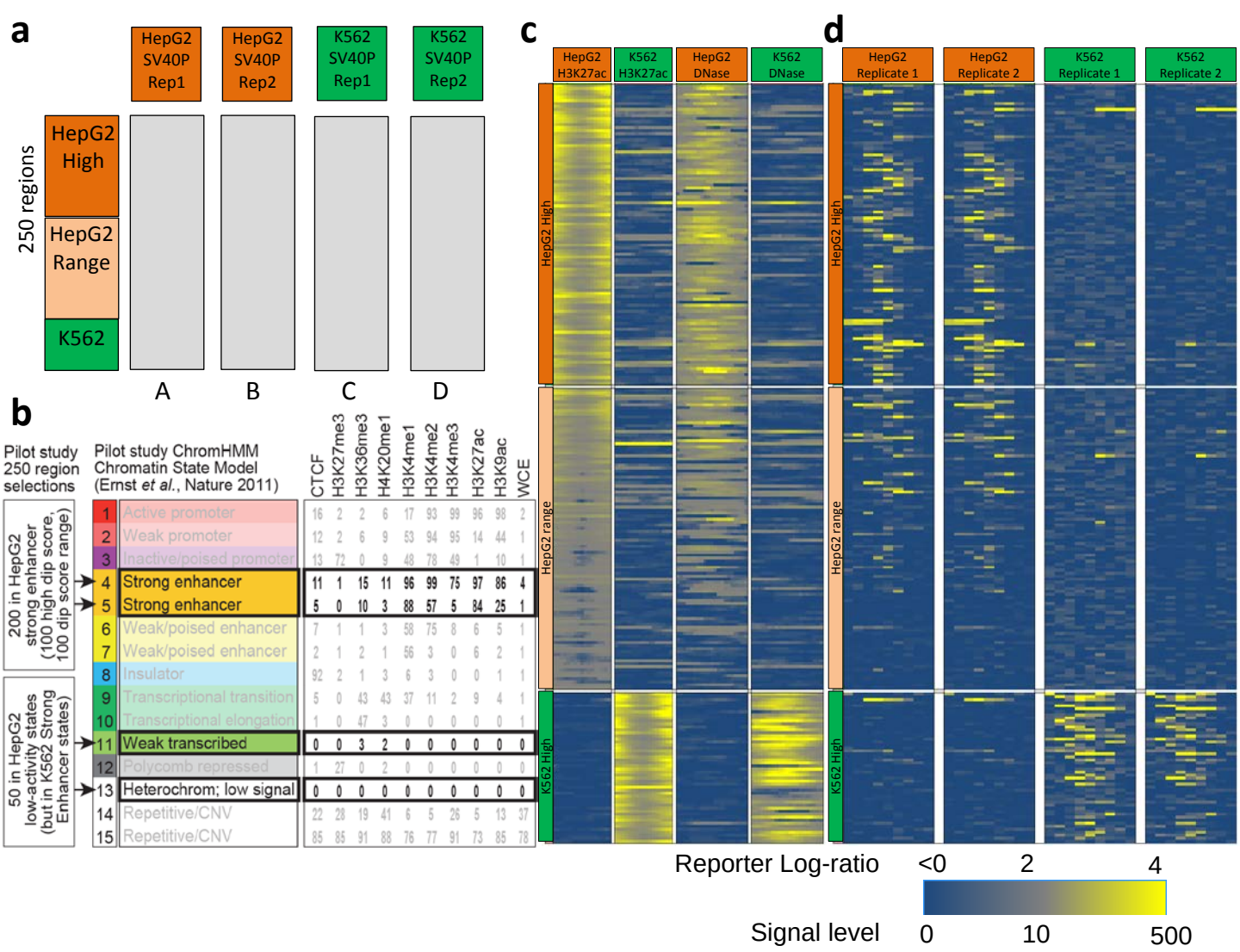
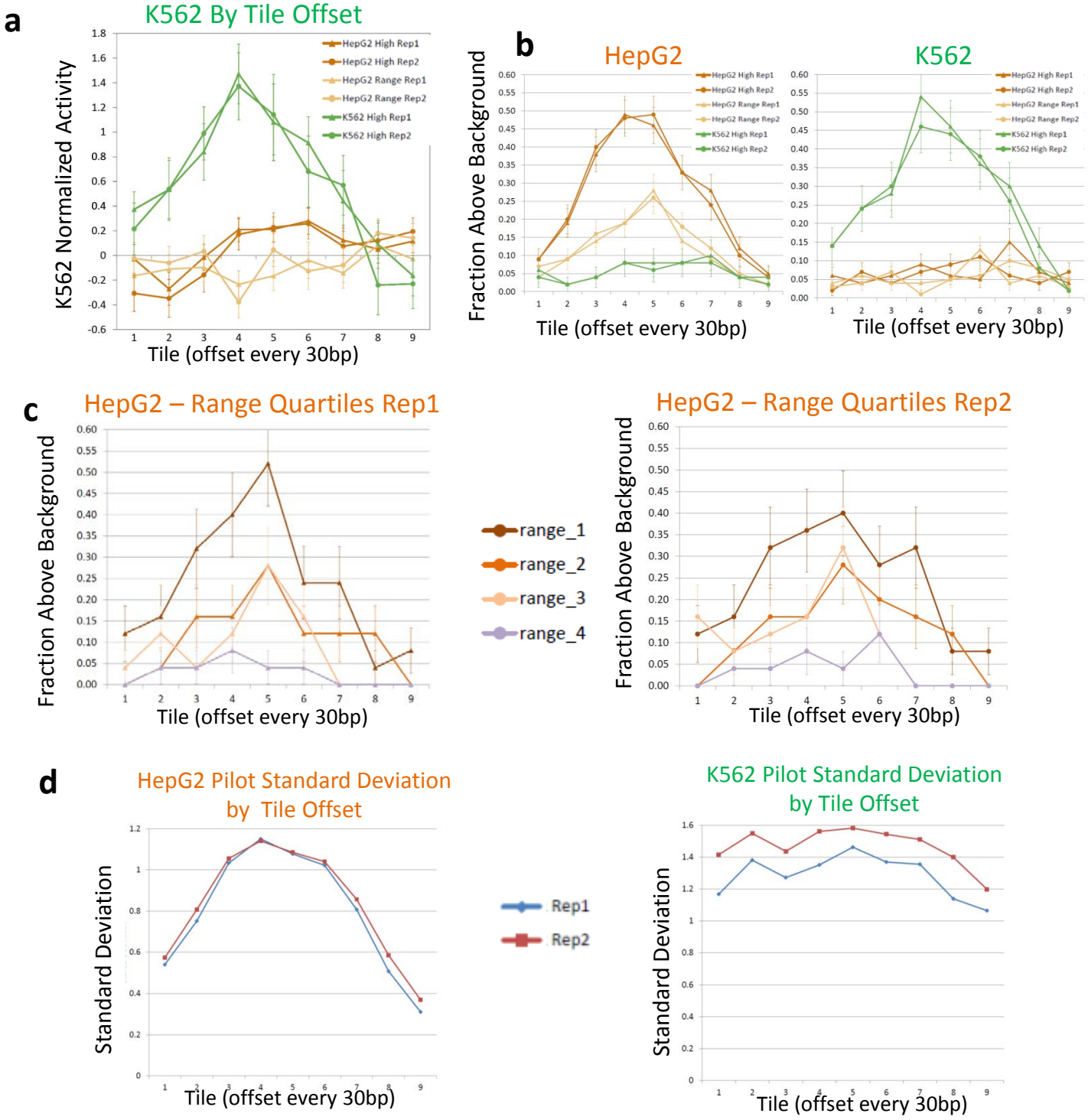


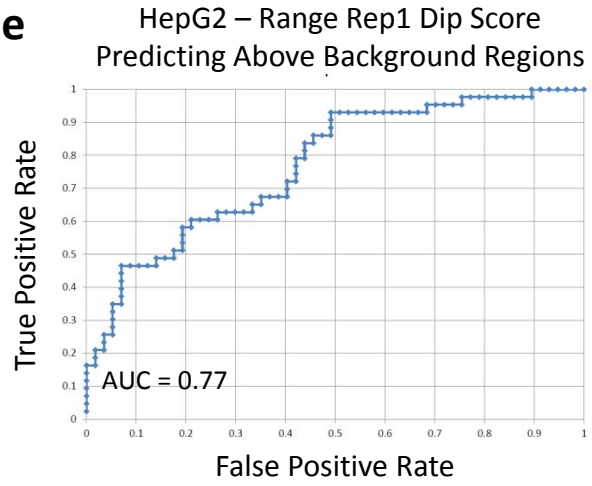
Supplementary Figure 1 – Large-step Pilot Design (a) Overview of the selection of regions and experiments. Two hundred fifty regulatory regions were selected to be tested with 100 being selected based on being in an HepG2 candidate enhancer state and having a high H3K27ac dip score, 100 being selected based on being in an HepG2 candidate enhancer state and covering a range of H3K27ac dip scores, and 50 for being in a K562 candidate enhancer state with a high H3K27ac dip score in K562 and in a low-activity state in HepG2 (see Methods). These regulatory regions were tested in both K562 and HepG2 using a SV40 promoter in replicate. **(b)** Chromatin state model used for selecting strong enhancer regions in HepG2 and K562, and low-activity regions in K562[2]. The frequency of each chromatin state mark is shown (scaled to be between 0 and 100). Note that for the scale-up model, we used a richer 25-state model to capture a higher diversity of chromatin states (shown in Supplementary Fig. 5). **(c)** The first two larger columns are the H3K27ac signal[2] in HepG2 and K562 cells respectively, while the next two larger columns are the DNase signal[7] in HepG2 and K562 cells respectively. Each larger column is separated by the white lines. Within each larger column is a heatmap of the signal across the 385-bp region based on the color scale shown at bottom. The first larger set of rows (separated by horizontal white lines) are those regions corresponding to the HepG2 tiled enhancer regions with a high dip score, the second are those regions corresponding to the HepG2 tiled enhancer regions with a range of dip scores, and the third set are the K562 based tiled enhancer regions. Within each set the regions are ordered in terms of decreasing dip score. **(d)** Heatmap of the reporter expression values. The first two larger columns show the results for the experiments in two HepG2 replicates and the next two columns show the experiments in the two K562 replicates. Within each larger column is the median log-base two ratio reporter expression values at each of the nine tile offsets normalized by taking the difference with the average value for the expression of the tiles in the outmost tile offsets (tiles #1 and #9) colored based on the indicated color scale shown at the bottom.

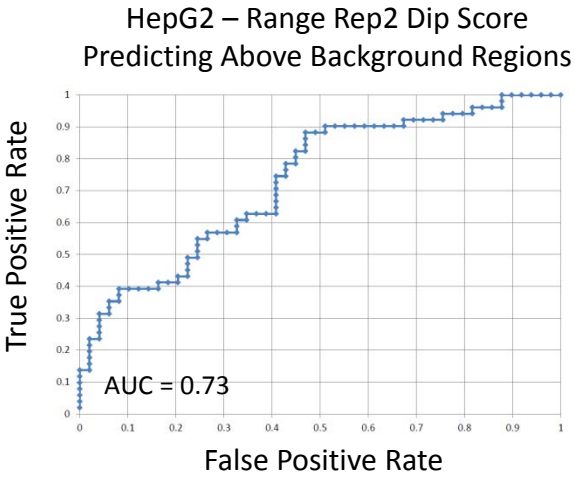


Supplementary Figure 2 – Large-step Pilot Design Analysis. (a) The same plot on average reporter expression by tile offset and group of tiled regions as in **Fig. 2a** but for the experiments conducted in K562 cells. **(b)** The fraction of reporter values that met the threshold at which only 5% of outmost tile offsets (tiles #1 and #9) reporter values did for (left) HepG2 and (right) K562 cells experiments. **(c)** The same as (b) except only showing the range values broken down into four quartiles and showing in separate graphs replicate 1 (left) and replicate 2 (right). **(d)** Standard deviation (y-axis) of the reporter expression values at each tile offset (x-axis) for HepG2 experiments (left) and K562 experiments (right) for each replicate (color).

