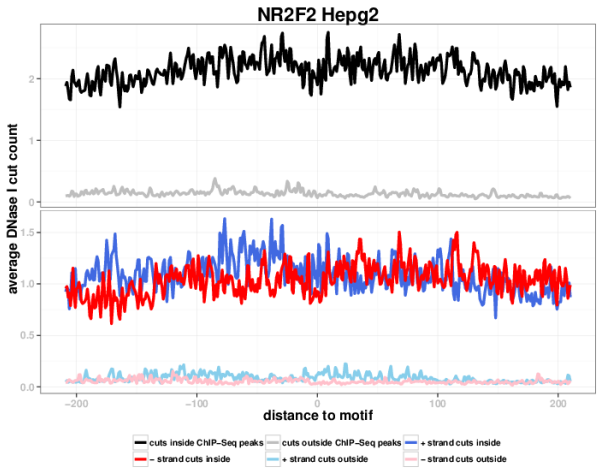

MYC Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

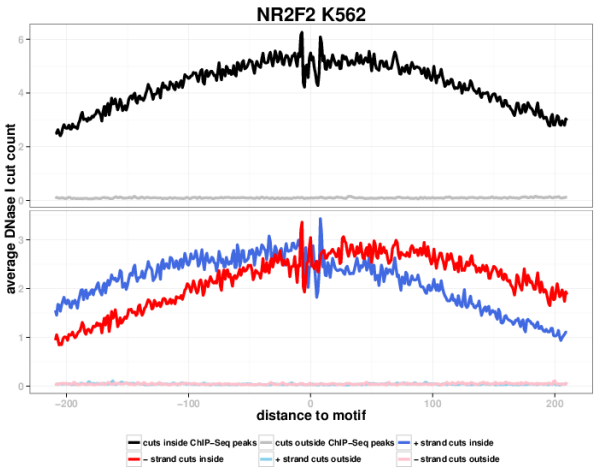

MYC K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

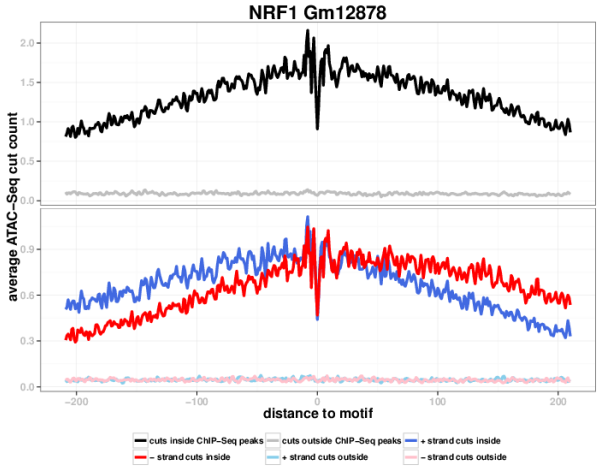

NFYA Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

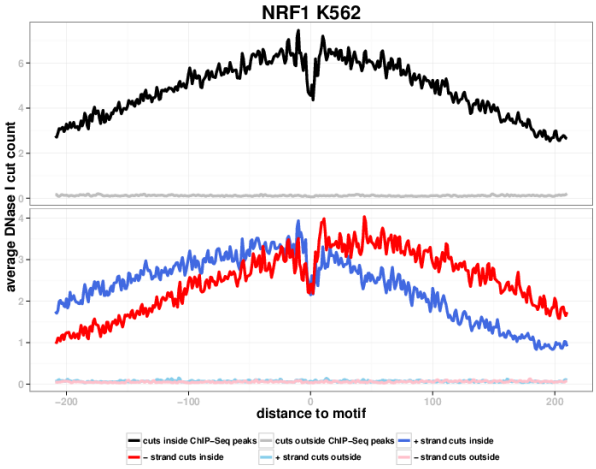

NFYA K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

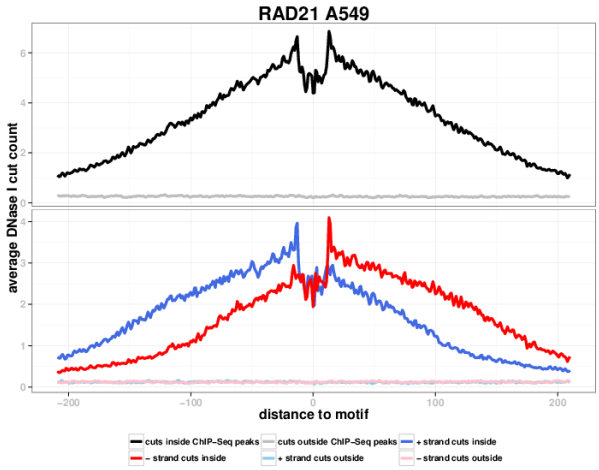

NR2C2 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

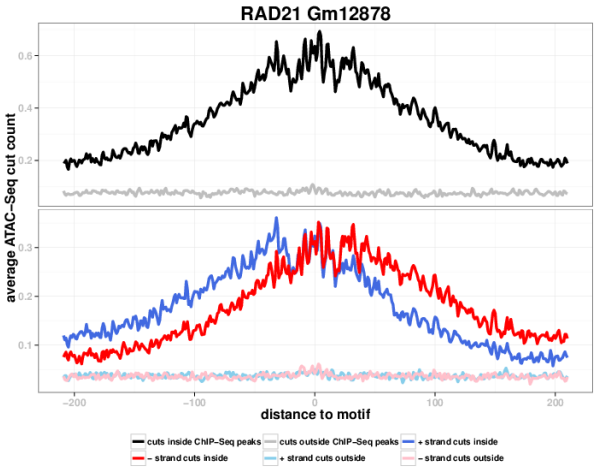

NR2F2 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

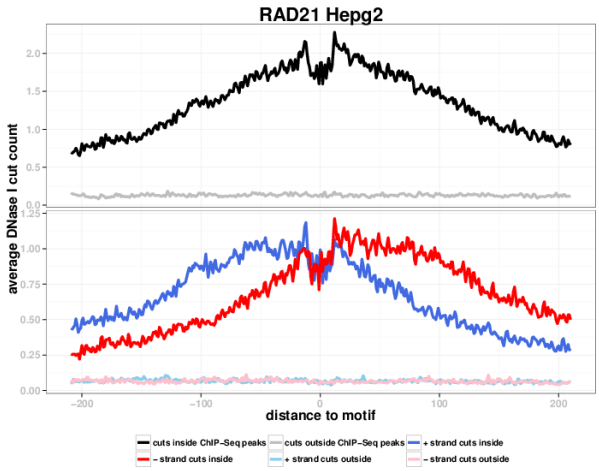

NR2F2 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

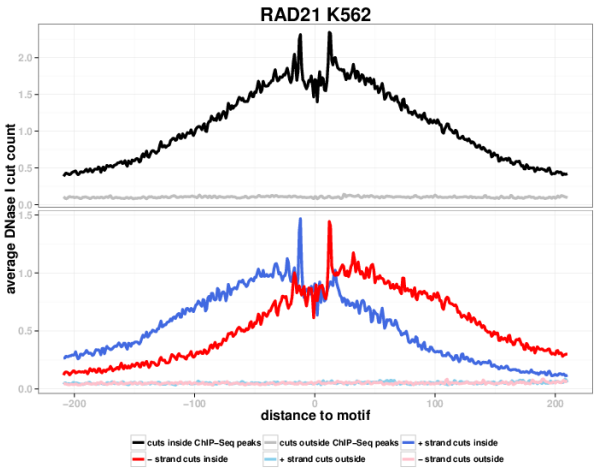

NRF1 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

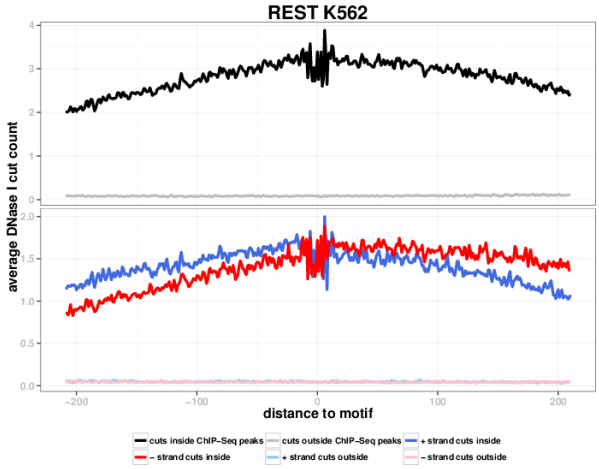

NRF1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

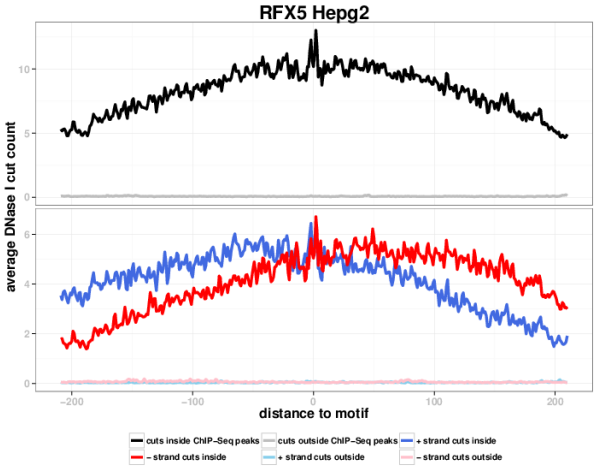

RAD21 A549
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

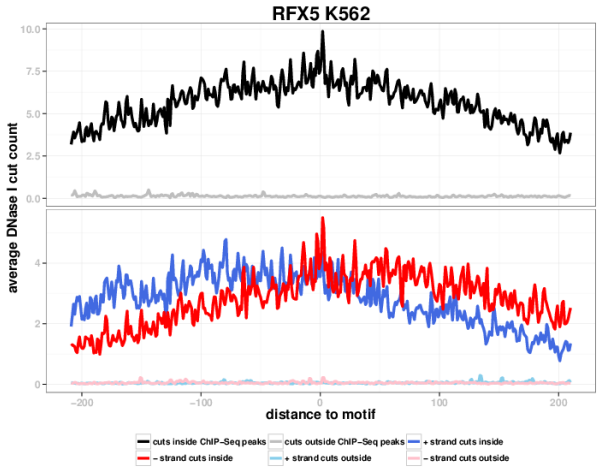

RAD21 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

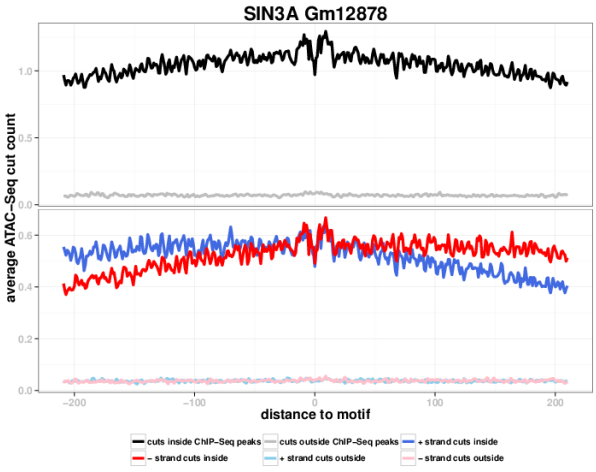

RAD21 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

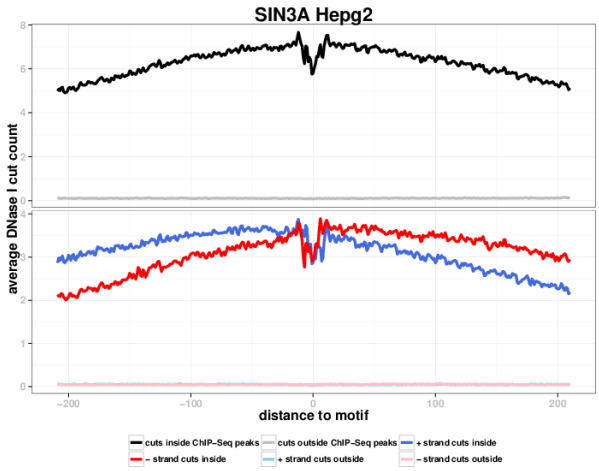

RAD21 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

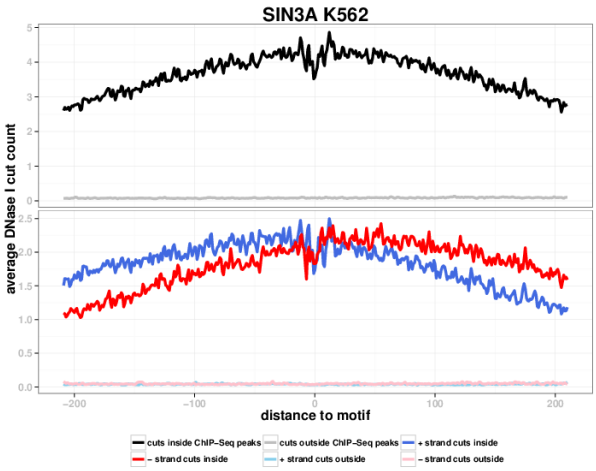

REST K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

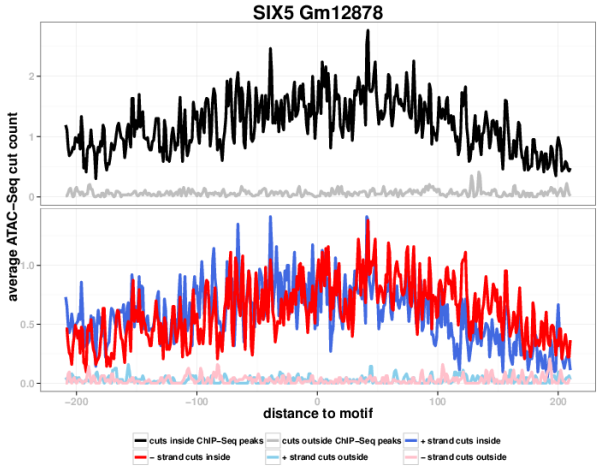

RFX5 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

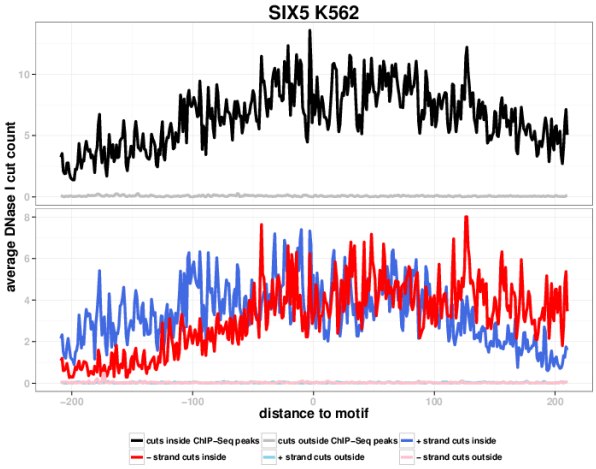

RFX5 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

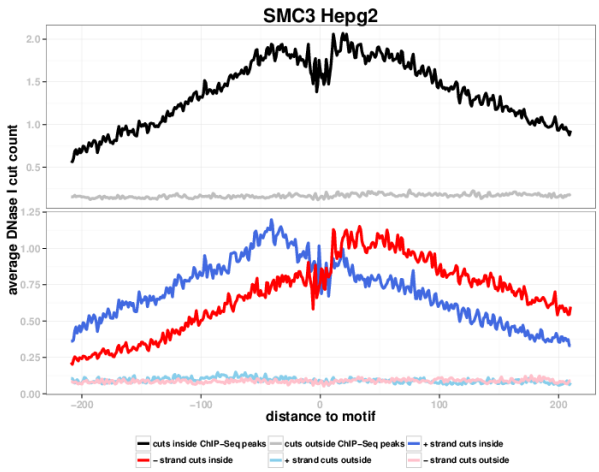

SIN3A Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

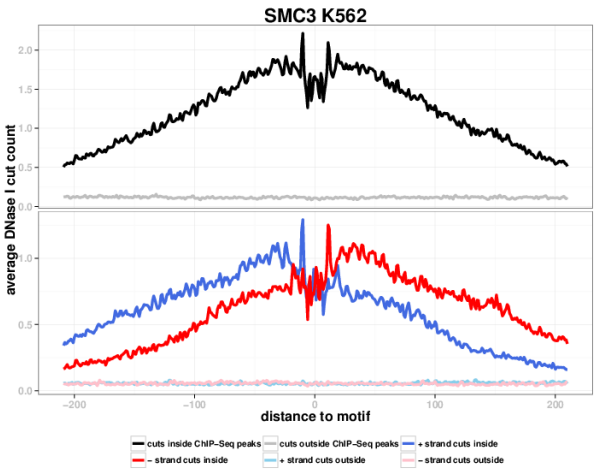

SIN3A Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

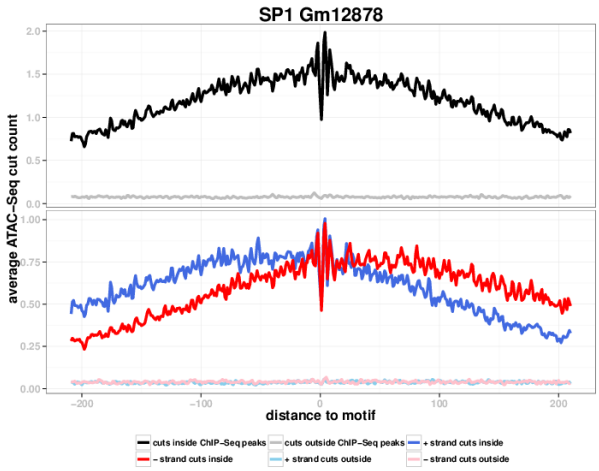

SIN3A K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

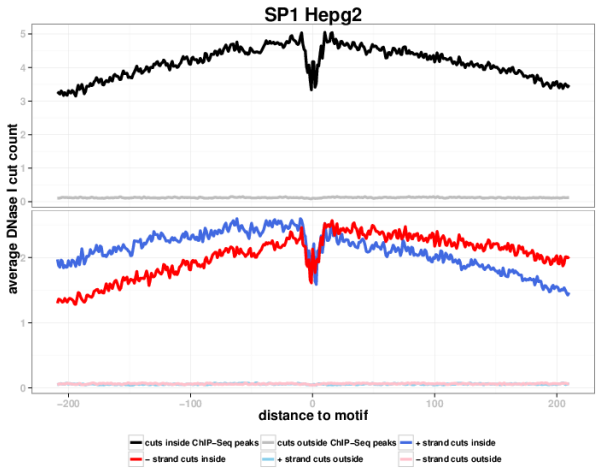

SIX5 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

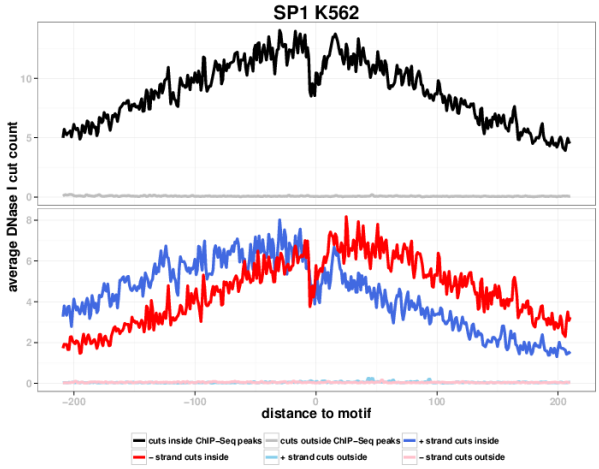

SIX5 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

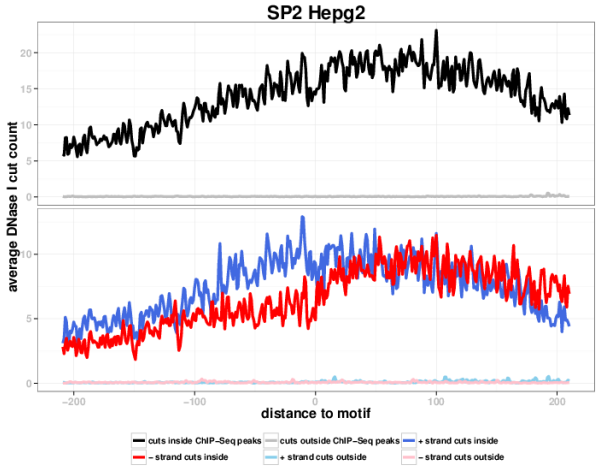

SMC3 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

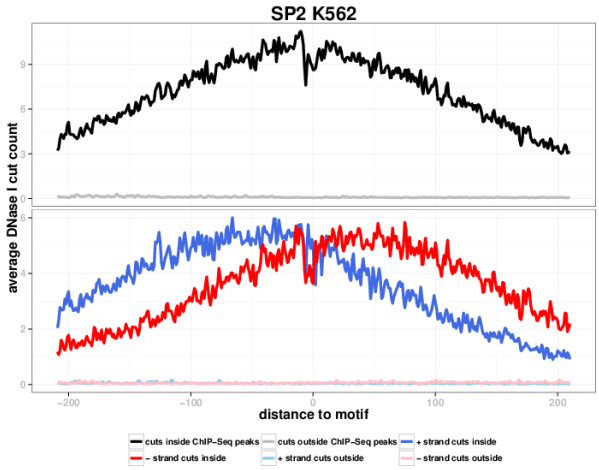

SMC3 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

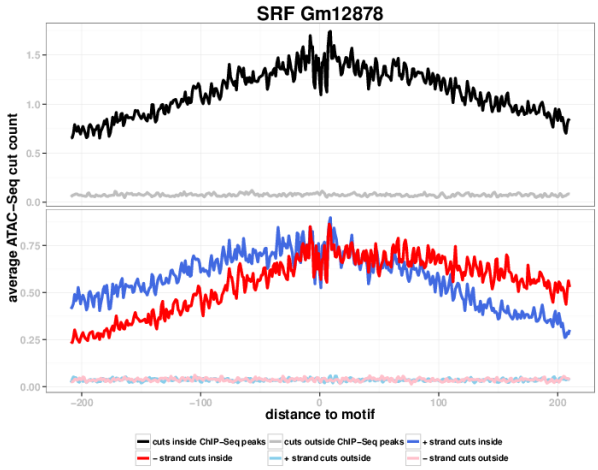

SP1 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

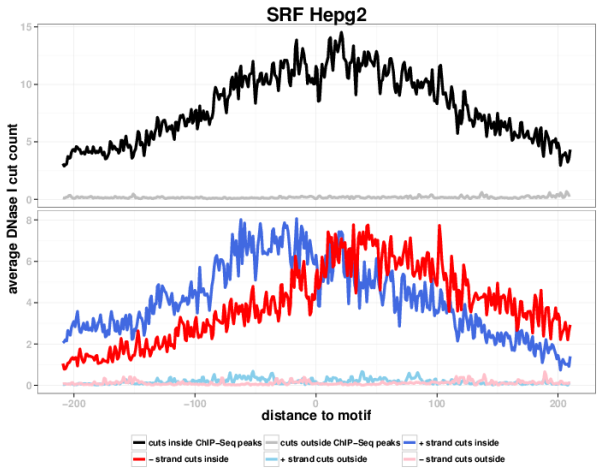

SP1 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

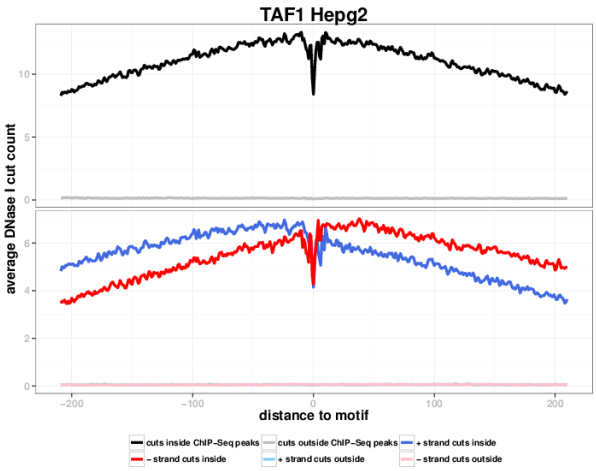

SP1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

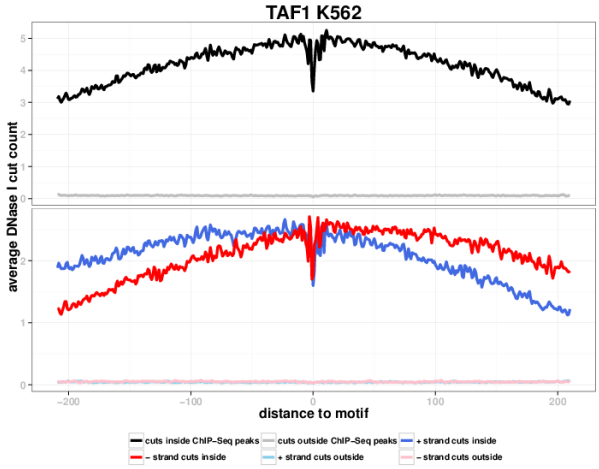

SP2 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

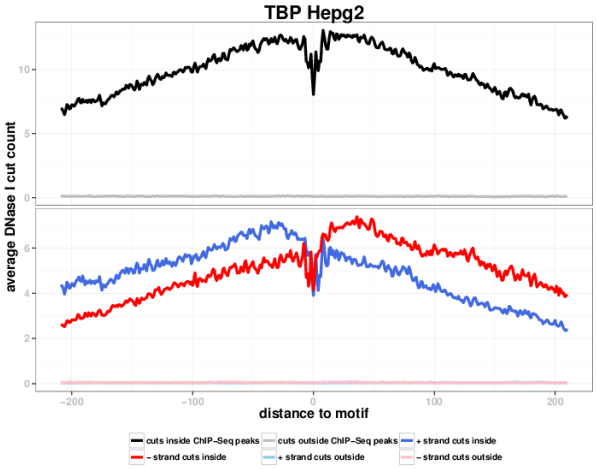

SP2 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

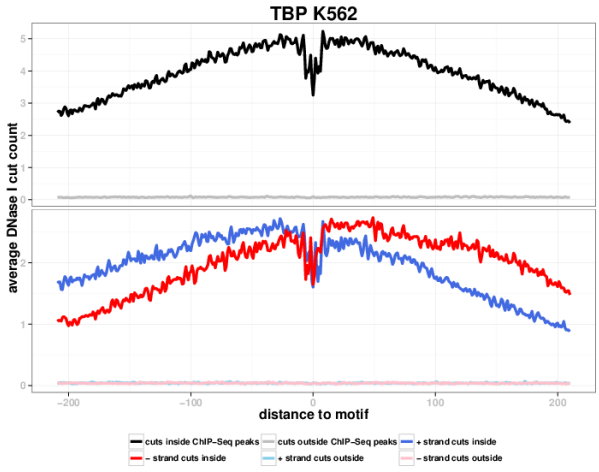

SRF Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

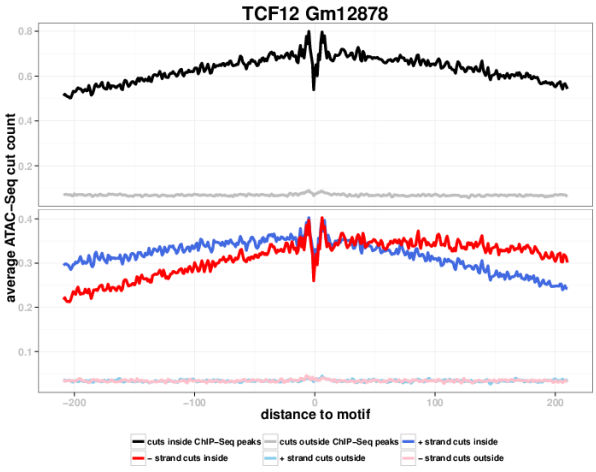

SRF Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

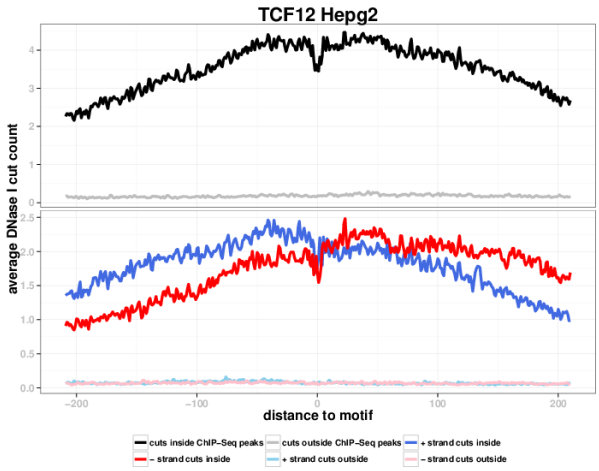

TAF1 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

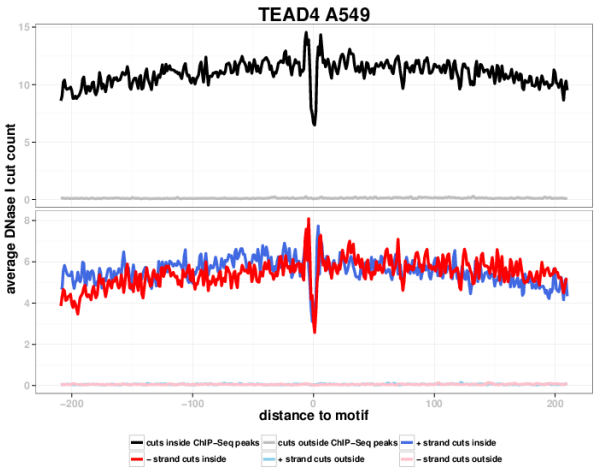

TAF1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

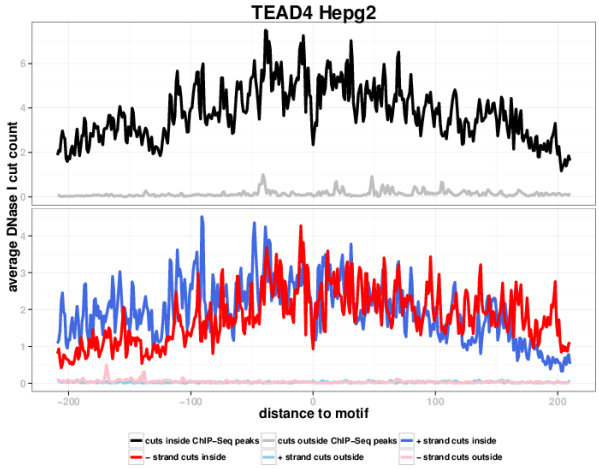

TBP Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

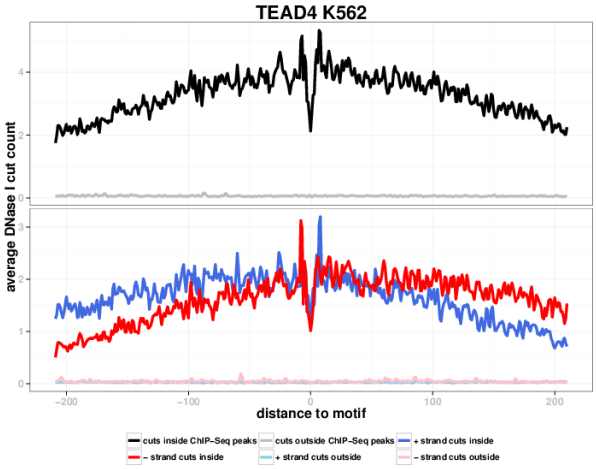

TBP K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

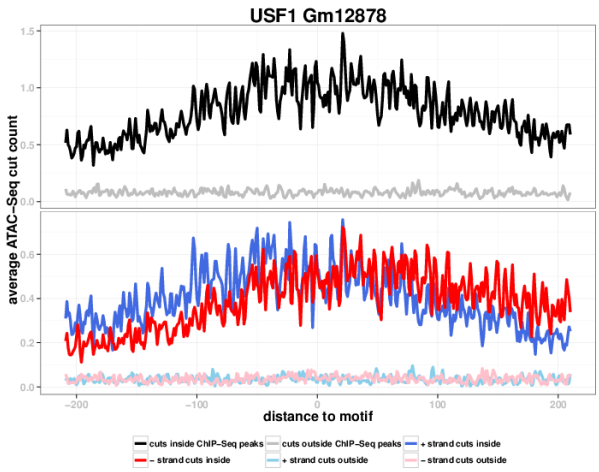

TCF12 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

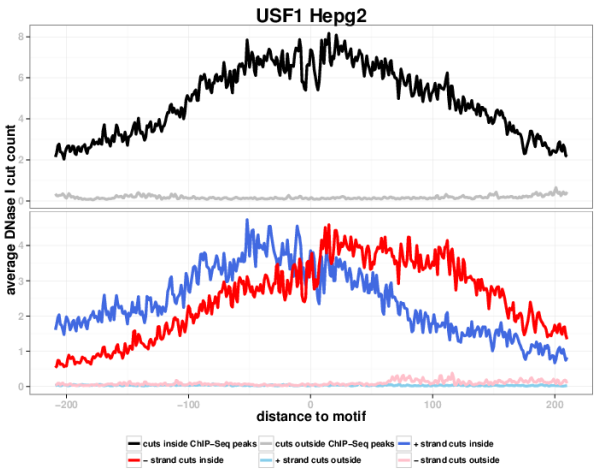

TCF12 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

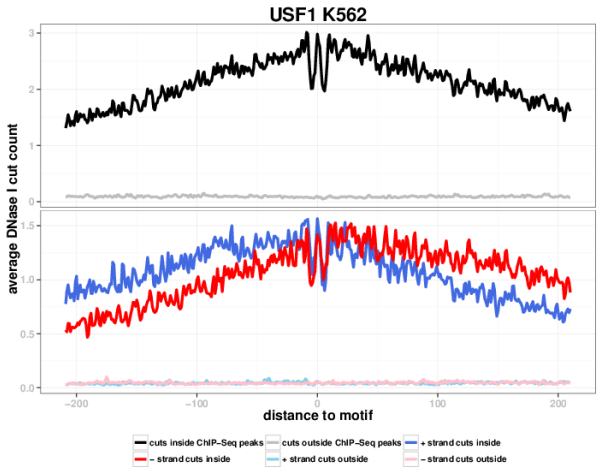

TEAD4 A549
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

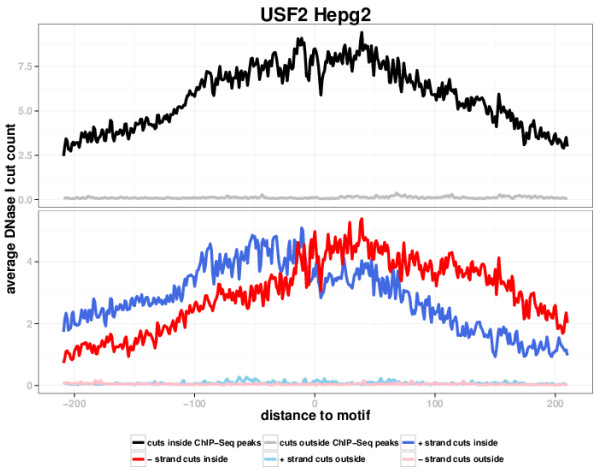

TEAD4 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

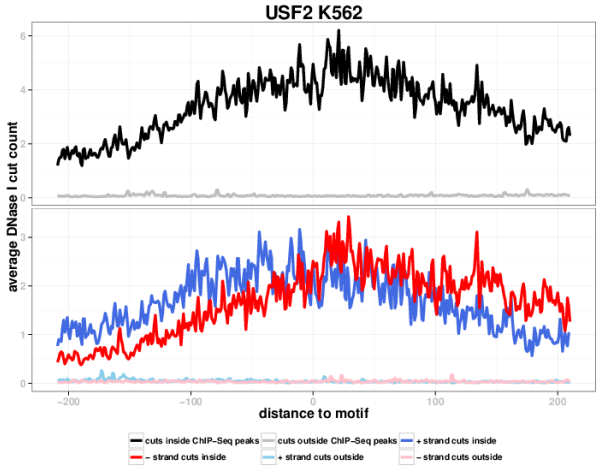

TEAD4 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

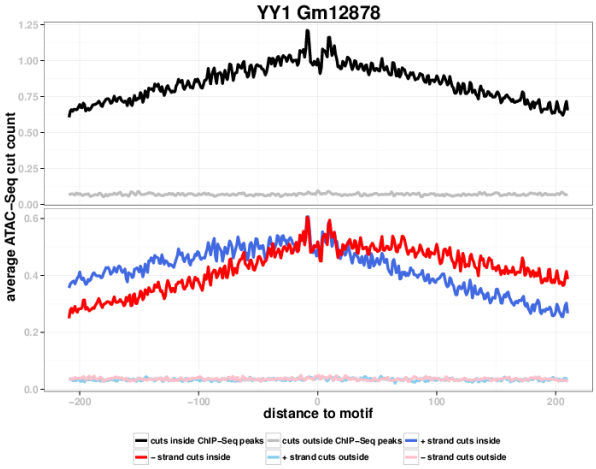