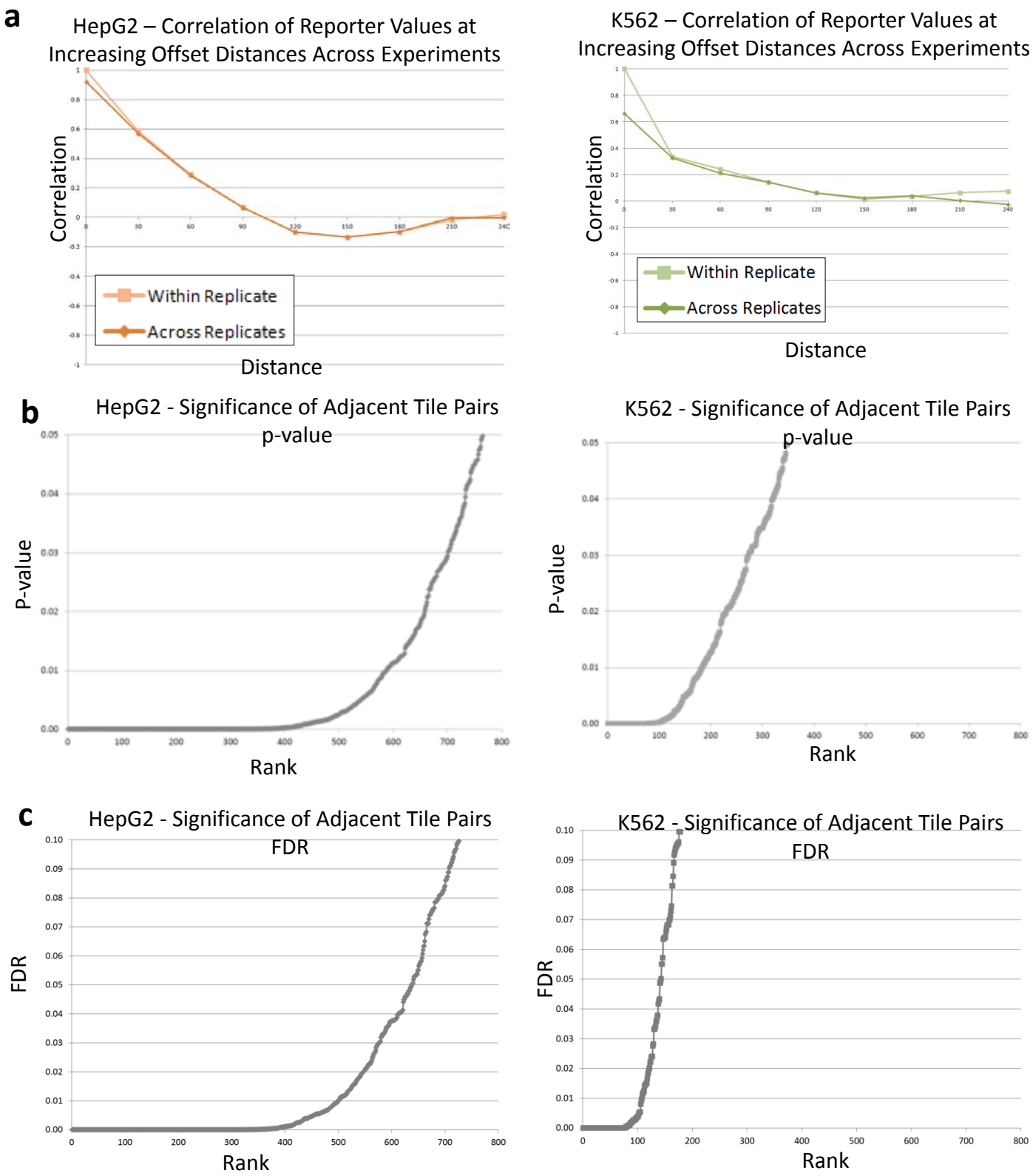
[Supplementary Figure 2 panel e is continued on next page]

e



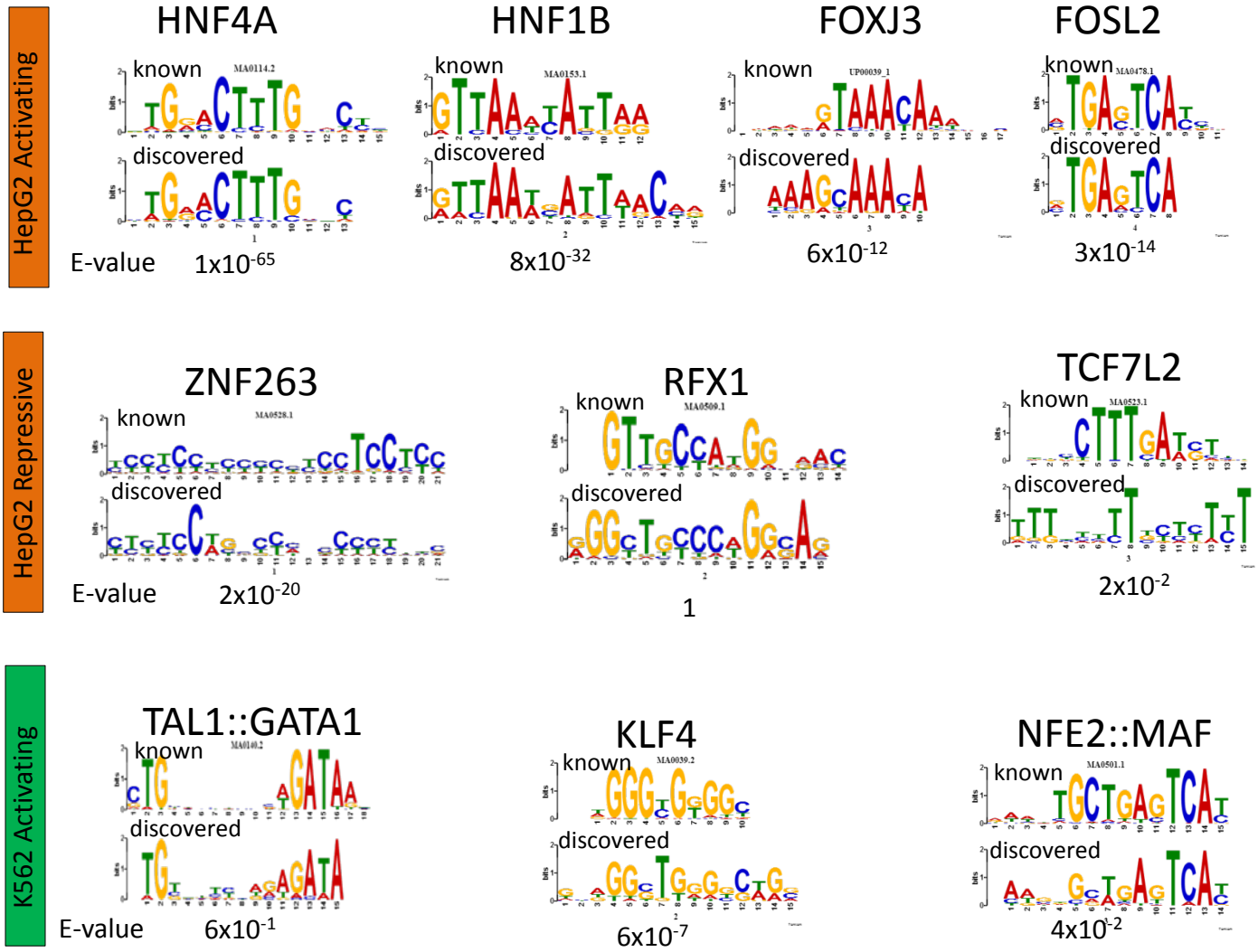


Supplementary Figure 2 continued from previous page: (e) For the 100 HepG2 tiled regions selected to span a range of H3K27ac dip scores, ROC curve for predicting regions with at least one tile above threshold, when ranking based on the dip score for the region for replicate 1 (left) and replicate 2 (right) experiments.



Supplementary Figure 3 – Large-step Pilot Correlation of Reporter Values as a Function of Distance. **(a)** Correlation of reporter expression values (y-axis) for each indicated offset (x-axis) both within (boxes) and across replicates (diamonds) in HepG2 (left) and K562 (right). **(b)** Cumulative number of adjacent Large-step pairs of tiles (x-axis) that showed a difference in expression at a given uncorrected p-value (y-axis) based on a Mann-Whitney Test (see Methods) in HepG2 (left) experiments and K562 (right) experiments. **(c)** The FDR values corresponding to the p-values shown in b.

a



b

	HepG2 activating	HepG2 repressive	K562 activating	K562 repressive
GATA_known14	1.0	1.0	6.4	1.0
LMO2_2	1.0	1.0	4.4	1.0
TAL1_disc1	1.0	1.0	3.0	1.0
CPHX_1	1.0	1.0	2.7	1.0
JDP2_2	1.0	1.0	2.4	1.0
NFE2L2_3	1.0	1.0	2.0	1.0
HNF4_known9	4.6	0.3	1.0	1.0
NR2F6_2	3.7	1.0	1.0	1.0
HNF1B_4	3.5	0.6	1.0	1.0
RXRA_known10	3.4	0.9	1.0	1.0
PPARA_4	3.1	1.0	1.0	1.0
HNF1A_4	2.8	1.0	1.0	1.0
HNF1_4	2.6	1.0	1.0	1.0
TCF7L2_disc1	2.4	1.0	1.0	1.0
AP1_known4	2.3	1.0	1.6	1.0
SMARC_disc1	2.2	1.0	1.1	1.0
CEBPB_known1	2.1	1.0	1.0	1.0
TLX2_2	2.0	1.0	1.0	1.0
FOXJ2_4	2.0	1.0	1.0	1.0
EGR1_disc2	2.0	1.0	1.0	1.0
RFX2_3	1.0	3.0	1.0	1.0
RFX5_known9	1.0	2.9	1.0	1.0

Supplementary Figure 4 – Large-step Pilot Design Motifs. (a) Motifs discovered based on running MEME software[72] on the identified at a 5% false discovery rate (top) HepG2 activating, (center) HepG2 repressive, and (bottom) K562 activating 30-bp sequences extended 10-bp to include the common sequence (see Methods). Only up to the top four discovered motifs that had a reported E-value of 1 or less are shown. Motifs are ordered left to right as returned by the MEME program. There were no motifs discovered for the K562 negative set that met the criterion. Each discovered motif is matched to a best matching known motif using TOMTOM[73] that is shown above the discovered motif. **(b)** Motifs (rows) enriched in at least one of the four sets (HepG2 Activating, HepG2 Repressive, K562 Activating, and K562 Repressive) at least 2 fold based on the program of Ref [13] extended to use a background of all nucleotides tested. The columns show the enrichment of the motif in the four sets considered using the same program and background. Only motifs with at least three instances in the background are included and if multiple motifs corresponding to the same transcription factor met the threshold only the one with the highest enrichment is listed.

a

Emission Probability for each mark in each state (Percent)

State	H3K36me3	H4K20me1	Control	FAIRE	POL2	H3K27ac	H3K9ac	H3K4me3	H3K4me2	H3K4me1	DNase_UW	DNase_DUKE	CTCF	H3K27me3
1_Tss	1.8	6.4	3.0	33.1	98.5	96.1	99.0	99.9	95.9	43.0	85.9	92.2	36.1	4.9
2_TssF	21.4	23.9	0.7	4.4	55.1	72.4	86.5	97.8	99.4	73.3	24.4	15.4	3.9	1.7
3_PromF	3.3	6.2	0.2	0.6	7.5	2.3	8.2	32.2	89.1	43.8	4.8	4.2	1.7	2.2
4_PromP	0.1	7.1	0.5	7.7	38.4	2.2	37.3	88.0	96.7	61.5	51.3	86.9	31.7	42.1
5_Enh	11.4	7.4	4.0	49.5	68.9	98.4	72.7	33.9	97.1	98.2	89.0	67.3	17.2	0.1
6_EnhF	5.0	2.3	0.6	3.7	10.8	94.3	28.7	1.0	53.9	93.4	20.0	8.1	2.3	0.1
7_EnhWF	0.9	2.2	0.2	0.8	1.4	0.6	1.3	0.0	7.0	88.4	4.4	3.7	1.4	0.4
8_EnhW	0.2	8.3	0.7	16.6	18.9	10.8	6.3	0.6	61.6	93.6	62.6	59.9	17.0	0.7
9_DNaseU	0.4	1.4	0.1	2.8	3.0	1.6	0.5	0.1	2.7	9.1	95.1	13.6	0.3	0.3
10_DNaseD	1.1	3.5	0.2	0.3	1.6	0.5	0.7	0.1	0.3	3.1	2.5	69.5	0.9	0.3
11_FaireW	1.5	1.9	7.9	22.5	0.8	10.1	5.6	1.7	0.2	0.3	0.5	1.5	0.6	2.0
12_CtcfO	5.9	8.8	1.0	34.6	18.6	1.8	1.4	0.3	10.3	16.7	71.4	86.9	99.8	2.0
13_Ctcf	0.3	2.1	0.1	0.5	0.3	0.1	0.2	0.0	0.7	1.4	6.0	16.1	99.1	1.7
14_Gen5'	70.4	35.9	0.7	3.2	52.3	26.2	9.1	1.2	23.4	83.6	9.5	11.6	3.6	0.0
15_Elon	93.2	4.2	0.3	0.7	1.4	0.3	0.2	0.0	0.1	0.3	0.4	1.4	1.0	0.0
16_ElonW	0.0	0.4	0.0	0.3	0.8	0.1	0.0	0.0	0.0	0.1	0.1	0.2	0.0	0.0
17_Gen3'	93.1	17.6	0.4	1.2	83.5	1.2	0.6	0.1	0.1	0.7	2.3	7.5	2.8	0.0
18_Pol2	0.7	5.6	0.1	0.7	82.6	0.9	0.6	0.0	0.1	2.2	2.7	7.1	1.2	0.3
19_H4K20	3.4	85.7	0.3	1.0	2.1	0.2	0.3	0.0	0.2	2.1	0.4	1.4	0.5	0.6
20_ReprD	0.5	15.5	0.5	1.4	2.0	0.5	0.6	1.0	11.4	20.7	4.4	47.9	11.7	90.4
21_Repr	0.0	2.1	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	1.1	0.1	89.8
22_ReprW	0.0	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0
23_Low	0.1	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0
24_Quies	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25_Art	77.0	73.0	40.8	70.2	59.2	34.7	38.3	25.3	31.8	52.4	40.7	63.5	64.5	47.5

b

Genome Coverage and Annotation Fold Enrichments (median across cell types)