USF1 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

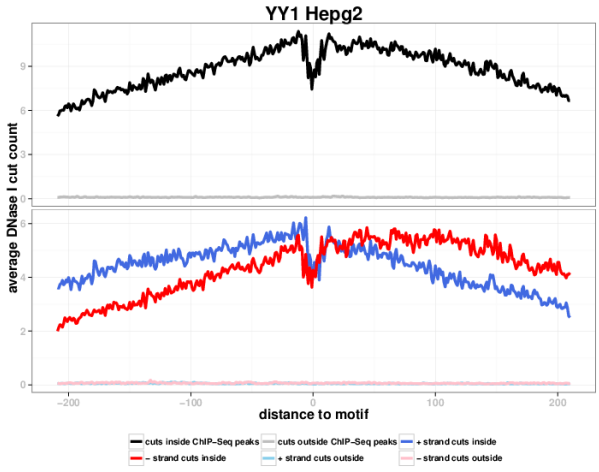

USF1 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

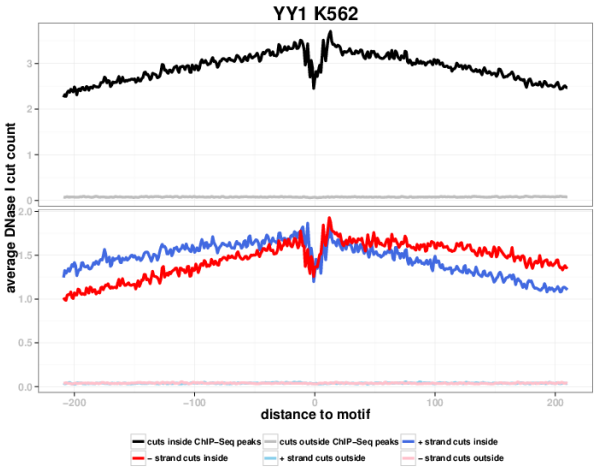

USF1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

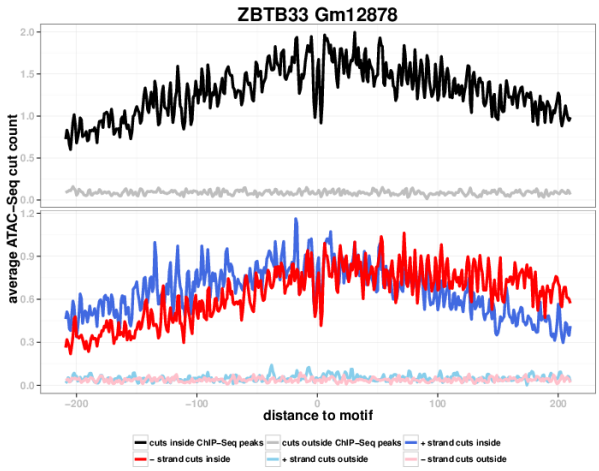

USF2 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

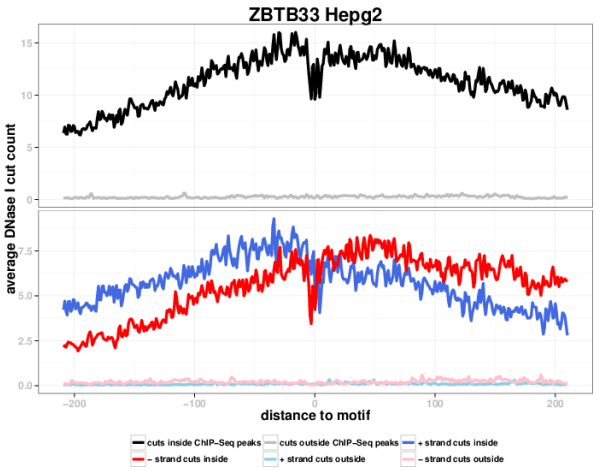

USF2 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

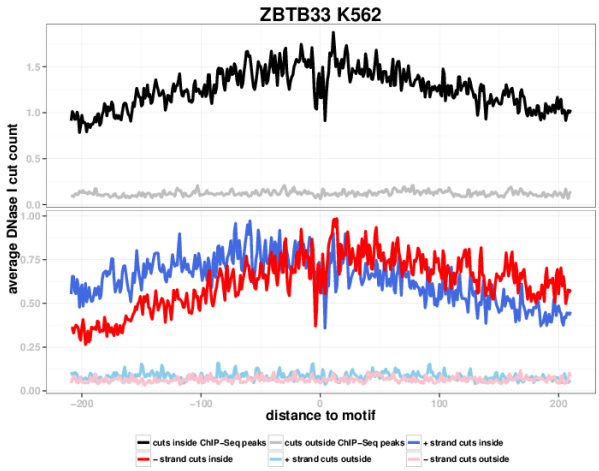

YY1 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

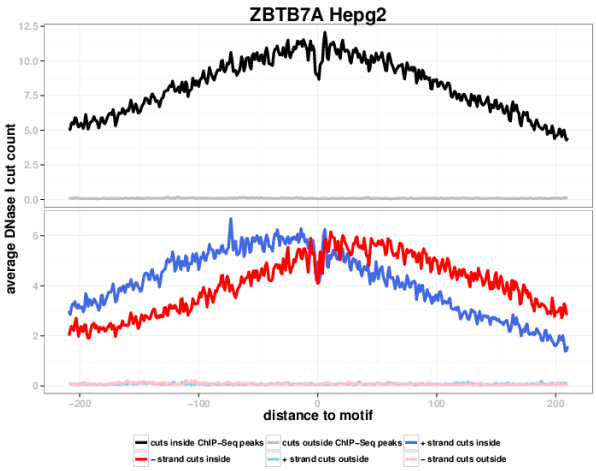

YY1 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

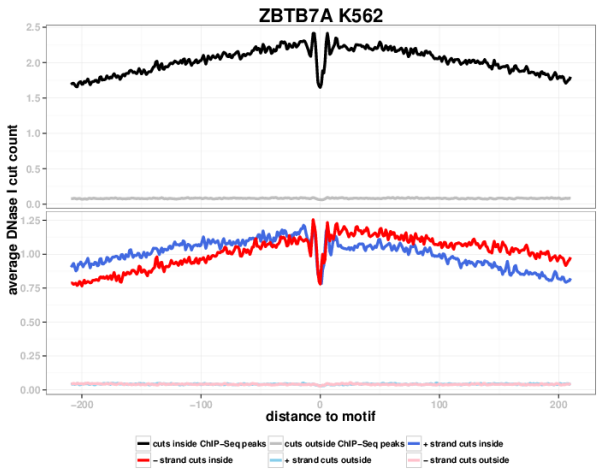

YY1 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

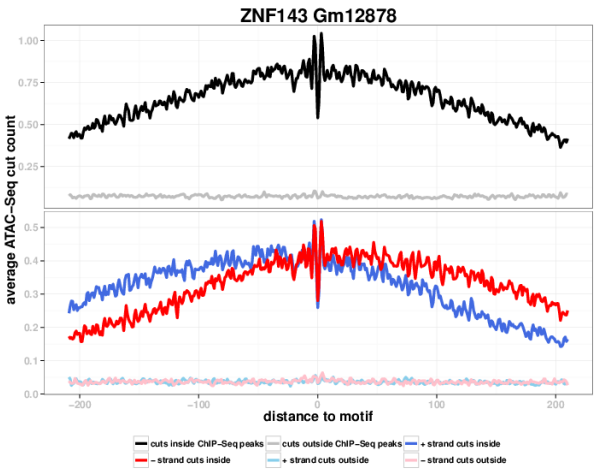

ZBTB33 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

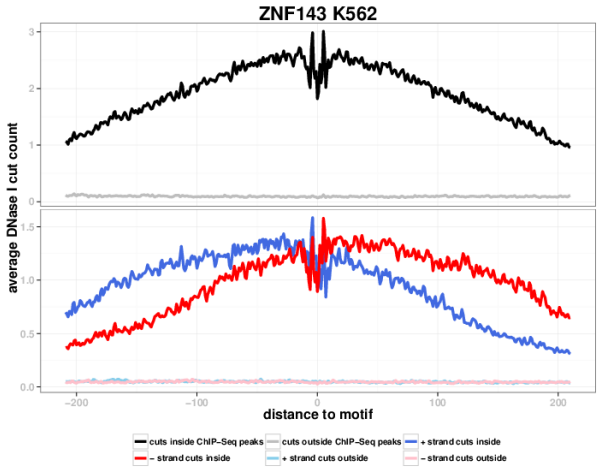

ZBTB33 Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

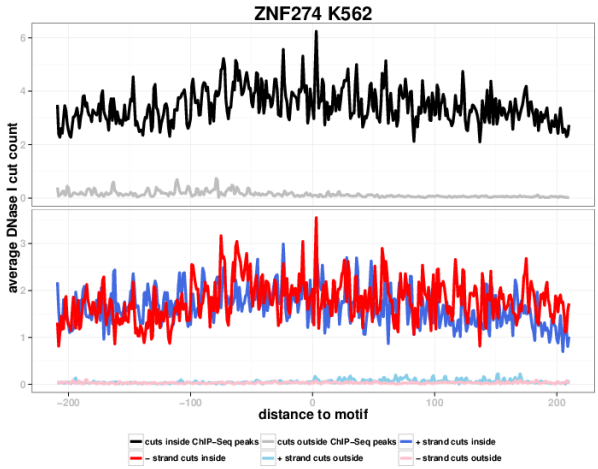

ZBTB33 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

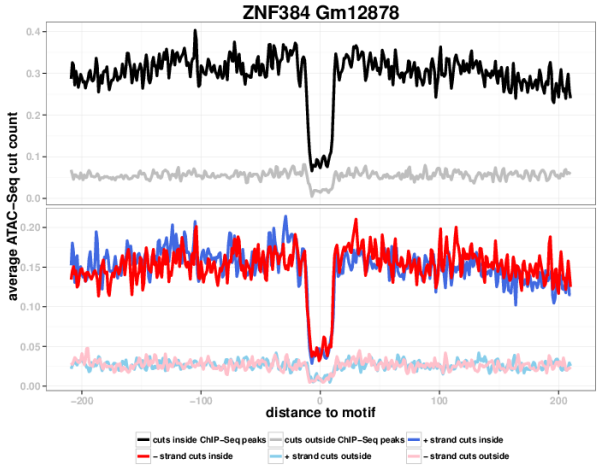

ZBTB7A Hepg2
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

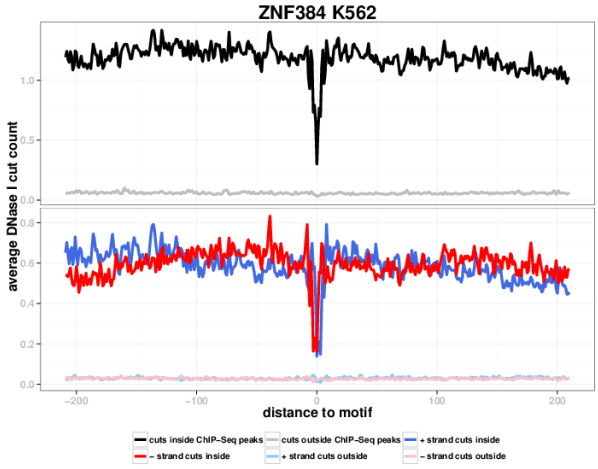

ZBTB7A K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside


ZNF143 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

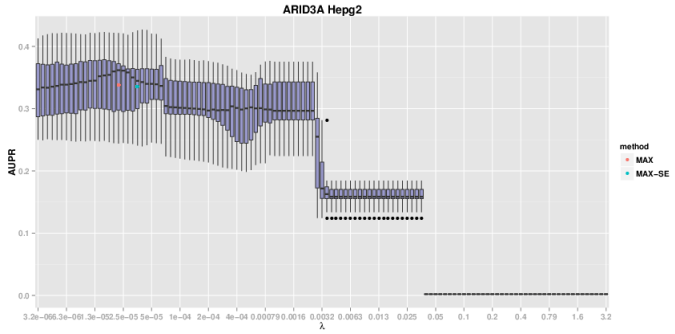

ZNF143 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

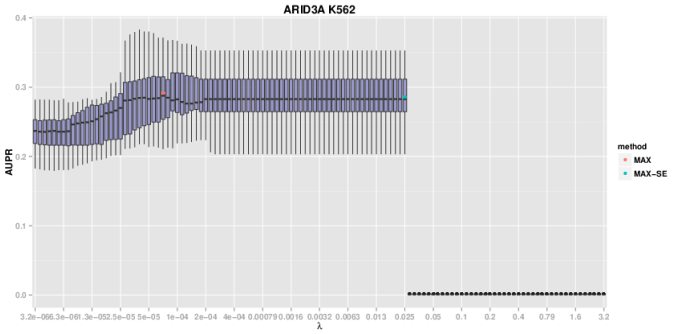

ZNF274 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

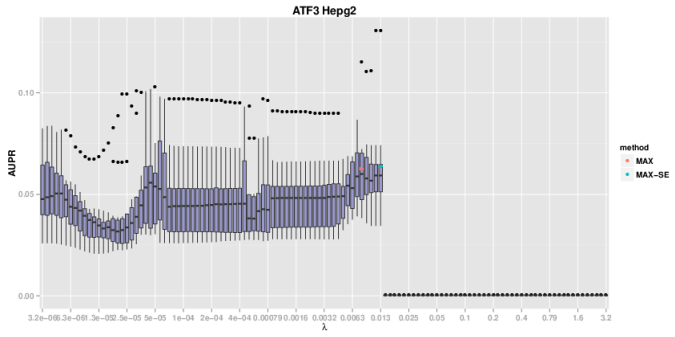

ZNF384 Gm12878
average ATAC−Seq cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

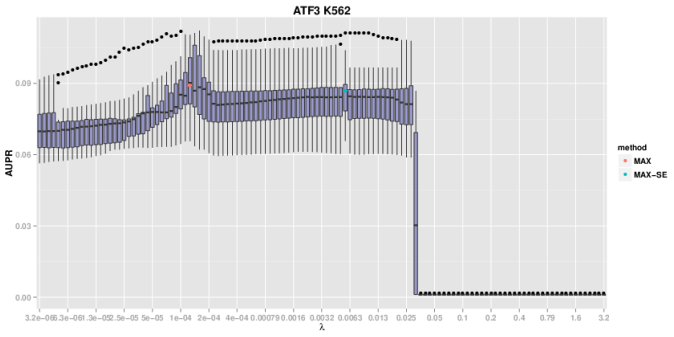

ZNF384 K562
average DNase I cut count
distance to motif
cuts inside ChIP−Seq peaks
cuts outside ChIP−Seq peaks
+ strand cuts inside
− strand cuts inside
+ strand cuts outside
− strand cuts outside

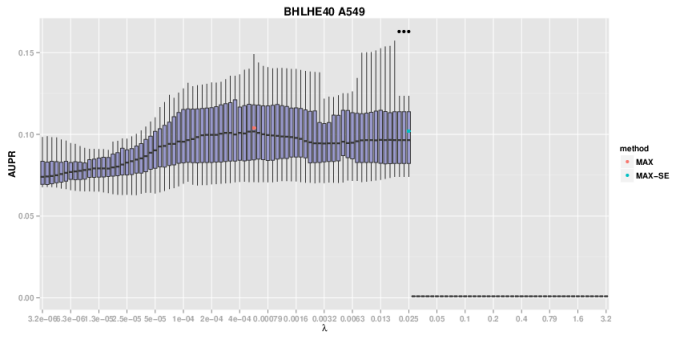
Appendix B Sparse logistic regression training plots