State	mean Genome %	TSS (+/-2KB 5' End)	Conserved Element (pi)	TF binding sites	Nuclear lamina
1_Tss	0.75	25.8	4.4	50.3	0.2
2_TssF	0.60	17.5	2.0	2.5	0.3
3_PromF	0.60	9.6	1.5	1.3	0.5
4_PromP	0.17	18.3	4.0	17.4	0.4
5_Enh	0.44	1.3	2.3	47.0	0.4
6_EnhF	0.66	1.3	1.3	3.0	0.5
7_EnhWF	1.02	1.2	1.4	1.3	0.5
8_EnhW	0.42	1.4	2.0	22.4	0.5
9_DNaseU	0.32	0.8	1.6	9.4	0.8
10_DNaseD	0.72	3.4	1.7	1.0	0.3
11_FaireW	0.71	1.2	0.7	0.4	0.8
12_CtcfO	0.26	1.7	2.2	24.4	0.6
13_Ctcf	0.75	1.1	1.4	4.3	1.0
14_Gen5'	0.41	1.4	1.9	1.7	0.2
15_Elon	2.75	0.3	2.0	0.1	0.3
16_ElonW	2.65	0.3	1.2	0.1	0.3
17_Gen3'	1.04	0.7	2.5	0.3	0.1
18_Pol2	0.74	2.7	1.3	0.4	0.2
19_H4K20	1.56	0.6	1.0	0.2	0.4
20_ReprD	0.22	7.4	2.3	1.3	0.5
21_Repr	1.62	2.2	1.4	0.1	0.9
22_ReprW	6.36	1.0	1.1	0.1	1.0
23_Low	27.15	0.9	0.9	0.1	0.7
24_Quies	48.04	0.3	0.7	0.0	1.3
25_Art	0.05	1.3	1.3	9.0	0.2

c

Candidate State Annotations

State	Candidate State Annotations
1_Tss	Active promoter, TSS/CpG island region
2_TssF	Active promoter, flanking TSS/CpG islands
3_PromF	Promoter flanking
4_PromP	Inactive/Poised promoter, highly conserved
5_Enh	Candidate Strong Enhancer, open chromatin
6_EnhF	Candidate Strong Enhancer, flanking open chromatin
7_EnhWF	Candidate poised/weak enhancer; flanking open chromatin of candidate enhancers
8_EnhW	Candidate weak enhancer and open chromatin
9_DNaseU	Primarily UW DNase, weaker open chromatin sites
10_DNaseD	Primarily Duke DNase, candidate regulatory elements in more likely repressive locations
11_FaireW	Modest Faire/Control enrichments, potential CNV
12_CtcfO	Distal CTCF/Candidate Insulator with open chromatin
13_Ctcf	Distal CTCF/Candidate Insulator without open chromatin
14_Gen5'	Transcription transition, highly expressed genes towards 5' end
15_Elon	Transcriptional Elongation, stronger H3K36me3, more exonic
16_ElonW	Transcriptional Elongation, weaker H3K36me3
17_Gen3'	Transcription 3' end of genes, highly expressed; more exonic
18_Pol2	Pol2 specific locations, majority in genes but substantial portion in intergenic locations
19_H4K20	Transcription, primarily H4K20me1, more intronic
20_ReprD	Polycomb Repression w. Duke DNase/promoter & conservation enriched (except HeLa)
21_Repr	Strong Polycomb Repression
22_ReprW	Weaker Polycomb Repression
23_Low	Low signal proximal to active elements
24_Quies	Heterochromatin/Dead Zone
25_Art	Potential CNV or repetitive artifacts

d

Percent Genome Coverage

State	HepG2	K562	Huvec	H1-hESC	GM12878	HeLa-s3
1_Tss	0.75	0.61	0.62	0.46	0.77	0.74
2_TssF	0.64	0.58	0.35	0.22	0.77	0.44
3_PromF	0.70	0.32	0.29	0.46	0.74	0.47
4_PromP	0.57	0.17	0.28	0.72	0.21	0.12
5_Enh	0.28	0.50	0.85	0.14	0.45	0.43
6_EnhF	0.52	0.55	0.94	0.27	0.70	0.61
7_EnhWF	1.23	2.09	1.27	1.62	1.19	0.84
8_EnhW	0.36	0.53	0.47	0.62	0.38	0.46
9_DNaseU	0.10	0.20	0.20	0.64	0.26	0.38
10_DNaseD	0.68	0.46	0.67	1.90	0.22	1.21
11_FaireW	0.18	0.30	0.46	0.42	0.28	1.14
12_CtcfO	0.23	0.24	0.27	0.43	0.23	0.30
13_Ctcf	0.65	0.75	0.60	0.70	0.63	0.86
14_Gen5'	0.70	1.08	0.61	0.28	0.53	0.28
15_Elon	4.75	2.45	4.15	1.36	3.71	1.78
16_ElonW	3.40	2.82	3.29	1.96	3.49	1.82
17_Gen3'	1.17	1.48	0.83	0.77	1.30	0.77
18_Pol2	0.51	0.69	0.77	2.65	0.57	0.91
19_H4K20	0.71	0.28	0.51	0.68	0.11	3.01
20_ReprD	1.00	1.06	0.45	0.46	0.21	0.23
21_Repr	3.27	4.19	4.33	0.44	2.05	1.20
22_ReprW	8.98	9.29	9.02	0.81	8.34	4.38
23_Low	13.64	12.64	17.39	36.19	13.44	40.86
24_Quies	54.87	56.49	51.35	45.75	59.38	36.70
25_Art	0.07	0.23	0.04	0.04	0.04	0.07

e

TSS Enrichments (+/-2KB 5' end)

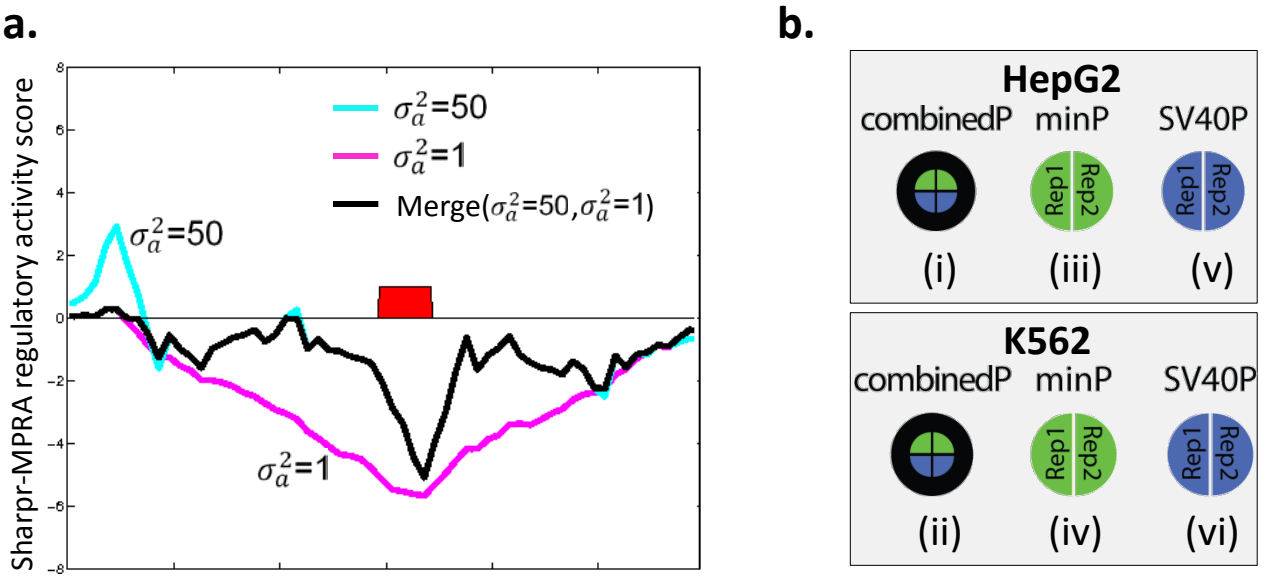
State	HepG2	K562	Huvec	H1-hESC	GM12878	HeLa-s3
1_Tss	24.8	27.6	28.9	30.1	25.0	26.5
2_TssF	18.0	20.7	23.2	24.4	13.1	21.9
3_PromF	7.8	9.1	15.9	12.7	7.9	11.2
4_PromP	13.6	19.2	21.9	20.7	20.4	16.1
5_Enh	1.1	2.9	1.5	2.8	1.2	1.5
6_EnhF	1.2	2.4	1.1	1.2	1.0	1.7
7_EnhWF	1.4	2.6	1.3	1.0	1.3	1.2
8_EnhW	1.7	3.1	1.8	3.1	1.6	1.2
9_DNaseU	1.9	1.0	0.6	0.9	1.0	0.7
10_DNaseD	2.8	4.7	3.3	2.5	3.5	3.4
11_FaireW	0.7	1.0	1.1	0.7	1.5	0.8
12_CtcfO	1.6	1.7	1.6	1.6	1.6	1.7
13_Ctcf	0.8	1.0	1.1	0.7	1.0	1.2
14_Gen5'	1.1	2.3	1.6	2.0	0.9	2.0
15_Elon	0.3	0.3	0.4	0.3	0.3	0.2
16_ElonW	0.3	0.3	0.4	0.4	0.3	0.3
17_Gen3'	0.7	0.6	0.8	0.7	0.8	0.6
18_Pol2	3.1	1.7	1.8	1.4	2.9	2.4
19_H4K20	0.3	0.4	0.2	0.4	0.7	0.6
20_ReprD	4.3	4.1	11.0	9.5	13.4	1.3
21_Repr	1.8	1.3	1.9	4.4	3.4	1.0
22_ReprW	0.9	0.9	1.0	2.2	1.1	0.9
23_Low	0.8	0.9	0.8	0.6	1.0	0.7
24_Quies	0.3	0.3	0.3	0.3	0.3	0.3
25_Art	2.4	2.0	1.9	1.8	1.4	1.3