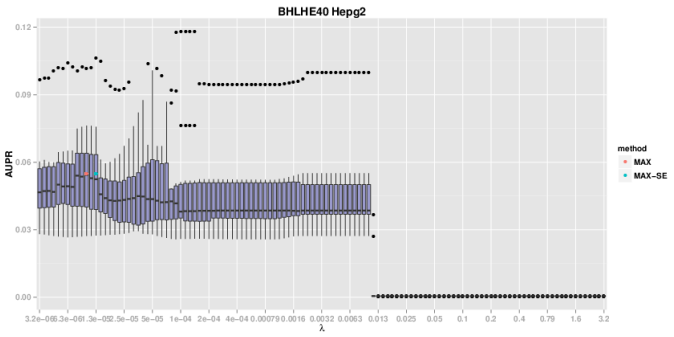


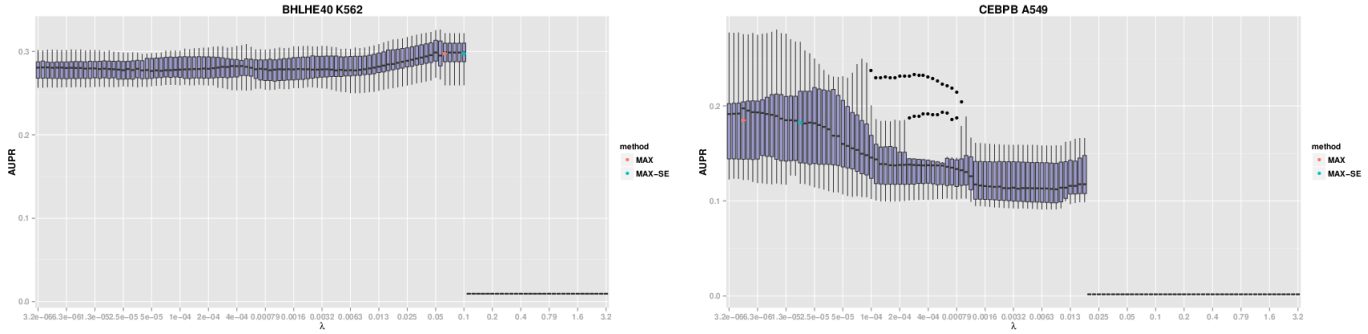


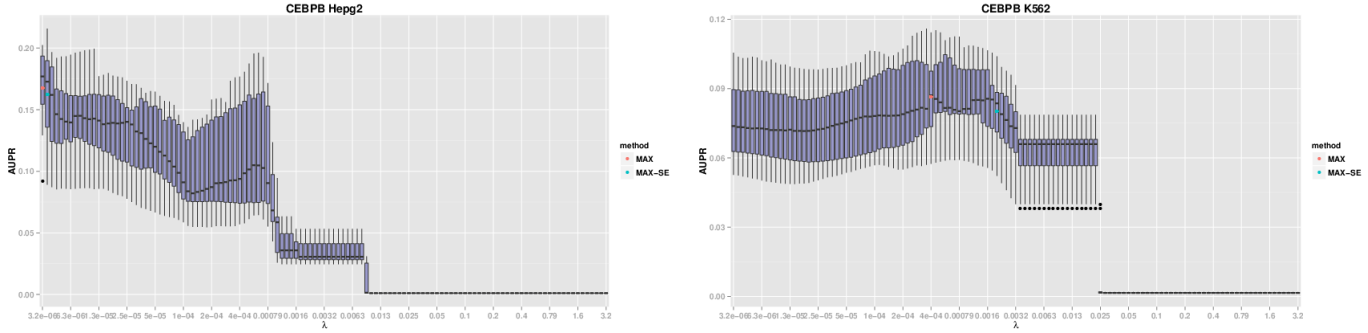


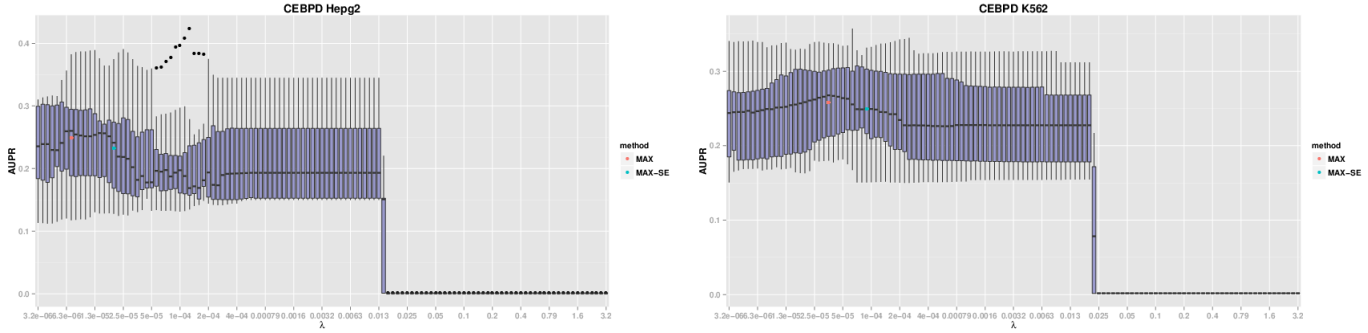


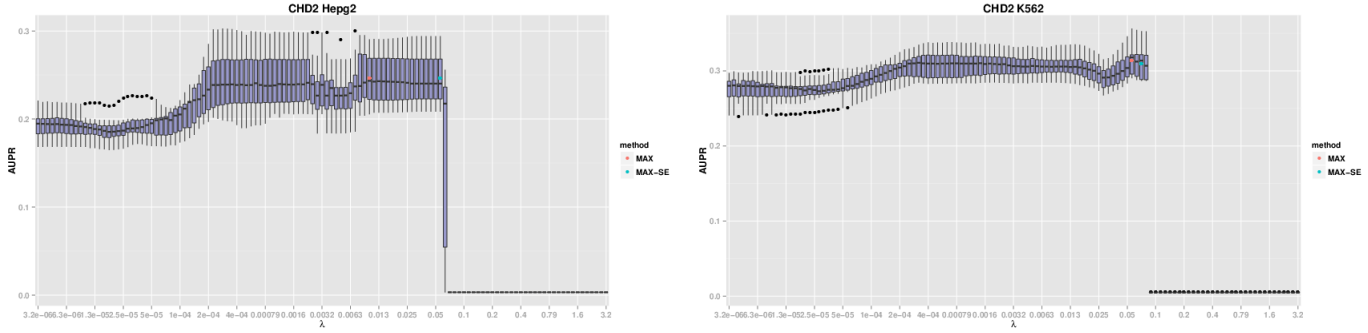


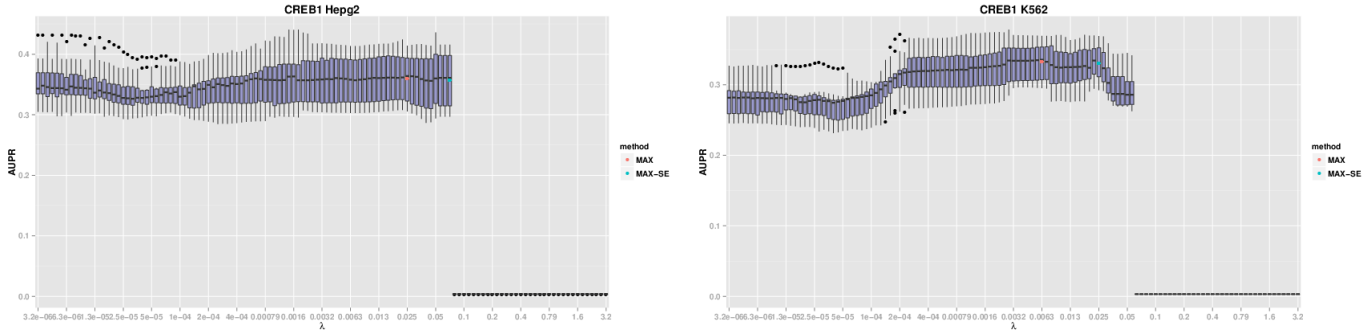



BHLHE40 K562
CEBPB A549
AUPR
λ
method
MAX
MAX−SE

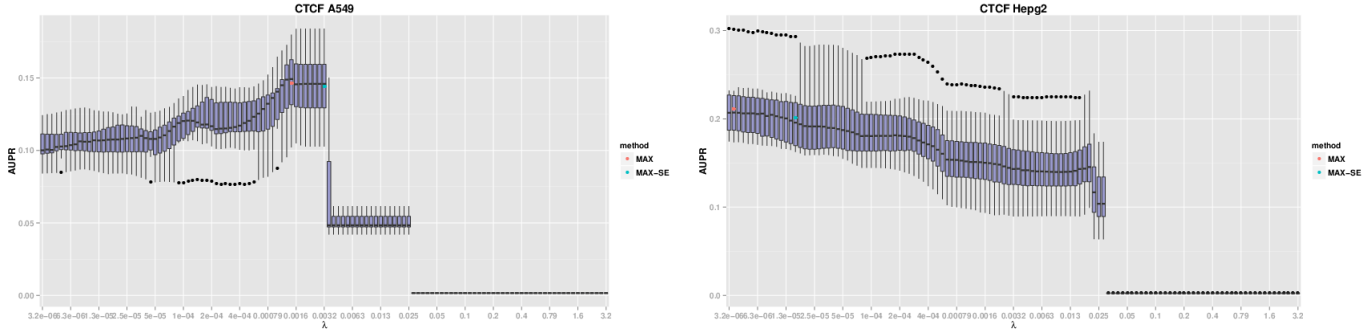

CEBPB Hepg2
CEBPB K562
AUPR
λ
method
MAX
MAX−SE


CEBPD Hepg2
CEBPD K562
AUPR
λ
method
MAX
MAX−SE

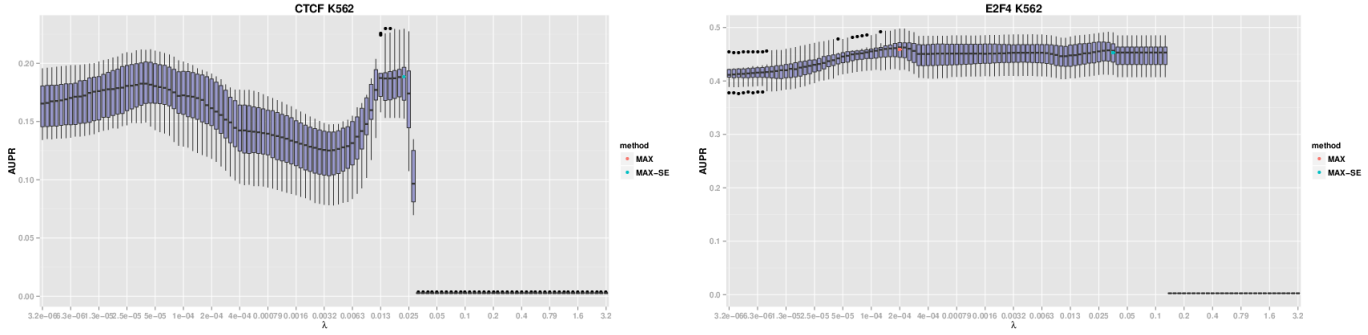

CHD2 Hepg2
CHD2 K562
AUPR
λ
method
MAX
MAX−SE


CREB1 Hepg2
AUPR
λ
method
MAX
MAX−SE

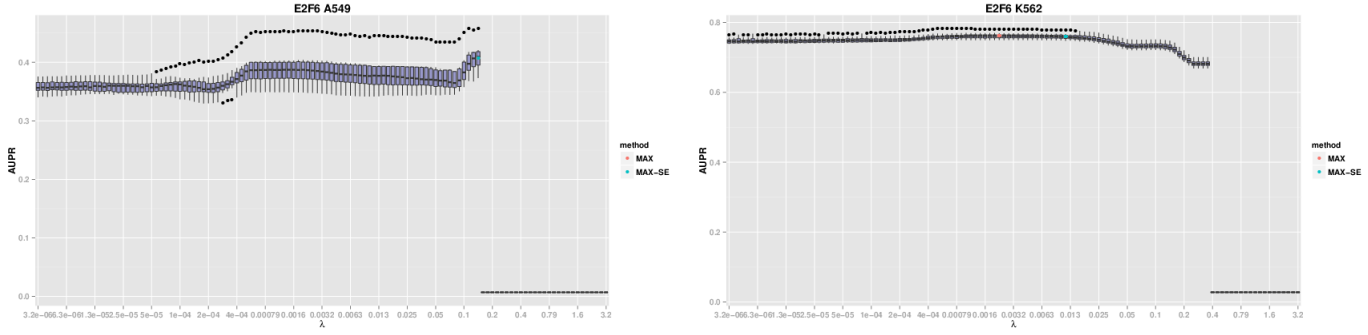
CREB1 K562
AUPR
λ
method
MAX
MAX−SE


CTCF A549
AUPR
λ
method
MAX
MAX−SE

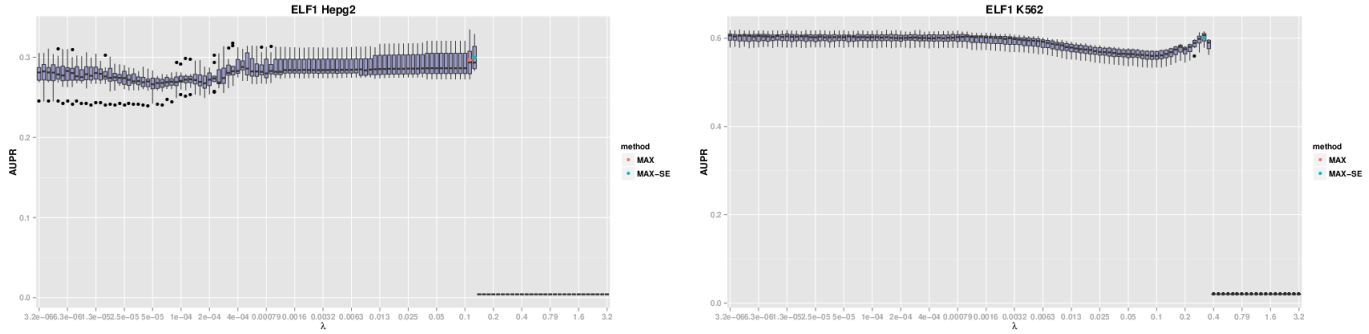
CTCF Hepg2
AUPR
λ
method
MAX
MAX−SE


CTCF K562
AUPR
λ
method
MAX
MAX−SE

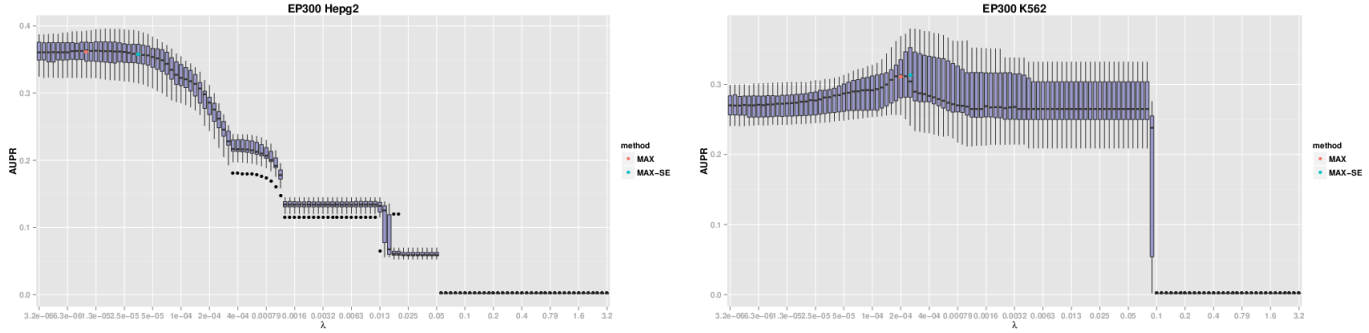
E2F4 K562
AUPR
λ
method
MAX
MAX−SE


E2F6 A549
AUPR
λ
method
MAX
MAX−SE

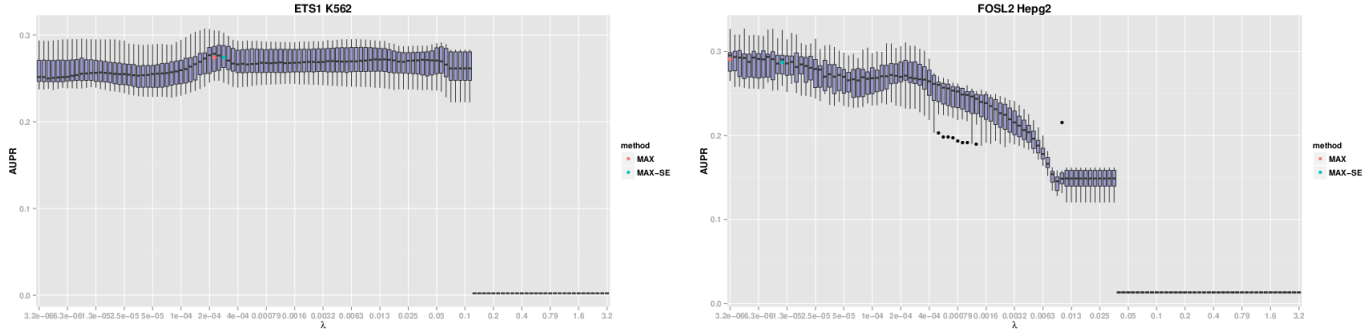
E2F6 K562
AUPR
λ
method
MAX
MAX−SE


ELF1 Hepg2
AUPR
λ
method
MAX
MAX−SE

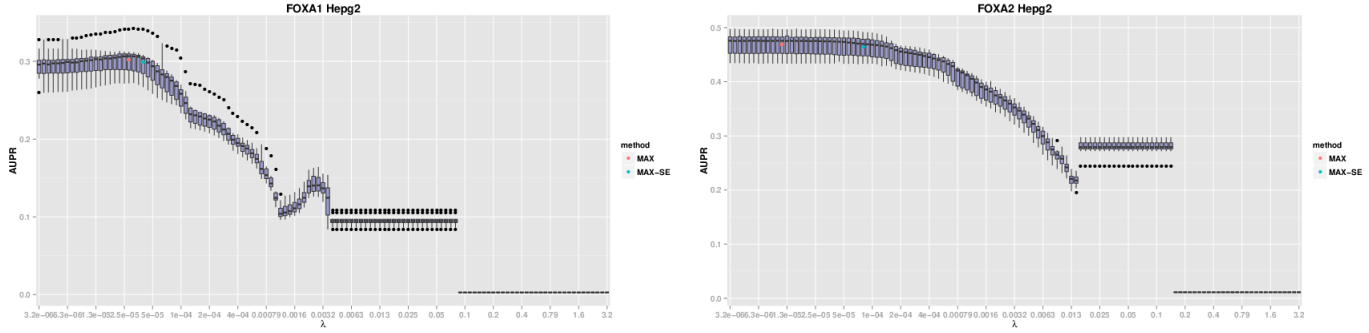
ELF1 K562
AUPR
λ
method
MAX
MAX−SE


EP300 Hepg2
AUPR
λ
method
MAX
MAX−SE

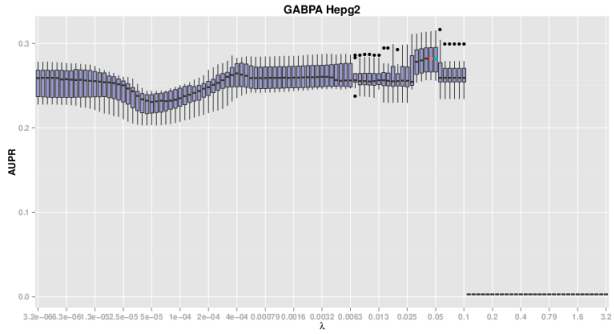
EP300 K562
AUPR
λ
method
MAX
MAX−SE

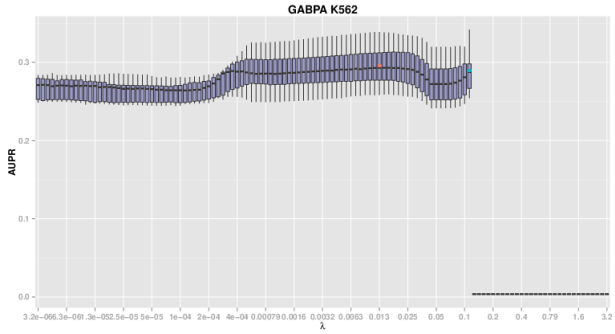

ETS1 K562
AUPR
λ
method
MAX
MAX−SE

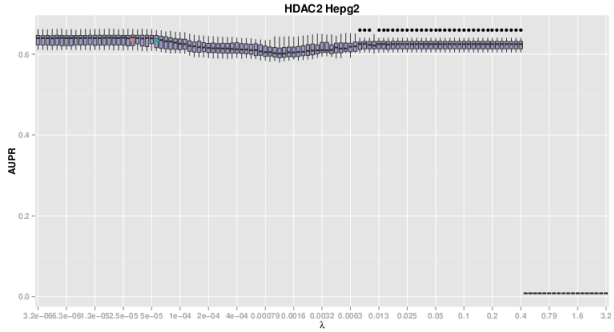
FOSL2 Hepg2
AUPR
λ
method
MAX
MAX−SE

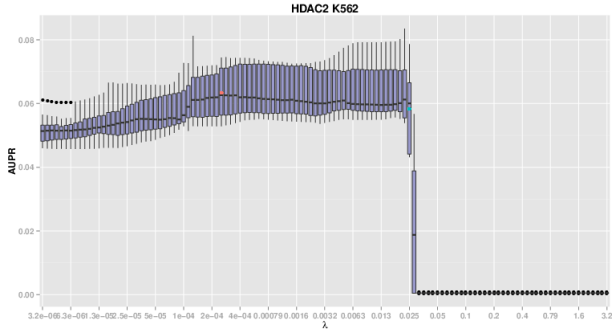

FOXA1 Hepg2
AUPR
λ
method
MAX
MAX−SE

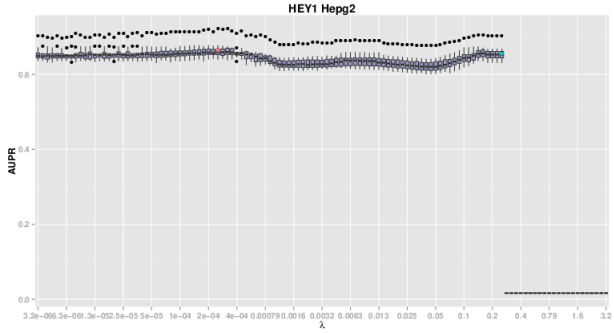
FOXA2 Hepg2
AUPR
λ
method
MAX
MAX−SE

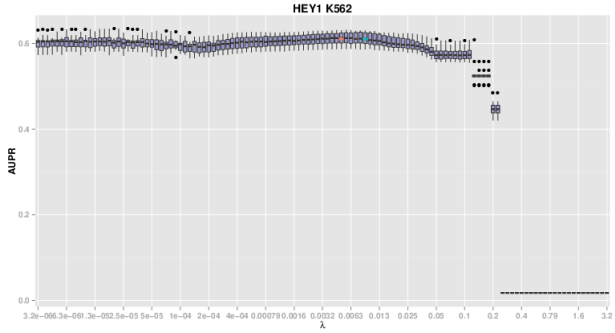

GABPA Hepg2
AUPR
λ
method
MAX
MAX−SE

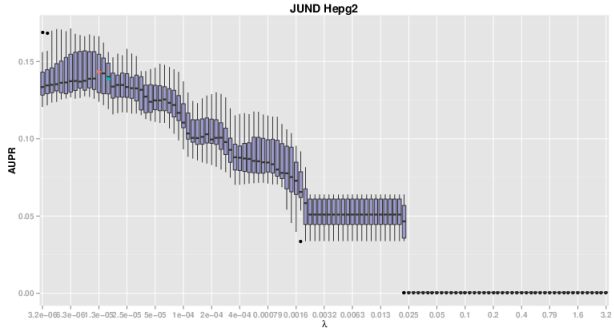

GABPA K562
AUPR
λ
method
MAX
MAX−SE

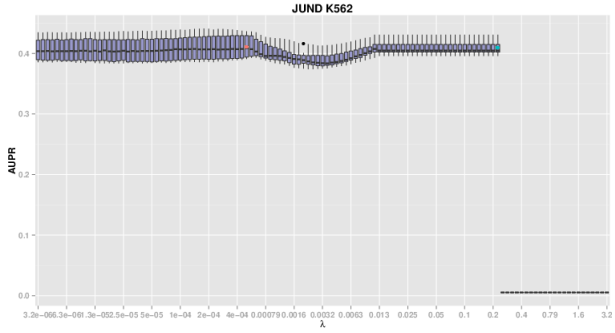

HDAC2 Hepg2
AUPR
λ
method
MAX
MAX−SE

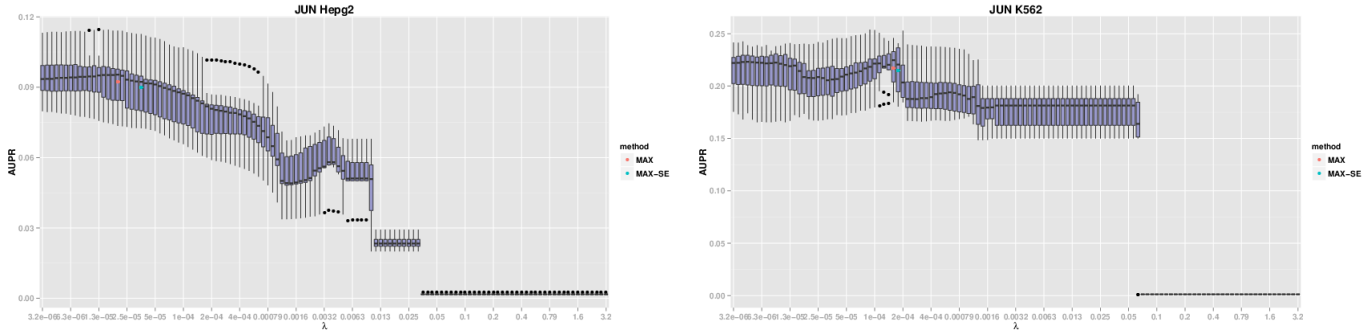

HDAC2 K562
AUPR
λ
method
MAX
MAX−SE


HEY1 Hepg2
AUPR
λ
method
MAX
MAX−SE

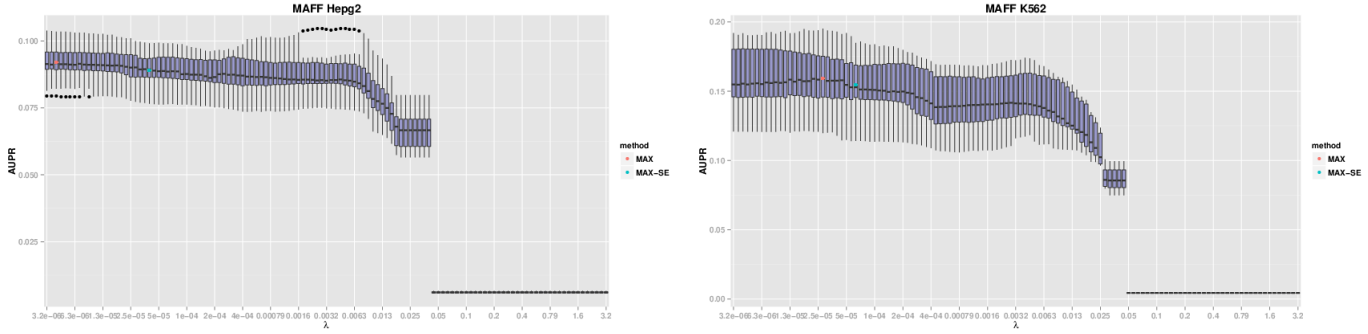

HEY1 K562
AUPR
λ
method
MAX
MAX−SE


JUND Hepg2
AUPR
λ
method
MAX
MAX−SE

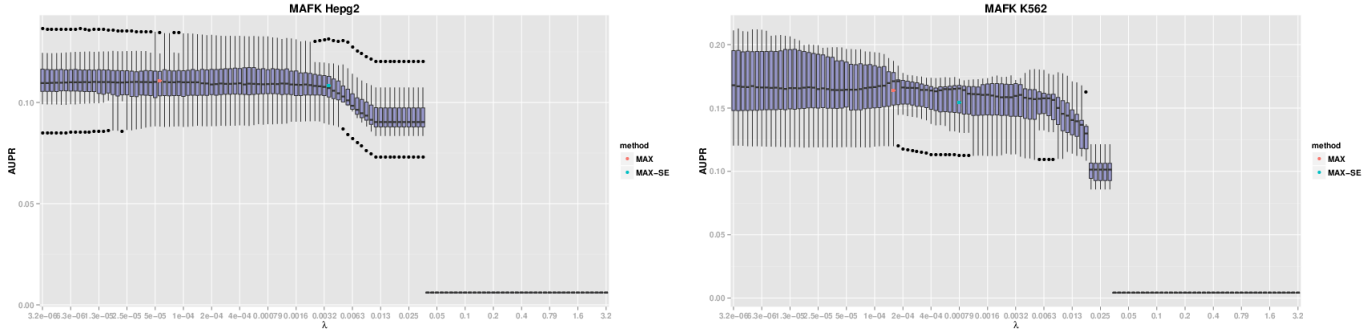

JUND K562
AUPR
λ
method
MAX
MAX−SE


JUN Hepg2
AUPR
λ
method
MAX
MAX−SE