f

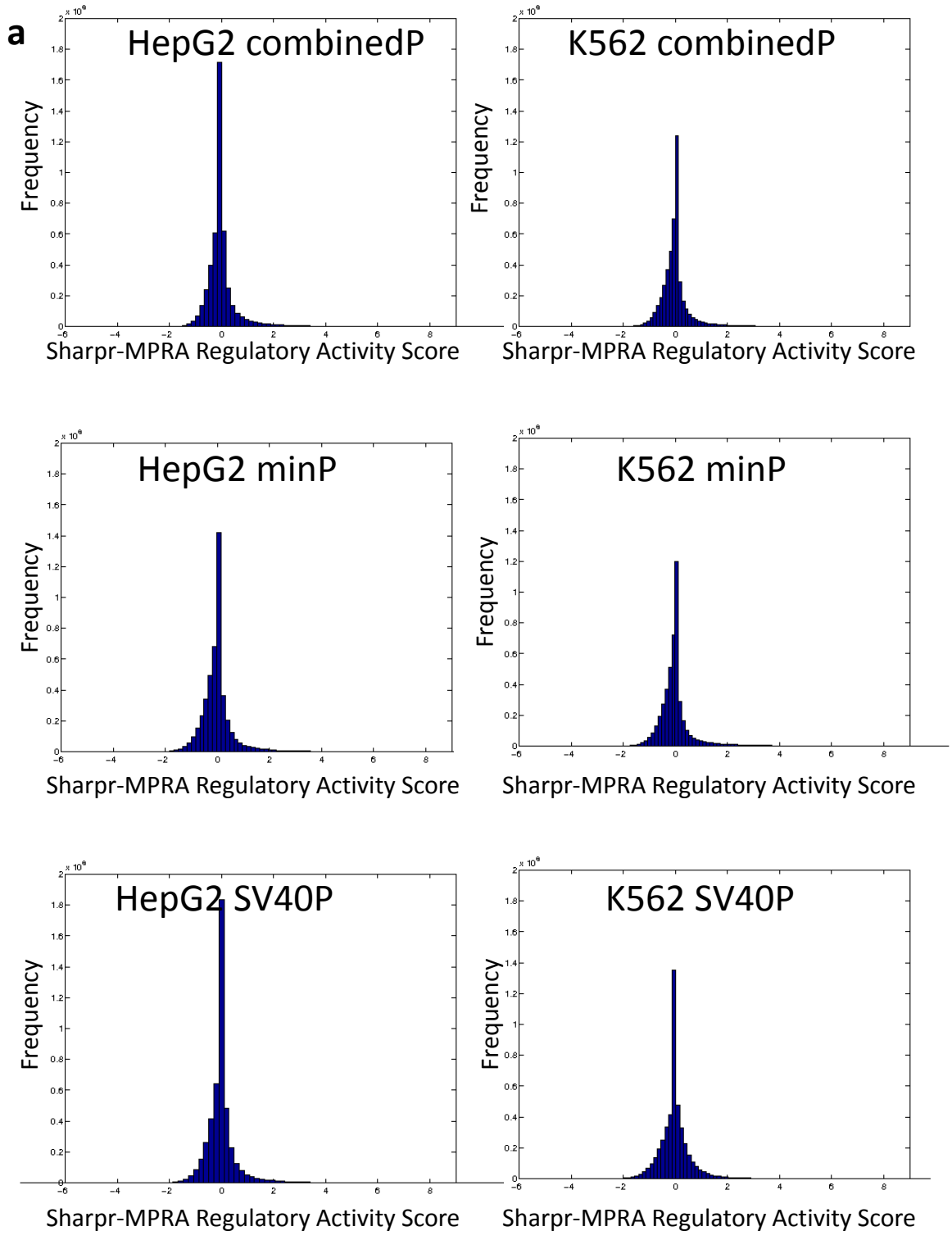
State-to-State Transition Parameters

State	1_Tss	2_TssF	3_PromF	4_PromP	5_Enh	6_EnhF	7_EnhWF	8_EnhW	9_DNaseU	10_DNaseD	11_FaireW	12_CtcfO	13_Ctcf	14_Gen5'	15_Elon	16_ElonW	17_Gen3'	18_Pol2	19_H4K20	20_ReprD	21_Repr	22_ReprW	23_Low	24_Quies	25_Art
1_Tss	83	12	1	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2_TssF	15	68	9	0	1	2	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
3_PromF	1	9	60	5	0	0	4	1	0	0	0	0	0	1	0	1	0	1	0	0	0	2	10	3	0
4_PromP	5	0	7	75	0	0	0	2	0	1	0	1	0	0	0	0	0	0	0	6	0	0	0	0	0
5_Enh	2	1	0	0	71	19	0	4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
6_EnhF	0	2	0	0	14	62	12	2	0	0	3	0	0	1	0	1	0	1	0	0	0	0	2	0	0
7_EnhWF	0	0	1	0	0	5	59	8	1	0	1	0	0	1	0	1	0	0	0	0	0	1	19	2	0
8_EnhW	0	0	1	2	3	2	22	58	1	2	0	1	0	1	0	0	0	1	0	0	0	0	2	0	0
9_DNaseU	0	0	0	0	0	0	4	2	44	0	1	1	0	0	0	0	0	1	1	0	0	2	32	10	0
10_DNaseD	0	0	0	0	0	0	1	1	0	75	0	1	0	0	0	0	0	1	1	1	0	4	12	1	0
11_FaireW	0	0	0	0	0	2	1	0	1	1	65	0	0	0	1	0	0	0	1	0	0	1	20	4	1
12_CtcfO	0	0	0	1	0	0	2	2	1	4	1	48	23	0	2	1	1	2	1	2	0	1	7	2	0
13_Ctcf	0	0	0	0	0	0	1	0	0	1	0	10	45	0	0	1	0	0	1	0	1	7	19	14	0
14_Gen5'	0	2	0	0	2	1	2	1	0	0	0	0	0	78	4	2	4	1	1	0	0	0	0	0	1
15_Elon	0	0	0	0	0	0	0	0	0	0	0	0	0	1	75	17	4	0	0	0	0	0	2	0	0
16_ElonW	0	0	0	0	0	0	1	0	0	0	0	0	0	0	18	77	1	2	1	0	0	0	0	0	0
17_Gen3'	0	0	0	0	0	0	0	0	0	0	0	0	0	2	10	3	80	4	0	0	0	0	0	0	0
18_Pol2	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	5	4	69	1	0	0	0	15	2	0
19_H4K20	0	0	0	0	0	0	1	0	0	1	1	0	0	1	1	2	0	1	57	0	0	2	30	2	0
20_ReprD	0	0	0	0	4	0	0	1	0	0	2	0	1	0	0	0	0	0	0	75	13	4	0	0	0
21_Repr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	70	24	0	2	0
22_ReprW	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	9	87	1	0	0
23_Low	0	0	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	1	1	0	0	0	94	0	0
24_Quies	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0
25_Art	0	0	0	0	0	0	0	0	0	5	1	0	5	1	0	3	0	0	1	0	0	0	0	0	80

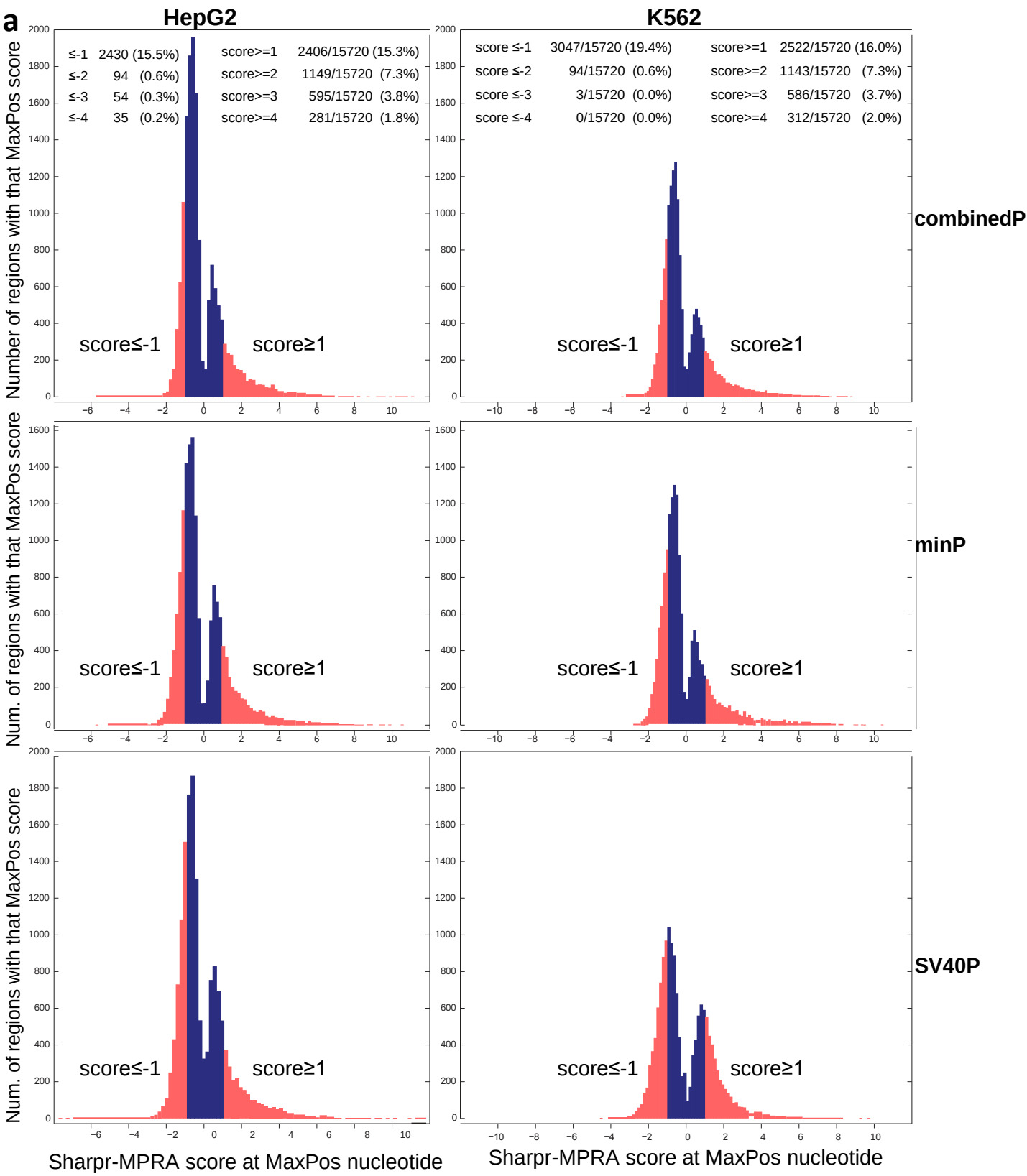
Supplementary Figure 5 – Chromatin State Model used for the Scale-up Experiments (Refs[38,39]). (a) Emission probability matrix, showing expected frequency (percent) of each mark (columns) in each chromatin state (rows). For the scale-up experiments, we used a richer 25-state ChromHMM model and considered a more diverse set of chromatin states, whereas for the pilot we only studied strong enhancer states and low-activity states from a 15-state model (Supplementary Fig. S1). **(b)** Average genome coverage (percent) and median fold enrichments as computed in Ref.[39] for Gencode Transcription Start Sites (TSS) +/- 2kb; Conserved elements using the SiPhy-PI method[42,65] ; (iii) ENCODE transcription factor binding data sets[1]; (iv) repression-associated nuclear lamina-associated domains (Guelen *et al*. *Nature* 453:948–951, 2008). **(c)** Candidate state annotations[39] . **(d)** Percent Genome Coverage of each state (rows) in six ENCODE cell types (columns). Shading highlights the cell types profiled here (HepG2, K562), and the cell types used to select our 15,720 tiled regions (HepG2, K562, Huvec, H1). **(e)** TSS Enrichment (+/- 2kb of GENCODE 5' Ends) for six ENCODE cell types. **(f)** State-to-State transition probabilities denoting the frequency (percent) with which state of the row transitions to the state of the column. Additional state enrichments shown in Refs[38,39].