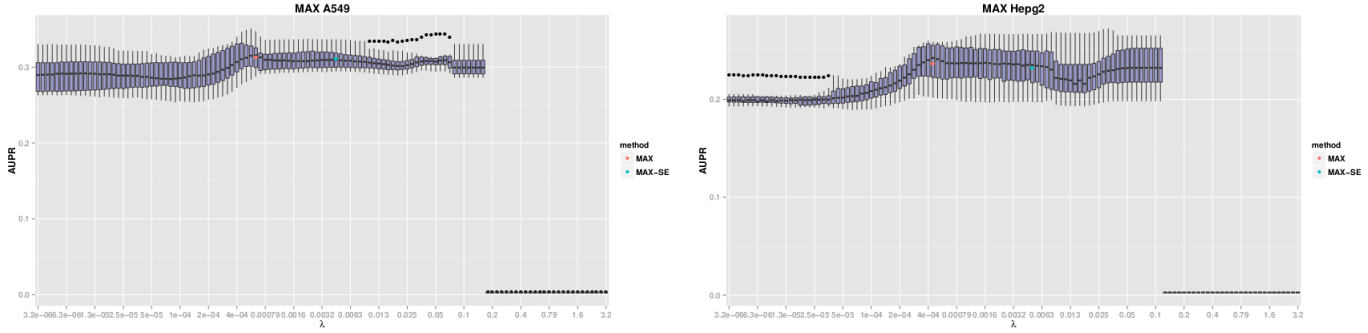
JUN K562
AUPR
λ
method
MAX
MAX−SE


MAFF Hepg2
AUPR
λ
method
MAX
MAX−SE

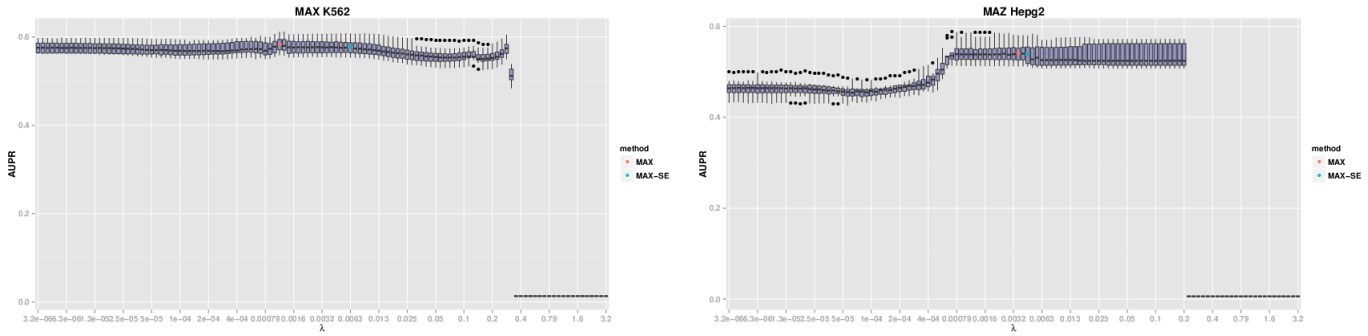
MAFF K562
AUPR
λ
method
MAX
MAX−SE

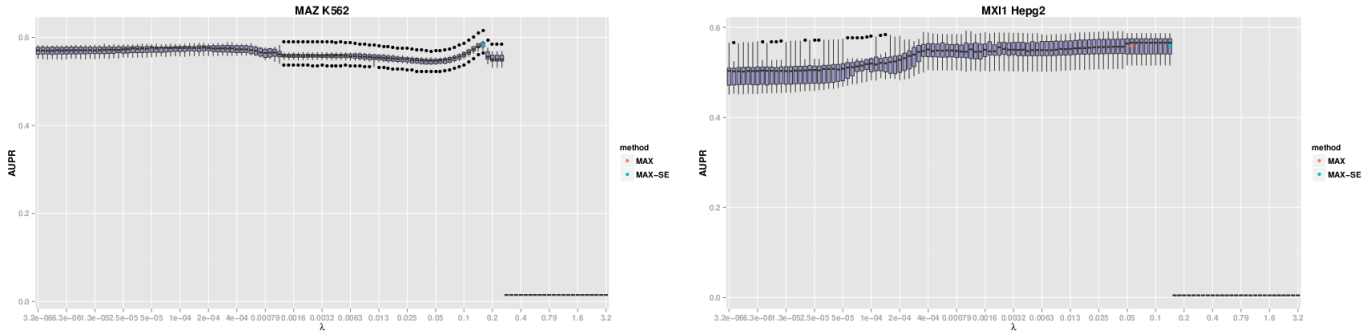

MAFK Hepg2
AUPR
λ
method
MAX
MAX−SE

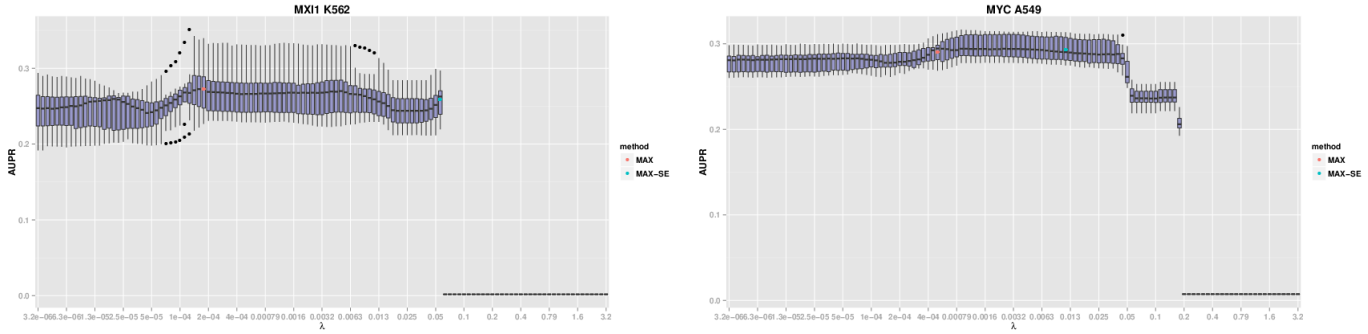
MAFK K562
AUPR
λ
method
MAX
MAX−SE

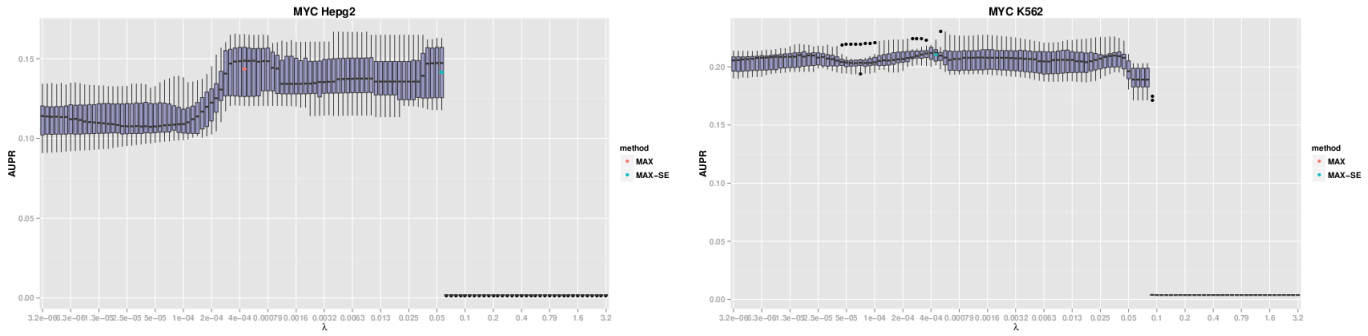

MAX A549
AUPR
λ
method
MAX
MAX−SE

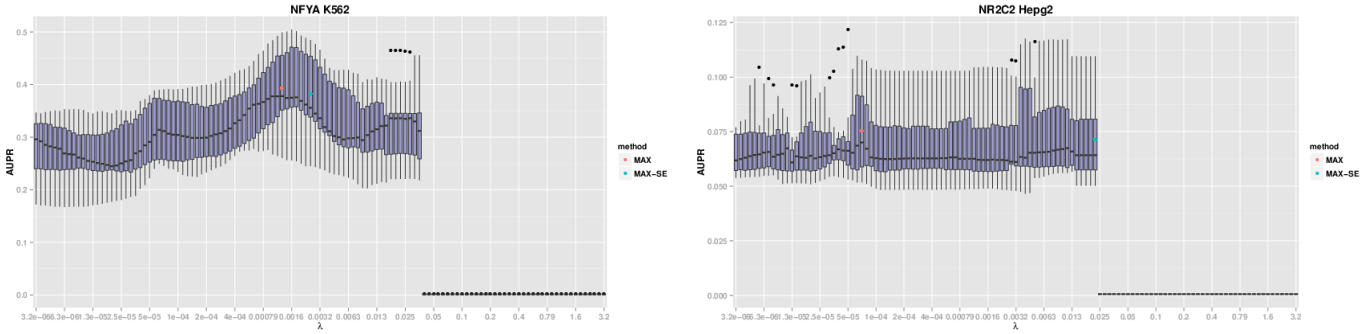
MAX Hepg2
AUPR
λ
method
MAX
MAX−SE


MAX K562
MAZ Hepg2
AUPR
λ
method
MAX
MAX−SE

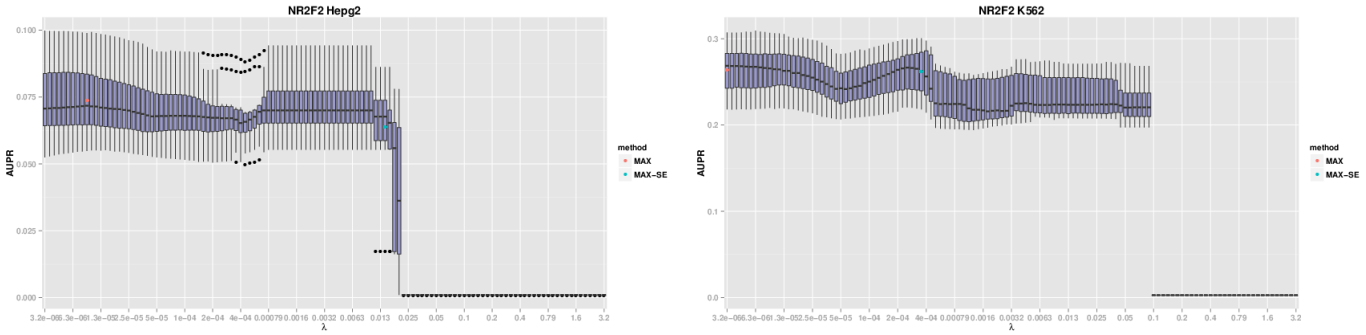

MAZ K562
MXI1 Hepg2
AUPR
λ
method
MAX
MAX−SE


MXI1 K562
MYC A549
AUPR
λ
method
MAX
MAX−SE

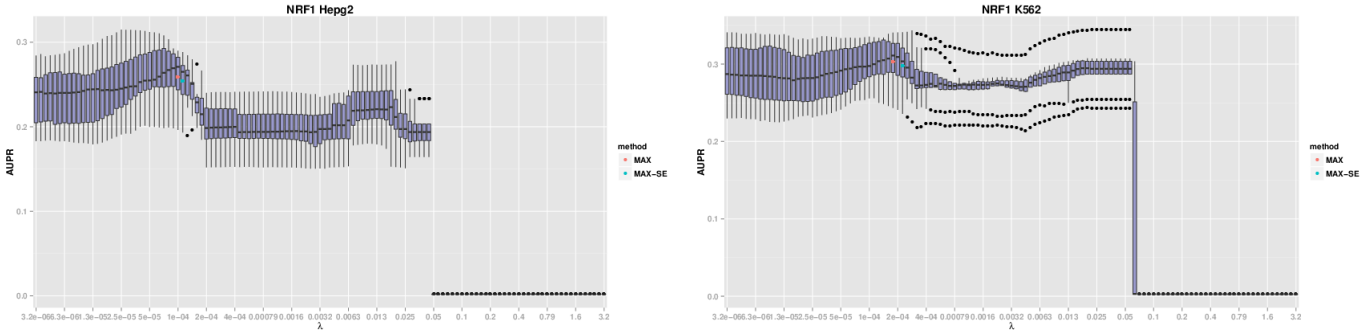

MYC Hepg2
MYC K562
AUPR
λ
method
MAX
MAX−SE


NFYA K562
AUPR
λ
method
MAX
MAX−SE

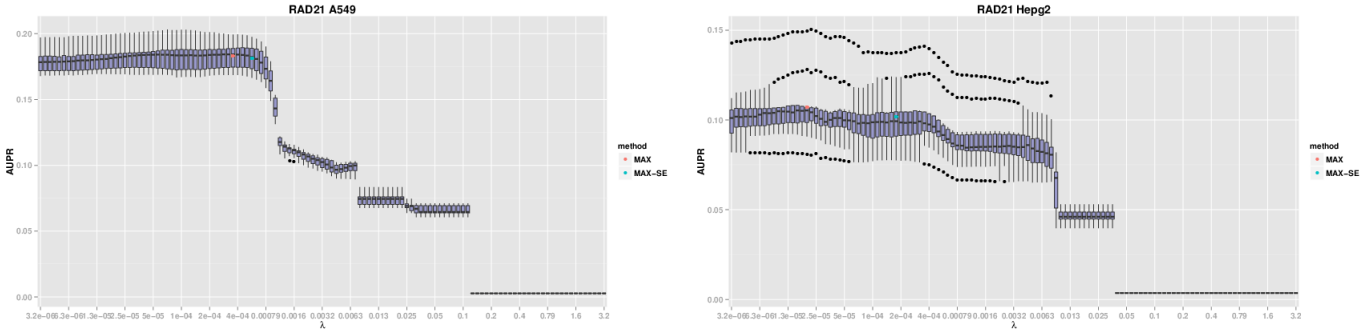
NR2C2 Hepg2
AUPR
λ
method
MAX
MAX−SE


NR2F2 Hepg2
AUPR
λ
method
MAX
MAX−SE

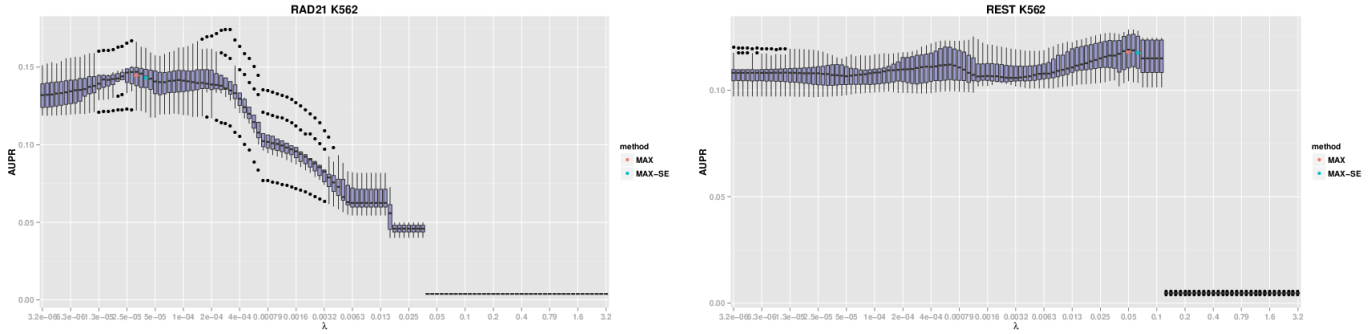
NR2F2 K562
AUPR
λ
method
MAX
MAX−SE

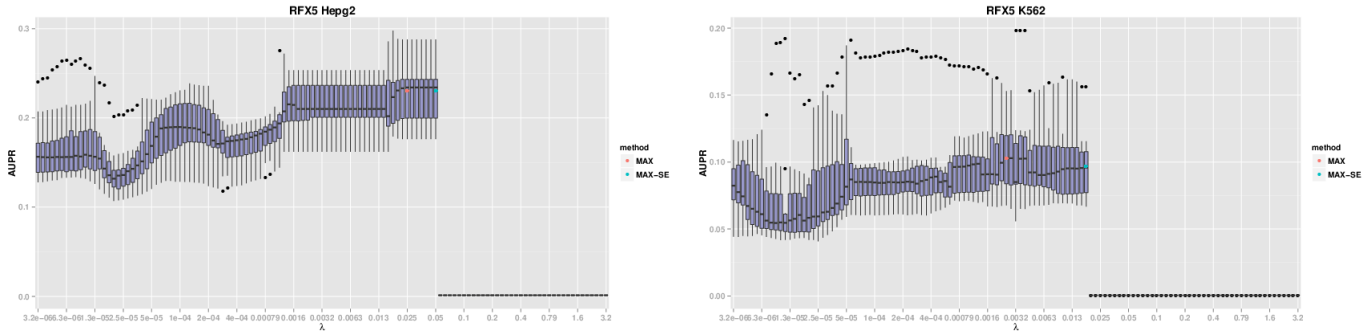

NRF1 Hepg2
AUPR
λ