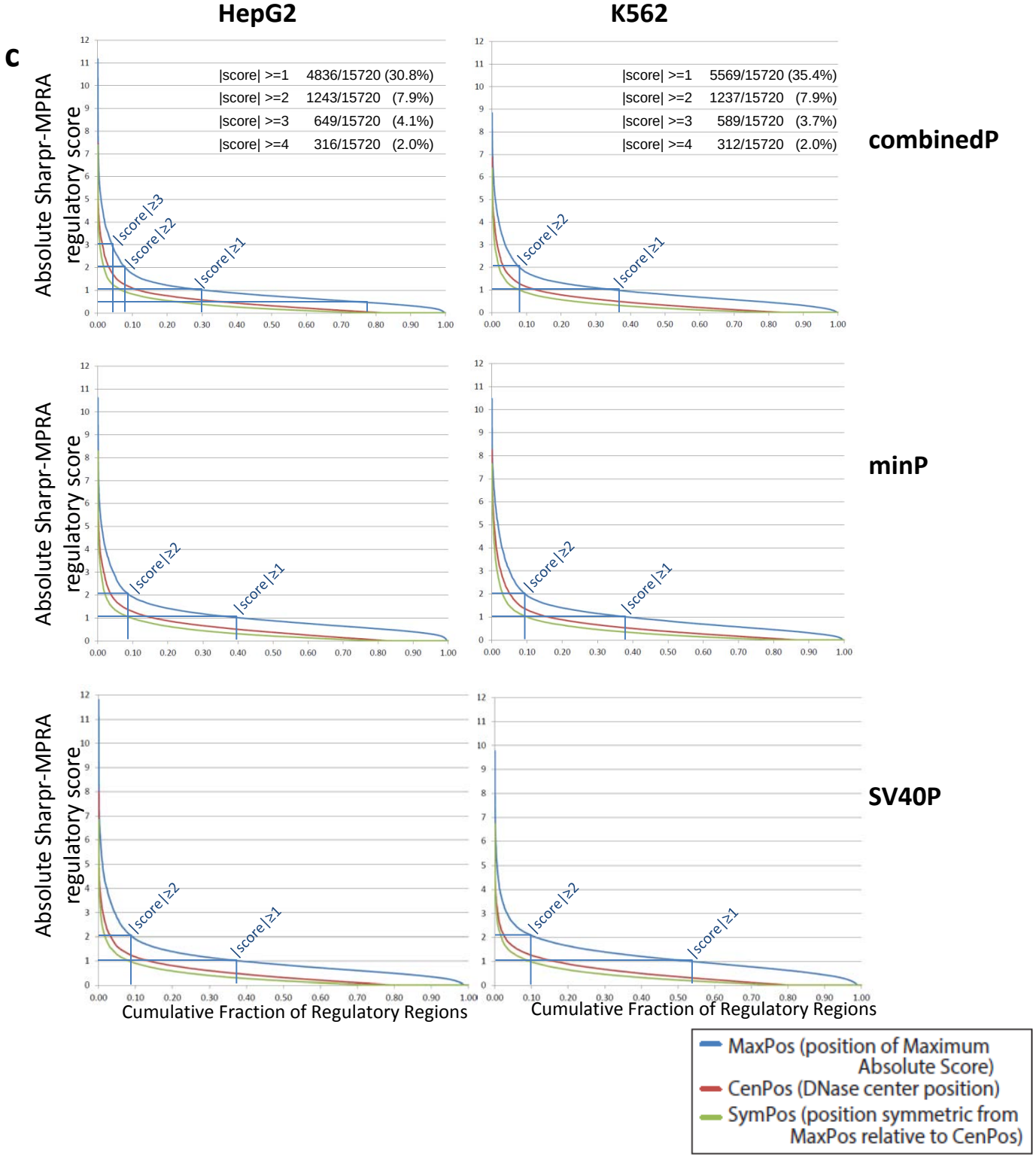
Supplementary Figure 6 – Inference using multiple variance priors, and for multiple replicates, promoters, and cell types. (a) Example Inference with Two Different Variance Parameter Priors. Effect of the variance prior on the output in the example in **Fig. 3b** right. If the variance prior is larger (turquoise line) the output can capture finer features, but can potentially overfit the data as seen by the activation peak on left unsupported by a transcription factor binding site prediction based on CENTIPEDE[8] (shown in red) in close proximity to the repressive peak. On the other hand if a smaller variance prior is used (pink line), then the output is more consistently activating or repressive, but can underfit providing a lower resolution output that can call additional nucleotides activating or repressive that are not. The strategy used was to apply the inference with both a relatively smaller variance prior (σ_α^2=1) and a relatively larger variance prior (σ_α^2=50) and merge the output of two inferences at each position, using the one with smaller absolute value between the two when their signs agreed, and using 0 when their signs disagreed. **(b) Summary of Sharpr-MPRA regulatory activity scores for each region.** The resulting track of applying the two variance parameters is applied in six settings for each regulatory region, resulting in six tracks of Sharpr-MPRA scores, each of 4.6 million nucleotides, which we release as genome browser tracks. The six released tracks correspond to: (i,ii) one combined promoter (combinedP) track for each cell type (black), incorporating a total of four experiments per region, including both minP replicates and both SV40P replicates; (iii,iv) one track for the minimal promoter (minP) in each cell type (green), combining the two replicates for the minP promoter; and (v,vi) one track for the strong promoter (SV40P) in each cell type (blue), combining the two replicates for the SV40P promoter. In addition to these six tracks, we ran SHARPR on each replicate separately for each cell type, and used the resulting inferences to evaluate the reproducibility of our inferences between individual replicates (**Fig. S10a**). For a subset of 44 genomic regions, two different barcode tilings overlaps perfectly, for all 295 nucleotides within them. For an additional 212 regions, multiple barcode tilings partially overlap, resulting in 256 regions with exact or partial overlap, of which 245 have two overlapping barcode sets, and 9 have three overlapping barcode sets. In forming the released browser tracks, scores for a base were averaged when the base was overlapped multiple times.



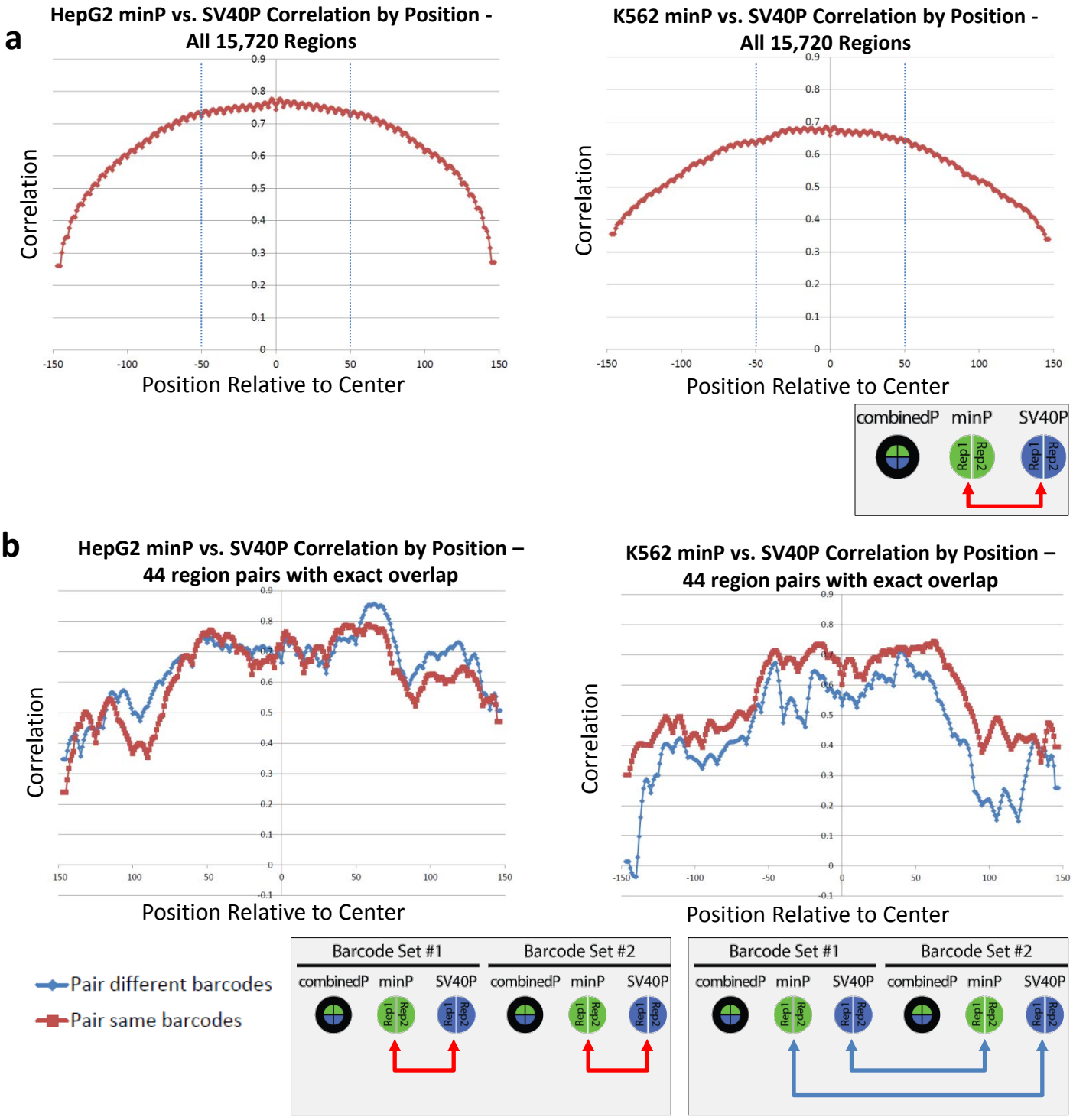
Supplementary Figure 7a – Sharpr-MPRA regulatory activity score distribution. (a) Nucleotide-level score distribution for 4.6 million nucleotide positions. The distribution of regulatory activity score at the nucleotide level for the combinedP promoter score (top row), minP score (middle row), and SV40P score (bottom row) in HepG2 cells (left) and K562 cells (right). The HepG2 distribution (top left) was also shown in gray in **Fig. 3c**. **(b,c)** Next two pages.



Supplementary Figure 7b – Sharpr-MPRA regulatory activity score distribution. (a) Previous page. **(b) MaxPos score distribution at the region level**. Score distribution for all 15,720 tiled region, using the score of a single nucleotide for each, MaxPos, the nucleotide with maximum absolute activity score, shown based on the combined minP and SV40P data (top), minP only (middle row), and SV40P only (bottom), each for HepG2 (left) and K562 (right). Pink bars highlight the subset of regulatory regions with MaxPos≥1 (primarily activating) and MaxPos≤-1 (primarily repressive) that are plotted in **Supplementary Data Files 6a-d**. **(c)** Next Page.



Supplementary Figure 7c – Sharpr-MPRA Regulatory Activity Score Distribution. (a,b) Previous pages. **(c) Absolute MaxPos, CenPos, and SymPos score distributions.** Distribution of the absolute Sharpr-MPRA regulatory score (y-axis) for the 15,720 regions at the nucleotide position of maximum absolute score (MaxPos, blue), the center nucleotide position (CenPos, red), and the symmetric nucleotide positions (SymPos, green), each ranked from highest to lowest absolute score (x-axis) for HepG2 (left) and K562 (right), using: combined minP and SV40P (combinedP) score (top row); the minP score only (middle row); and the SV40P score (bottom row). Blue tickmarks indicate the fraction of MaxPos nucleotides with absolute scores above the indicated values and are also listed for specific values based on the combinedP data (top). MaxPos, CenPos, and SymPos nucleotide positions are illustrated in the example of the bottom-left panel of **Fig. 3b**.



Supplementary Figure 8 – Correlation between minP and SV40P inferred regulatory activity score by nucleotide position. (a) Correlation between the inferred regulatory activities based on the minimal and SV40P promoter data (y-axis) as a function of nucleotide position relative to the DNase peak center position in the tiling (x-axis) for HepG2 (left) and K562 (right). We find higher correlation closer to the center where each nucleotide is covered by more reporter constructs (**Fig. 3a**), and where more meaningful regulation is likely to occur (see **Fig. 2a** and **Supplementary Fig. 2**). **(b)** Comparison of correlation (y-axis) between minP and SV40P using the same barcode set (red curves) and using different barcode sets (blue curves) for 88 pairs of inferences in the 44 regions that were tiled twice using the exact same set of reporter sequences for HepG2 (left) and K562 (right).

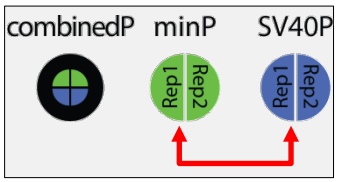
a. Correspondence between SV40P and minP scores: all 15,720 tiled regions, all nucleotides

HepG2

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	33	41	25	1	0	0	1,243,065
[-0.5,0)	16	43	39	2	0	0	3,160,190
[0,0.5)	8	31	54	5	1	0	3,963,513
[0.5,1)	2	12	45	24	11	6	465,209
[1,1.5)	0	3	26	27	22	22	186,738
>=1.5	0	0	6	11	16	67	256,085

K562

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	33	40	24	2	0	0	1,355,150
[-0.5,0)	17	42	37	3	1	0	3,232,903
[0,0.5)	9	33	50	6	2	1	3,673,088
[0.5,1)	6	21	41	17	8	7	528,112
[1,1.5)	2	9	32	20	15	21	209,038
>=1.5	0	2	13	14	16	55	276,509



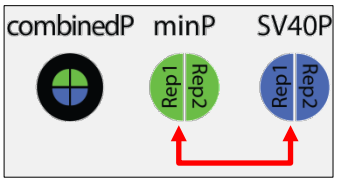
b. Correspondence between SV40P and minP scores: all 15,720 tiled regions, CENTIPEDE nucleotides

HepG2

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	45	33	21	1	0	0	21,331
[-0.5,0)	16	37	42	5	1	0	44,760
[0,0.5)	5	22	57	11	3	1	85,459
[0.5,1)	1	8	38	29	15	10	24,912
[1,1.5)	0	2	19	26	23	29	14,529
>=1.5	0	0	3	7	13	76	31,679

K562

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	39	32	24	4	1	0	17,997
[-0.5,0)	15	35	39	7	3	1	36,937
[0,0.5)	6	20	50	13	6	4	71,452
[0.5,1)	3	10	37	22	13	15	25,281
[1,1.5)	1	6	26	21	16	30	16,335
>=1.5	0	1	8	10	13	67	36,678



Supplementary Figure 9 (a,b) – Nucleotide-level Sharpr-MPRA score correspondence at individual bases across promoter and barcode sets. **(a)** Comparison of Sharpr-MPRA regulatory activity score across minP and SV40 experiments for all nucleotide positions assigned to each indicated score range for HepG2 (left) and K562 (right), as shown by the red arrow in the diagram. For each row, left-most column indicates the score range, right-most column (#bases) indicates total number of nucleotide positions in that score range, and six colored columns indicate percentage of bases assigned to each score range bin. Each comparison was considered in both directions between promoter types, and each base is counted twice, once for its minP score, and once for its SV40P score. For example, 83% of nucleotides with scores ≥1.5 in one promoter type show scores ≥1 in the other promoter type for HepG2 (71% for K562) (red boxes), and 74% of nucleotides with scores < -0.5 in one promoter show scores < 0 in the other promoter for HepG2 (73% for K562) (blue boxes). **(b)** Same as (a) except restricted to those nucleotides overlapping a CENTIPEDE base in the cell type. **(c-f)** Next pages

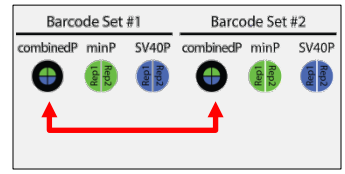
c. Score correspondence between barcode sets: 256 exact- or partial-overlap regions, all nucleotides

HepG2 - Combined Score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	27	48	24	1	0	0	9,431
[-0.5,0)	11	50	38	1	0	0	41,277
[0,0.5)	5	31	54	9	2	0	49,954
[0.5,1)	1	6	45	31	14	4	9,642
[1,1.5)	0	1	15	26	40	18	5,207
>=1.5	0	0	3	6	16	75	5,909

K562 - Combined Score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	24	44	31	1	0	0	12,909
[-0.5,0)	15	42	39	3	0	0	37,838
[0,0.5)	8	31	48	9	3	1	48,268
[0.5,1)	2	13	43	23	12	8	10,399
[1,1.5)	0	2	29	26	19	23	4,589
>=1.5	0	1	8	11	15	66	7,417

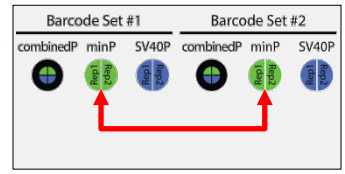


HepG2 - minP score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	21	41	36	2	0	0	12,998
[-0.5,0)	14	41	41	3	0	0	37,510
[0,0.5)	10	32	46	10	2	1	48,393
[0.5,1)	3	11	44	24	12	6	10,914
[1,1.5)	0	3	22	26	25	23	5,051
>=1.5	0	1	5	10	18	66	6,554

K562 - minP score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	29	43	26	2	0	0	14,521
[-0.5,0)	15	45	36	3	0	0	40,044
[0,0.5)	9	33	45	9	2	1	43,739
[0.5,1)	3	12	43	21	11	10	9,338
[1,1.5)	0	2	24	24	22	27	4,199
>=1.5	0	1	6	10	12	71	9,579

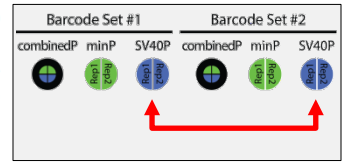


HepG2 - SV40P score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	25	43	31	1	0	0	11,644
[-0.5,0)	15	42	41	3	0	0	34,690
[0,0.5)	7	26	56	8	2	1	53,540
[0.5,1)	1	10	44	26	15	4	10,300
[1,1.5)	0	1	19	28	35	17	5,556
>=1.5	0	0	5	7	17	71	5,690

K562 - SV40P score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	28	29	36	6	1	0	17,140
[-0.5,0)	17	29	43	8	2	1	29,611
[0,0.5)	12	26	45	11	4	2	49,464
[0.5,1)	7	18	41	18	9	6	12,878
[1,1.5)	3	10	35	19	15	18	5,955
>=1.5	0	3	14	12	17	53	6,372



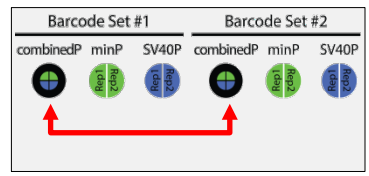
d. Score correspondence between barcode sets: 44 exact-overlaps region pairs, all nucleotides

HepG2 - Combined Score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	38	44	18	0	0	0	1,710
[-0.5,0)	10	50	40	1	0	0	7,799
[0,0.5)	3	27	57	11	1	0	11,246
[0.5,1)	0	4	46	33	15	3	2,631
[1,1.5)	0	0	11	26	51	13	1,493
>=1.5	0	0	2	6	18	74	1,081

K562 - Combined Score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	32	45	22	0	0	0	2,225
[-0.5,0)	13	44	39	4	0	0	7,852
[0,0.5)	5	29	50	12	3	1	10,533
[0.5,1)	0	10	42	31	12	4	2,917
[1,1.5)	0	2	29	31	22	15	1,124
>=1.5	0	0	6	9	13	72	1,309

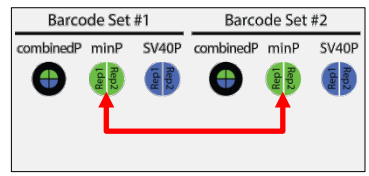


HepG2 - minP score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	18	37	43	1	0	0	2,294
[-0.5,0)	12	41	46	2	0	0	7,266
[0,0.5)	9	31	44	12	3	1	10,636
[0.5,1)	1	4	42	31	15	6	2,966
[1,1.5)	0	0	21	28	30	21	1,562
>=1.5	0	0	7	15	26	52	1,236

K562 - minP score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	45	33	20	2	0	0	2,931
[-0.5,0)	12	48	37	3	0	0	7,818
[0,0.5)	6	30	46	13	3	1	9,429
[0.5,1)	3	9	47	22	12	6	2,568
[1,1.5)	0	1	23	25	30	20	1,230
>=1.5	0	0	5	8	13	74	1,984

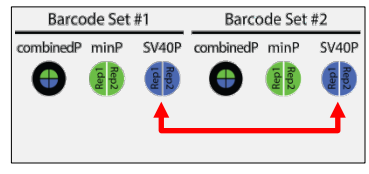


HepG2 - SV40P score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	29	52	19	0	0	0	1,828
[-0.5,0)	13	43	41	4	0	0	7,239
[0,0.5)	3	25	60	10	2	0	11,642
[0.5,1)	0	10	44	27	16	3	2,576
[1,1.5)	0	0	14	26	46	14	1,592
>=1.5	0	0	1	8	21	70	1,083

K562 - SV40P score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases
<-0.5	29	32	32	5	1	0	3,020
[-0.5,0)	15	31	46	6	1	1	6,579
[0,0.5)	9	28	46	12	4	1	10,873
[0.5,1)	5	13	41	27	10	5	3,106
[1,1.5)	2	6	35	23	18	17	1,362
>=1.5	0	3	11	14	22	50	1,020



Supplementary Figure 9 (c,d) – Sharpr-MPRA score correspondence at individual bases across promoter and barcode sets. **(a,b)** Previous page. **(c)** Comparison of Sharpr-MPRA scores at individual bases across different barcode sets for all 256 multiply-tiled regions, including both exact and partial overlaps. Comparisons are shown for the minP and SV40P combined data (top), minP data only (middle row), and SV40P data only (bottom row), with format similar to panel a. **(d)** Same as (c) except for only the subset of 44 regions that were covered by two sets of barcodes with exact tile overlap. **(e,f)** Next page.

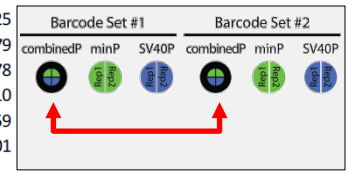
e. Score correspondence between barcode sets: 256 exact- or partial-overlap regions, CENTIPEDE nucleotides

HepG2 - Combined Score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	30	35	33	2	0	0	449
[-0.5,0)	12	41	43	4	1	0	1,329
[0,0.5)	5	20	60	11	3	1	2,807
[0.5,1)	1	7	40	32	15	5	810
[1,1.5)	0	1	13	22	41	22	547
>=1.5	0	0	3	4	12	81	1,022

K562 - Combined Score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	26	46	26	2	0	0	325
[-0.5,0)	14	21	52	9	2	3	1,079
[0,0.5)	3	23	49	18	5	2	2,478
[0.5,1)	1	10	50	12	13	13	910
[1,1.5)	0	4	26	26	17	26	459
>=1.5	0	2	3	9	9	76	1,301

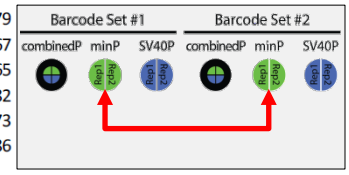


HepG2 - minP score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	15	33	46	6	0	0	615
[-0.5,0)	14	35	41	7	1	2	1,441
[0,0.5)	11	24	47	13	4	1	2,478
[0.5,1)	4	12	35	24	16	8	889
[1,1.5)	0	2	19	28	26	26	529
>=1.5	0	2	3	7	14	74	1,012

K562 - minP score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	21	55	23	1	1	0	379
[-0.5,0)	14	45	34	5	1	1	1,467
[0,0.5)	5	27	43	18	6	1	1,865
[0.5,1)	1	8	39	28	10	15	882
[1,1.5)	0	4	22	19	27	29	473
>=1.5	0	1	2	9	9	80	1,486

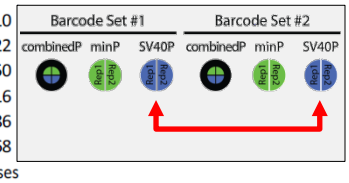


HepG2 - SV40P score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	42	29	29	0	0	0	432
[-0.5,0)	10	40	41	8	1	0	1,294
[0,0.5)	5	20	59	13	3	1	2,750
[0.5,1)	0	12	41	25	16	7	900
[1,1.5)	0	1	14	25	46	14	568
>=1.5	0	0	2	6	8	84	1,020
	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases

K562 - SV40P score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	33	15	39	10	3	0	510
[-0.5,0)	9	15	49	18	6	2	822
[0,0.5)	8	17	43	17	12	4	2,450
[0.5,1)	5	15	40	16	9	15	1,016
[1,1.5)	2	7	43	14	13	22	686
>=1.5	0	2	9	14	14	61	1,068
	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases



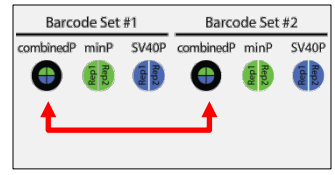
f. Score correspondence between barcode sets: 44 exact-overlap regions, CENTIPEDE nucleotides

HepG2 - Combined Score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	39	40	21	0	0	0	469
[-0.5,0)	11	53	36	0	0	0	1,669
[0,0.5)	4	26	57	12	1	0	2,323
[0.5,1)	0	1	46	31	20	2	620
[1,1.5)	0	0	5	30	52	13	424
>=1.5	0	0	0	5	21	74	265

K562 - Combined Score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	42	37	21	0	0	0	602
[-0.5,0)	12	52	31	4	0	0	1,806
[0,0.5)	7	29	48	12	4	0	1,913
[0.5,1)	0	13	37	32	16	2	607
[1,1.5)	0	3	25	33	21	19	297
>=1.5	0	0	2	4	15	79	369

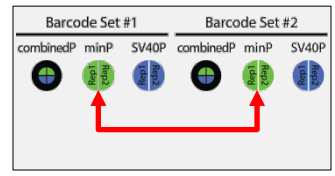


HepG2 - minP score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	13	44	42	1	0	0	553
[-0.5,0)	15	42	40	2	0	0	1,639
[0,0.5)	11	32	40	13	4	0	2,103
[0.5,1)	1	4	36	37	17	6	791
[1,1.5)	0	1	24	37	16	22	356
>=1.5	0	0	2	13	24	60	328