method
MAX
MAX−SE

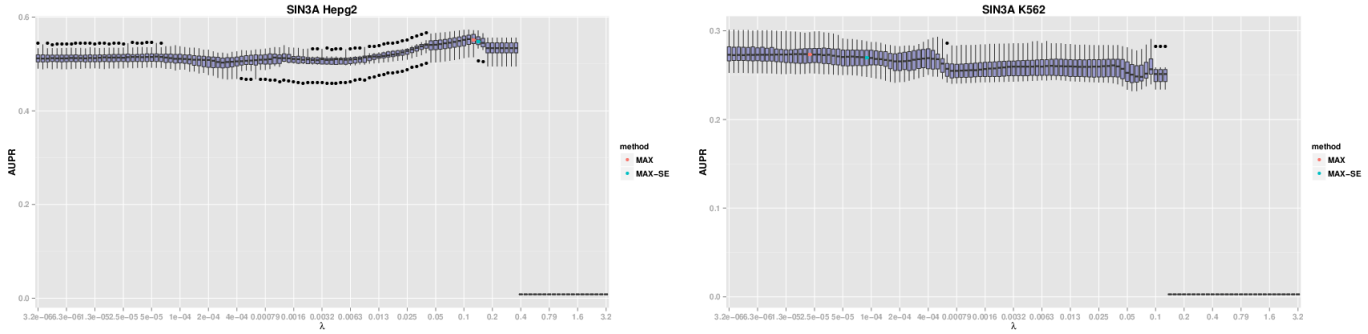
NRF1 K562
AUPR
λ
method
MAX
MAX−SE

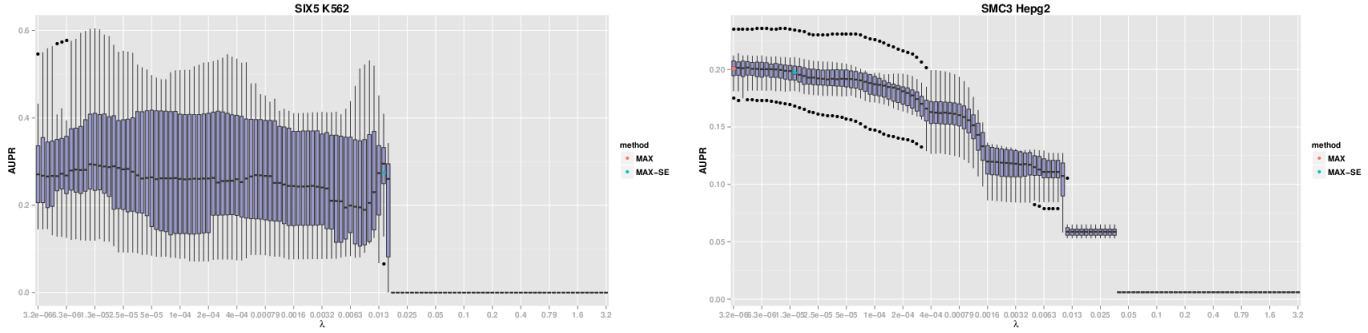

RAD21 A549
AUPR
λ
method
MAX
MAX−SE

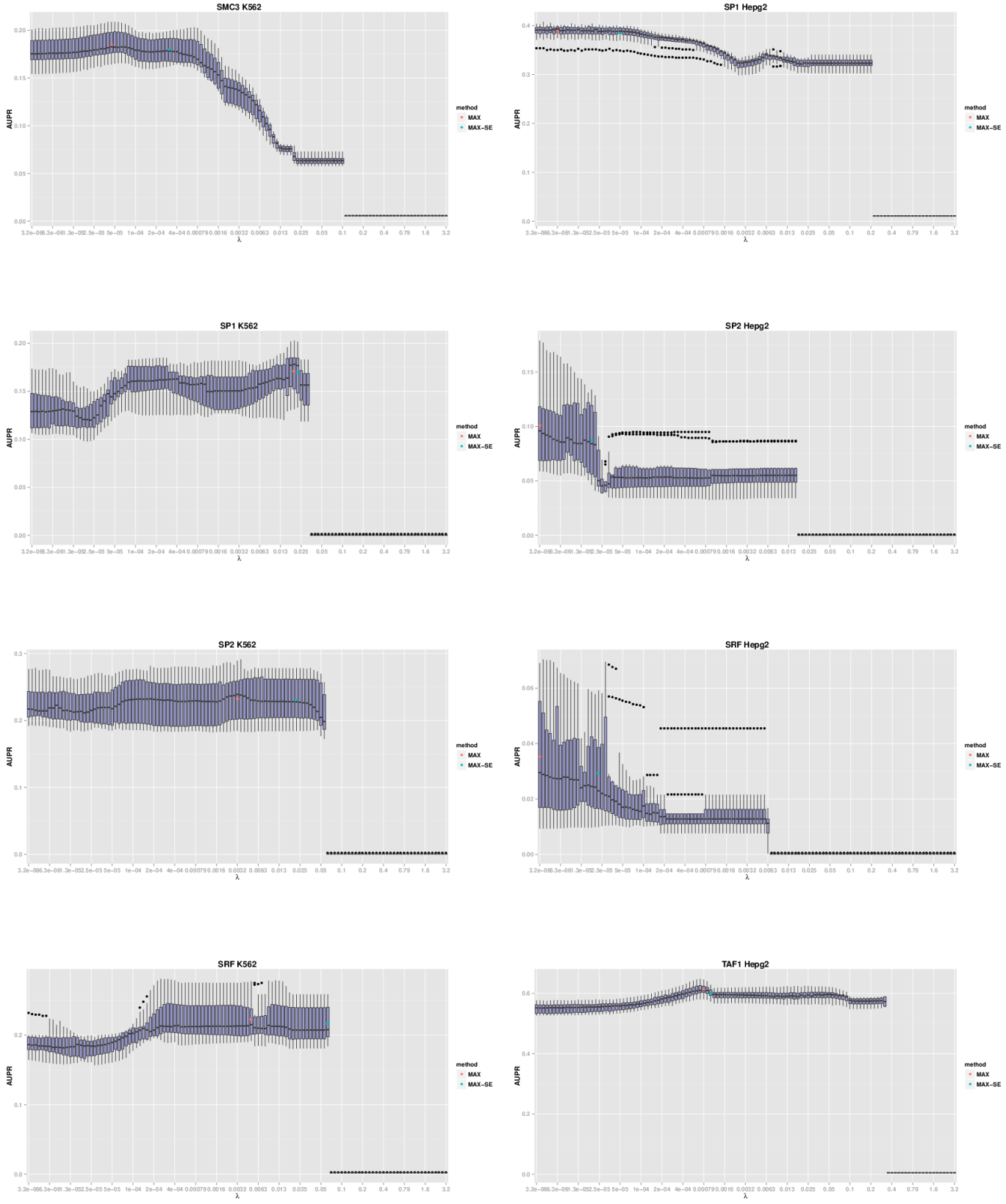
RAD21 Hepg2
AUPR
λ
method
MAX
MAX−SE

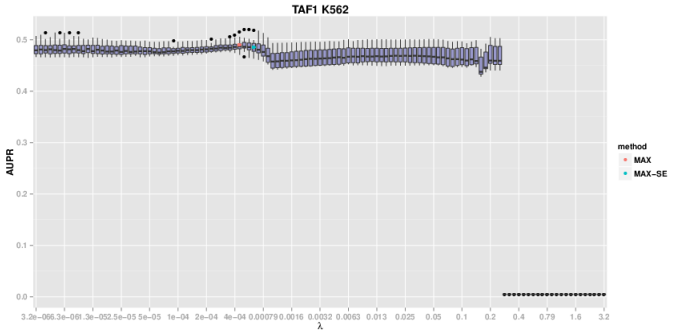

RAD21 K562
REST K562
AUPR
λ
method
MAX
MAX−SE

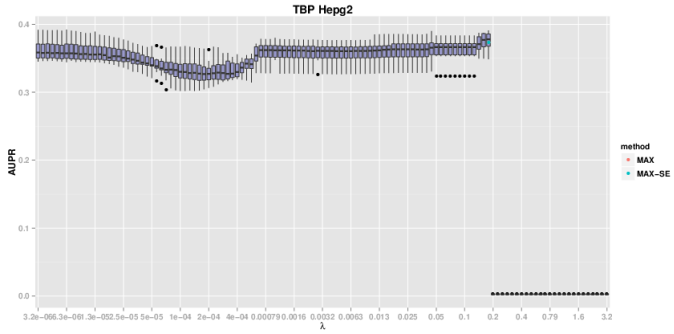

RFX5 Hepg2
RFX5 K562
AUPR
λ
method
MAX
MAX−SE

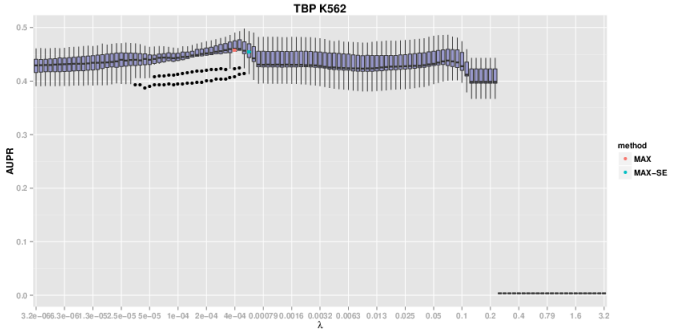

SIN3A Hepg2
SIN3A K562
AUPR
λ
method
MAX
MAX−SE

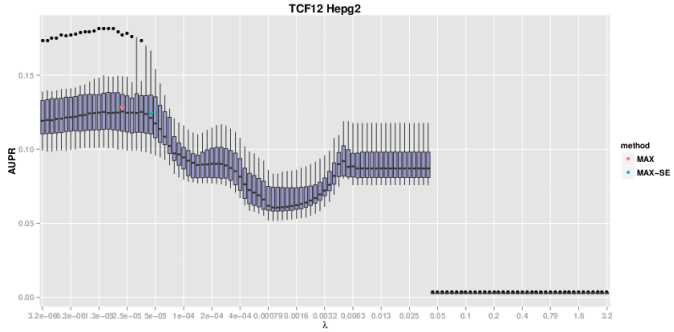

SIX5 K562
SMC3 Hepg2
AUPR
λ
method
MAX
MAX−SE

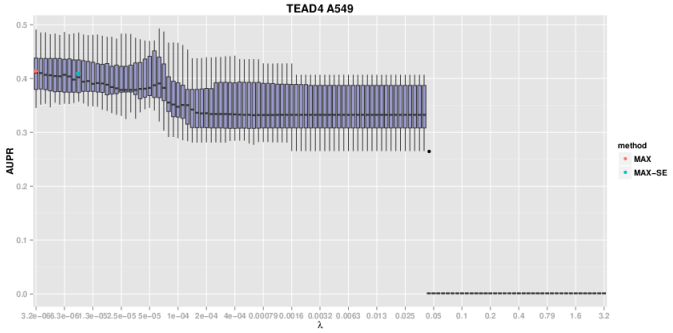

SMC3 K562
SP1 Hepg2
SP1 K562
SP2 Hepg2
SP2 K562
SRF Hepg2
SRF K562
TAF1 Hepg2
AUPR
λ
method
MAX
MAX−SE

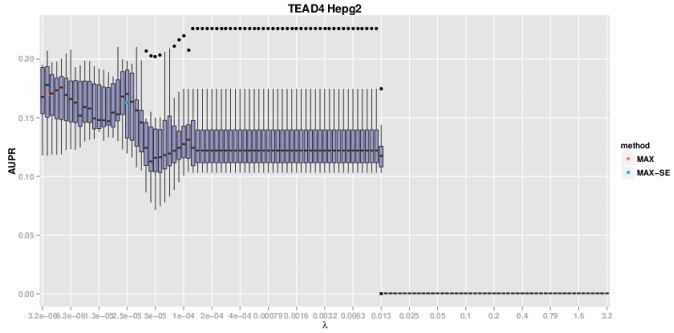

TAF1 K562
AUPR
λ
method
MAX
MAX−SE

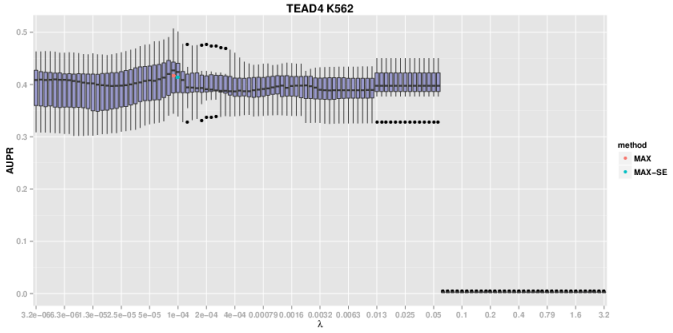

TBP Hepg2
AUPR
λ
method
MAX
MAX−SE

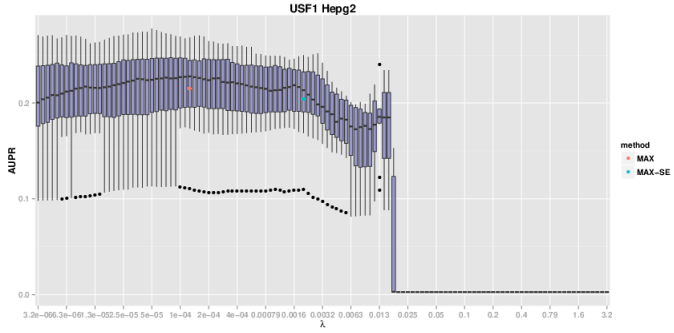

TBP K562
AUPR
λ
method
MAX
MAX−SE

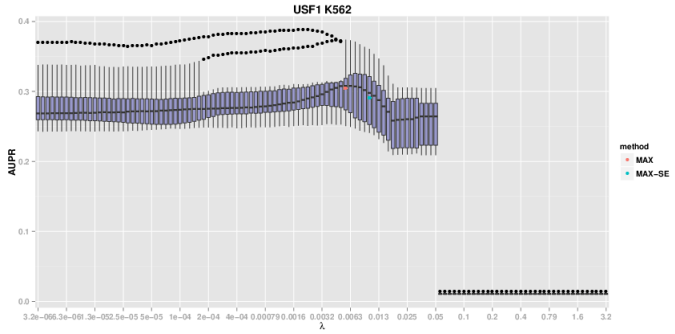

TCF12 Hepg2
AUPR
λ
method
MAX
MAX−SE