K562 - minP score

	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	47	35	16	2	0	0	763
[-0.5,0)	16	51	29	4	0	0	1,665
[0,0.5)	7	27	48	12	6	1	1,802
[0.5,1)	3	11	39	29	15	4	564
[1,1.5)	0	1	30	24	27	17	345
>=1.5	0	0	3	5	13	80	455

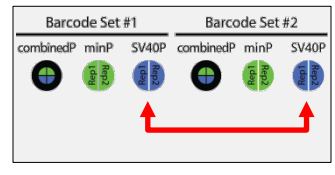


HepG2 - SV40P score

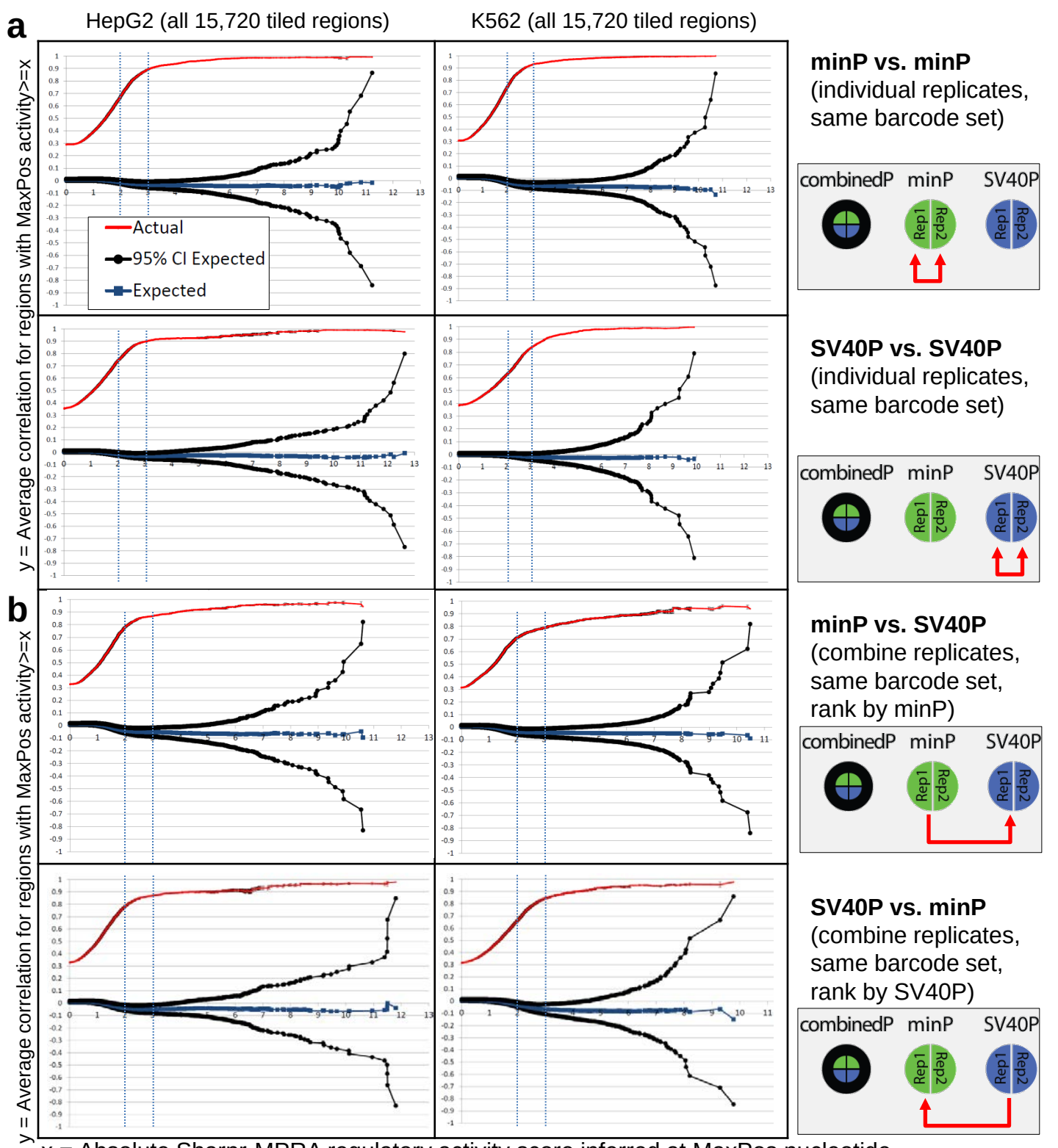
	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	41	35	24	0	0	0	478
[-0.5,0)	11	52	34	4	0	0	1,554
[0,0.5)	5	22	56	14	3	0	2,351
[0.5,1)	0	9	47	27	14	2	699
[1,1.5)	0	0	14	23	54	10	443
>=1.5	0	0	1	6	18	76	245
	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases

K562 - SV40P score

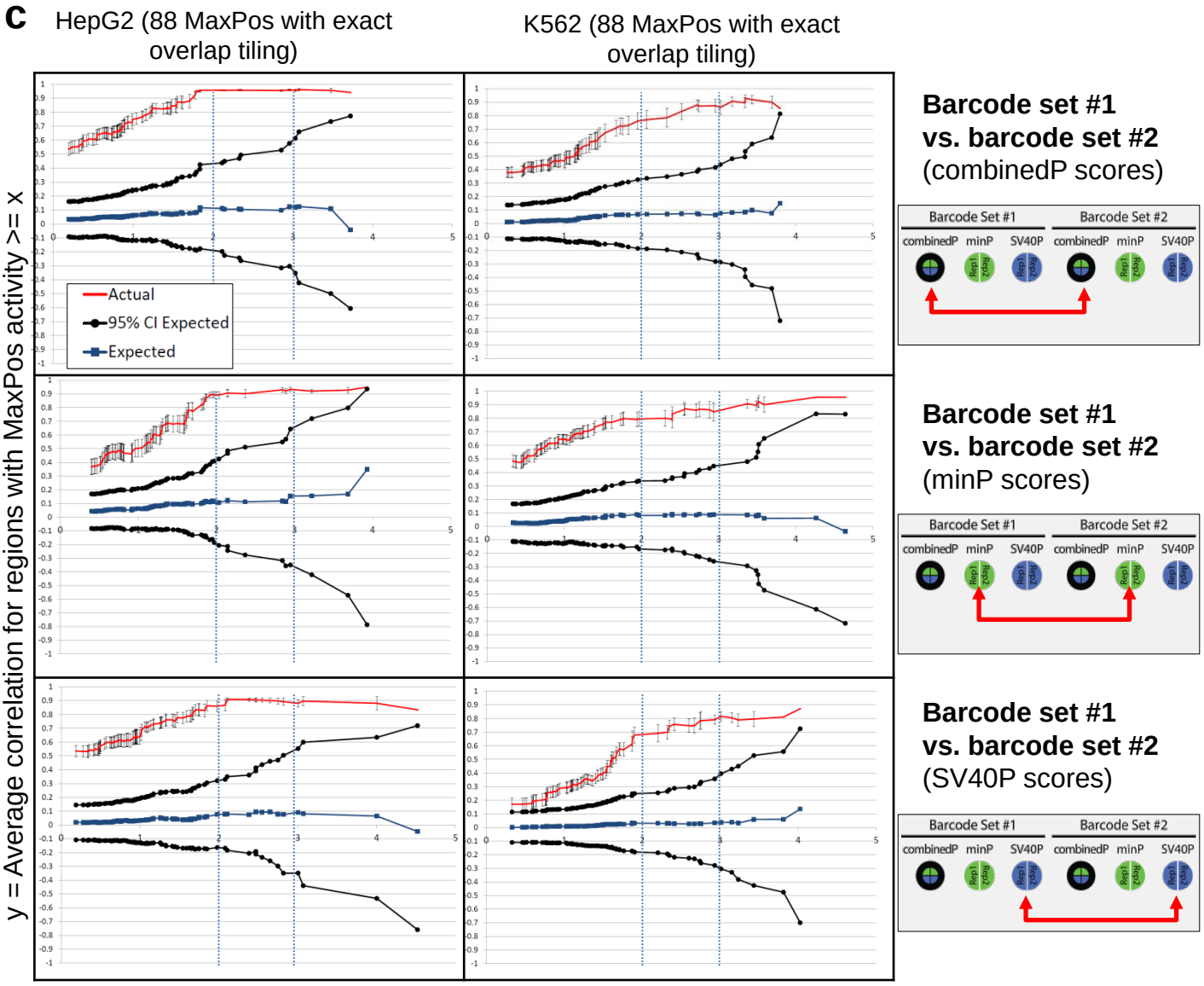
	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	
<-0.5	35	31	23	11	1	0	780
[-0.5,0)	17	37	38	6	1	1	1,439
[0,0.5)	9	28	44	15	4	1	2,002
[0.5,1)	11	11	38	24	11	6	784
[1,1.5)	1	4	24	25	22	25	344
>=1.5	0	3	6	19	36	36	245
	<-0.5	[-0.5,0)	[0,0.5)	[0.5,1)	[1,1.5)	>=1.5	#bases



Supplementary Figure 9 (e,f) – Sharpr-MPRA score correspondence at individual bases across promoter and barcode sets. (a-d) Previous pages. **(e)** Comparison of Sharpr-MPRA scores inferred across different barcode sets for all 256 multiply-tiled regions, including both exact and partial overlaps similar to (c), except restricted to those nucleotides overlapping a CENTIPEDE base in the cell type. **(f)** The same as (e), except for only the subset of 44 regions that were covered by two sets of barcodes with exact tile overlap.

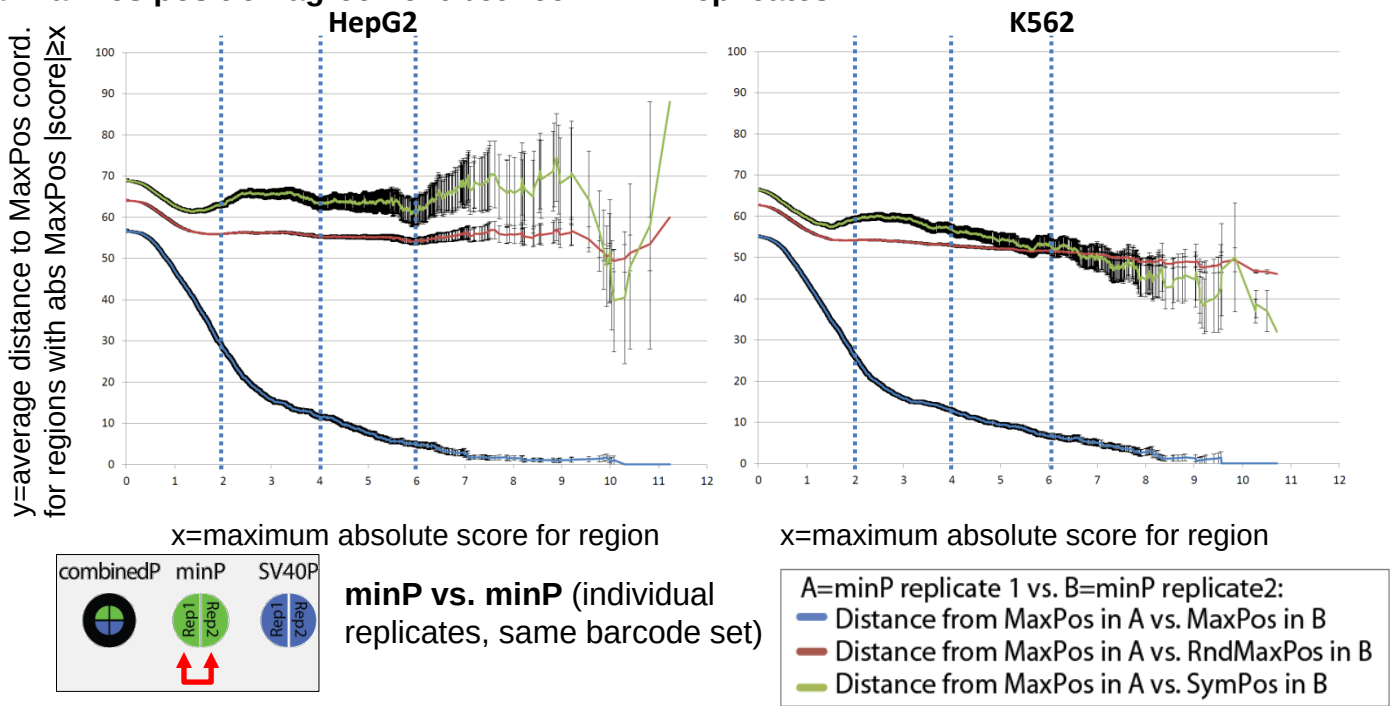


Supplementary Figure 10a,b – Sharpr-MPRA within-region reproducibility correlation analysis. **(a)** Average within region Pearson correlation in activity scores across individual replicates of minP and SV40P experiments (y-axis) for regions meeting or exceeding varying maximum absolute score values (x-axis) for HepG2 (left) and K562 (right), with comparison performed indicated by the red arrow in the corresponding diagram. Each region was included twice, once based on the MaxPos value from each replicate. Observed correlation (red curve) are compared to the expected value (blue) and 95% confidence interval (black) of 10,000 row-wise permutations of regions. Error bars indicate standard error. **(b)** Average Pearson correlation between minP and SV40P similar to panel (a). Each region was included only once and the MaxPos score on the x-axis is based on minP for the first row and SV40P for the second row. **(c)** Next two pages.