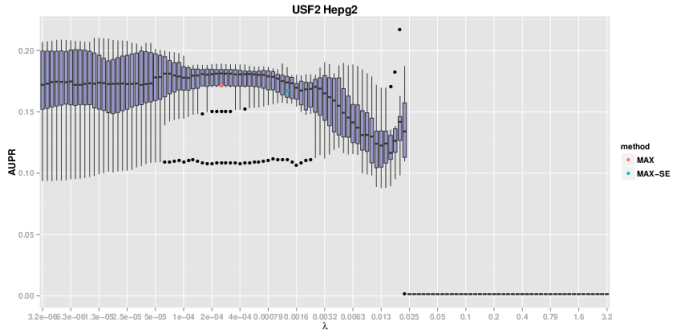

TEAD4 A549
AUPR
λ
method
MAX
MAX−SE

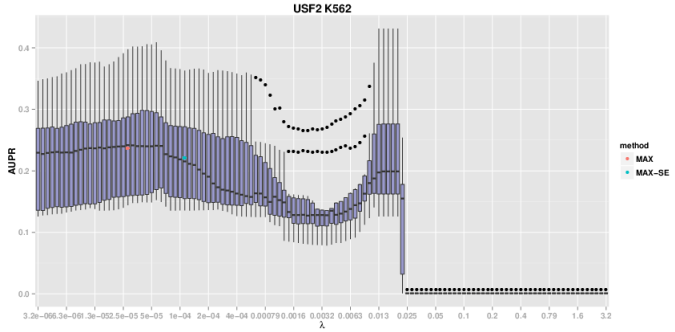

TEAD4 Hepg2
AUPR
λ
method
MAX
MAX−SE

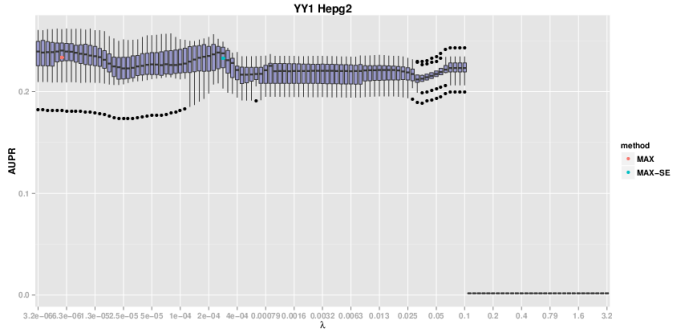

TEAD4 K562
AUPR
λ
method
MAX
MAX−SE

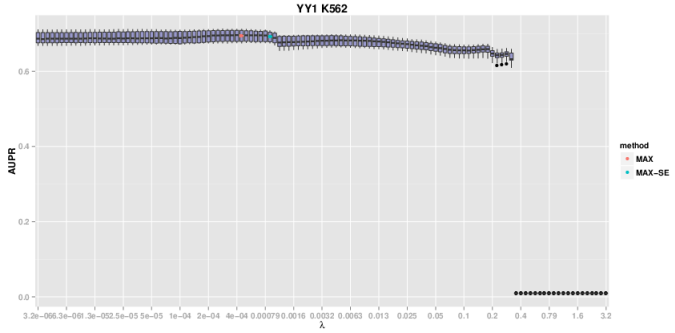

USF1 Hepg2
AUPR
λ
method
MAX
MAX−SE

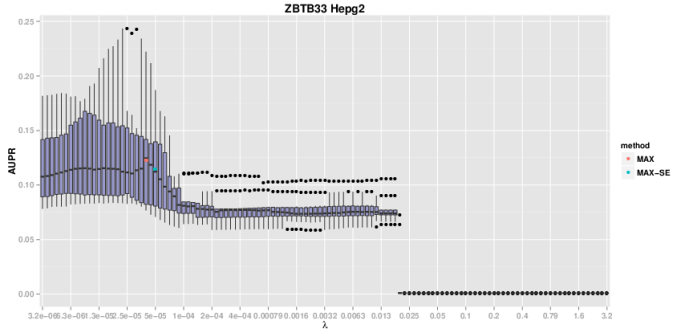

USF1 K562
AUPR
λ
method
MAX
MAX−SE

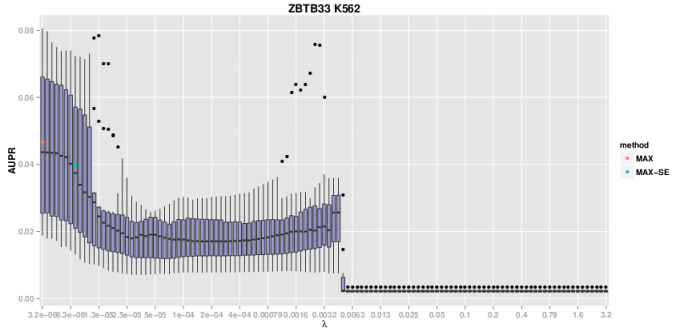

USF2 Hepg2
AUPR
λ
method
MAX
MAX−SE

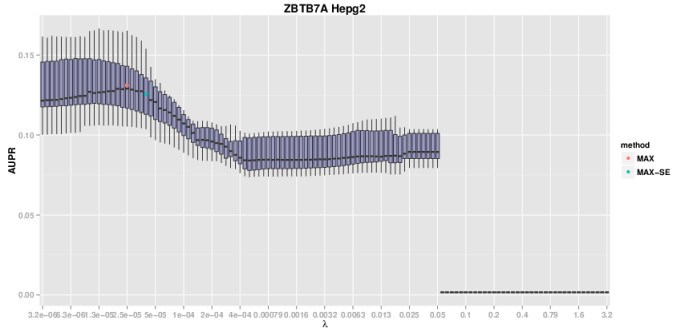

USF2 K562
AUPR
λ
method
MAX
MAX−SE

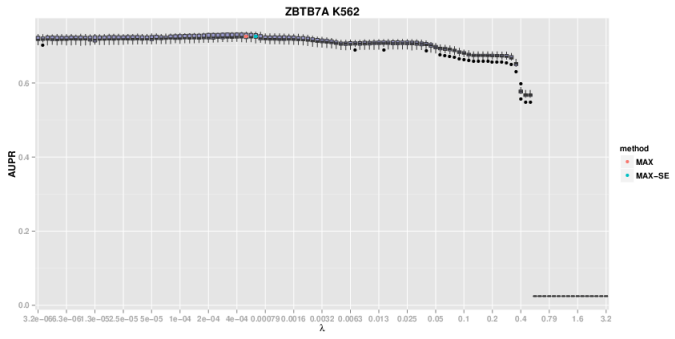

YY1 Hepg2
AUPR
λ
method
MAX
MAX−SE

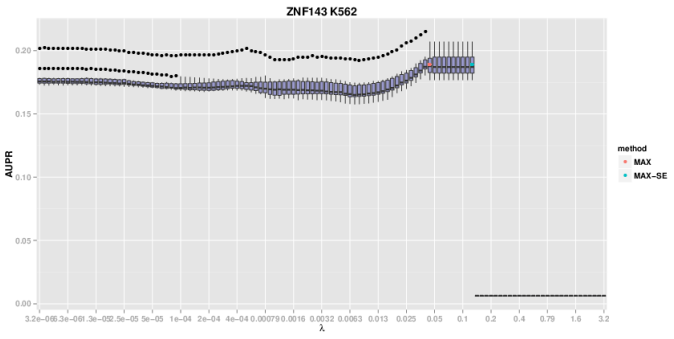

YY1 K562
AUPR
λ
method
MAX
MAX−SE

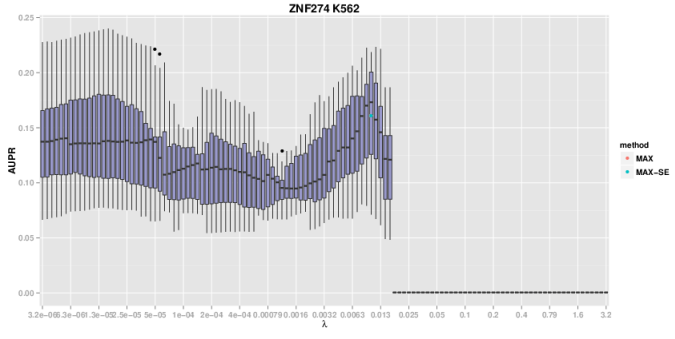

ZBTB33 Hepg2
AUPR
λ
method
MAX
MAX−SE

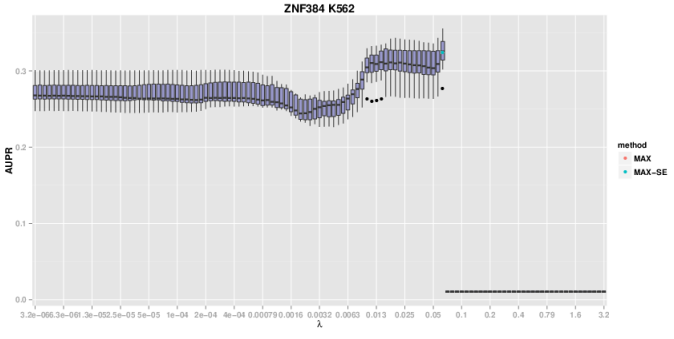

ZBTB33 K562
AUPR
λ
method
MAX
MAX−SE


ZBTB7A Hepg2
AUPR
λ
method
MAX
MAX−SE


ZBTB7A K562
AUPR
λ
method
MAX
MAX−SE


ZNF143 K562
AUPR
λ
method
MAX
MAX−SE


ZNF274 K562
AUPR
λ
method
MAX
MAX−SE


ZNF384 K562
AUPR
λ
method
MAX
MAX−SE