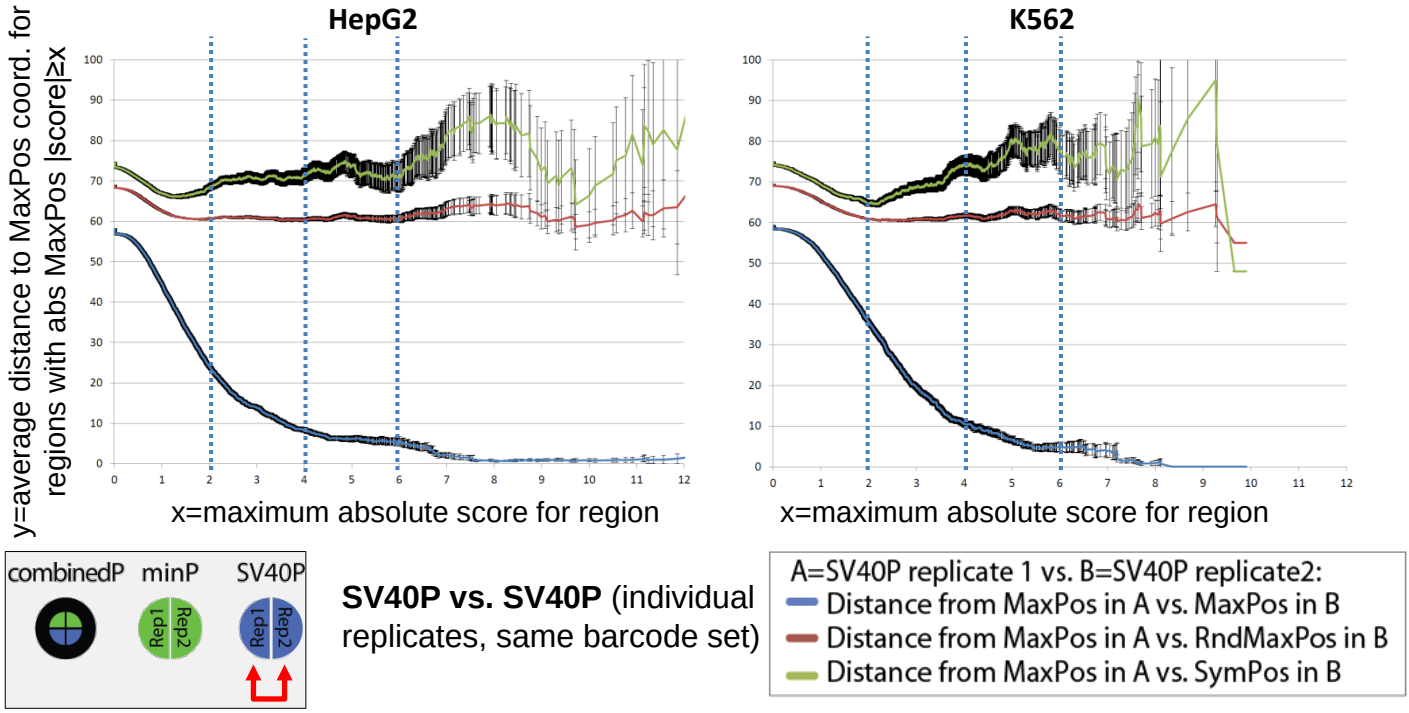
Supplementary Figure 10c. Sharpr-MPRA within-region reproducibility correlation analysis. (a-b) See previous page. **(c)** Similar to (a,b), but with all comparisons performed across different barcode sets of the same experiment. Top: Barcode comparison using combined minP and SV40P data. Middle row: Barcode comparison using minP data. Bottom: Barcode comparison using the SV40P data. Each panel shows 88 points corresponding to two MaxPos values for each of the 44 regions with exact overlap tiling.

a. MaxPos position agreement between minP replicates



minP vs. minP (individual replicates, same barcode set)

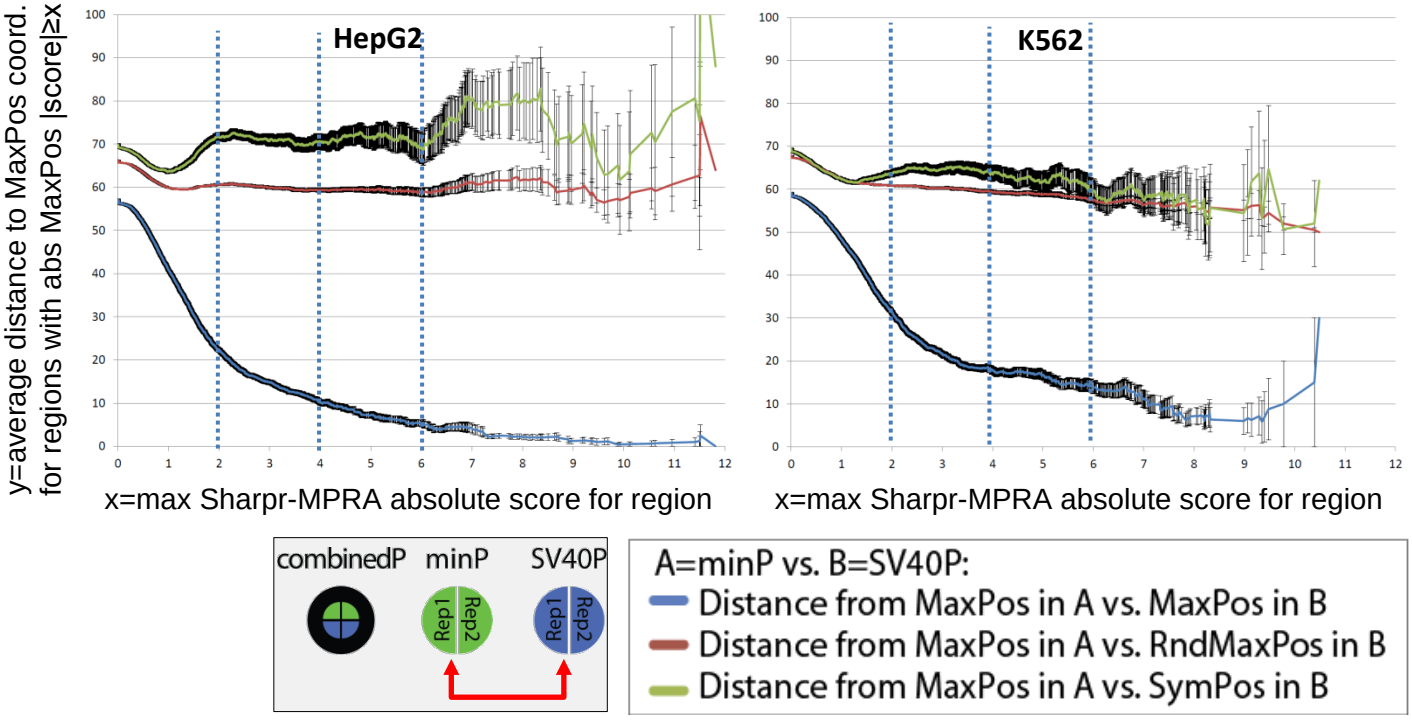
b. MaxPos position agreement between SV40P replicates



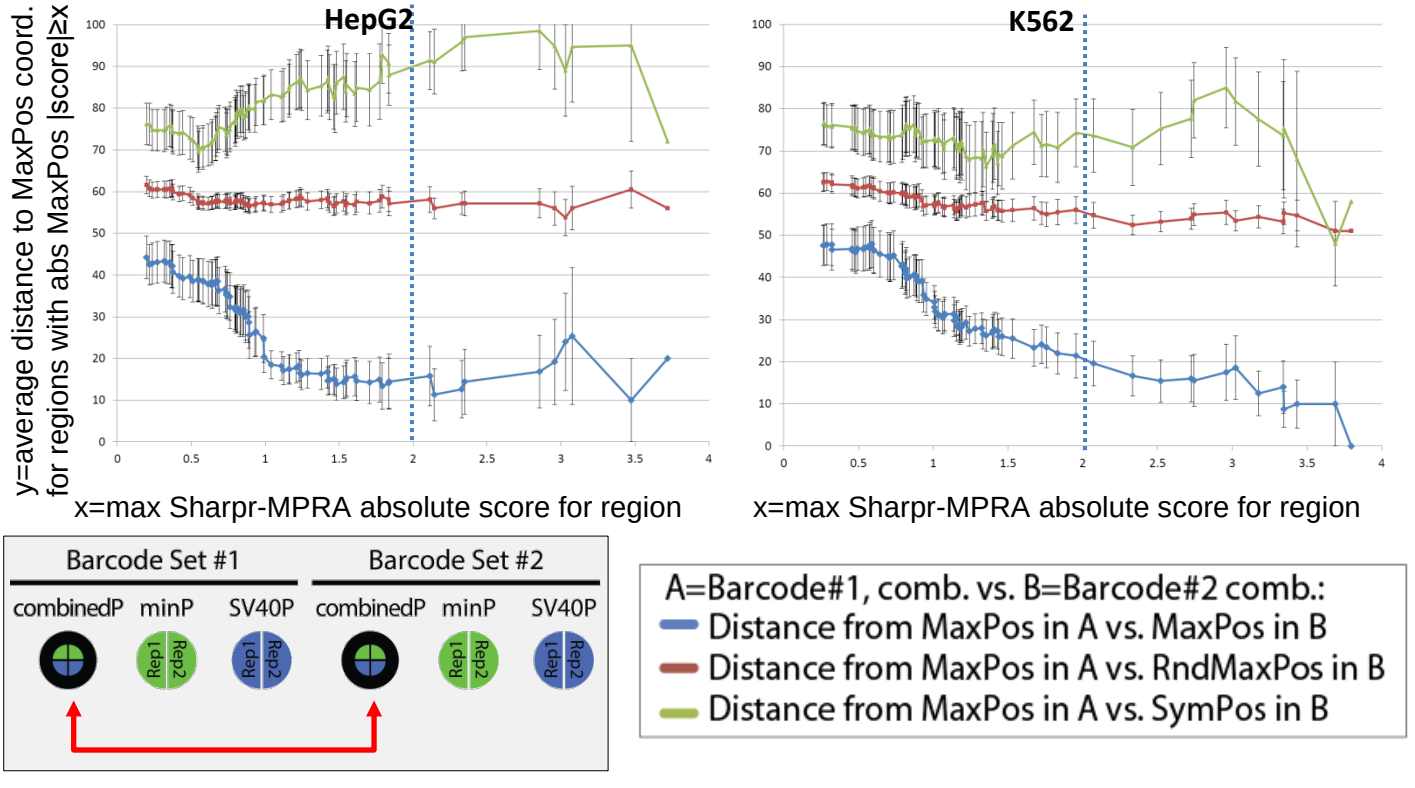
SV40P vs. SV40P (individual replicates, same barcode set)

Supplementary Figure 11a,b. Agreement in position of maximum absolute activity (MaxPos) between replicates, promoter types, and barcodes. **(a)** Average absolute distance (y-axis) between the position of maximum absolute activity (MaxPos) in one minP replicate and the other for those regions meeting or exceeding varying maximum absolute score values (x-axis) for HepG2 (left) and K562 (right) (MaxPos, blue line). Each distance was included twice, once based on each of the two MaxPos values in the pair. This is compared to the expected distance between MaxPos in one minP-replicate and a position in the other minP replicate if sampled randomly from the distribution of all MaxPos positions across all regions (RndMaxPos, red line), and also compared to the distance between MaxPos in one minP replicate and the symmetric position in minP replicate (SymPos, green line). This shows that even when MaxPos regions are highly off-centered, there is significantly higher agreement in the MaxPos position than expected by chance. **(b)** Same comparison as (a) for SV40P individual replicates for HepG2 (left) and K562 (right). **(c,d)** Next page.

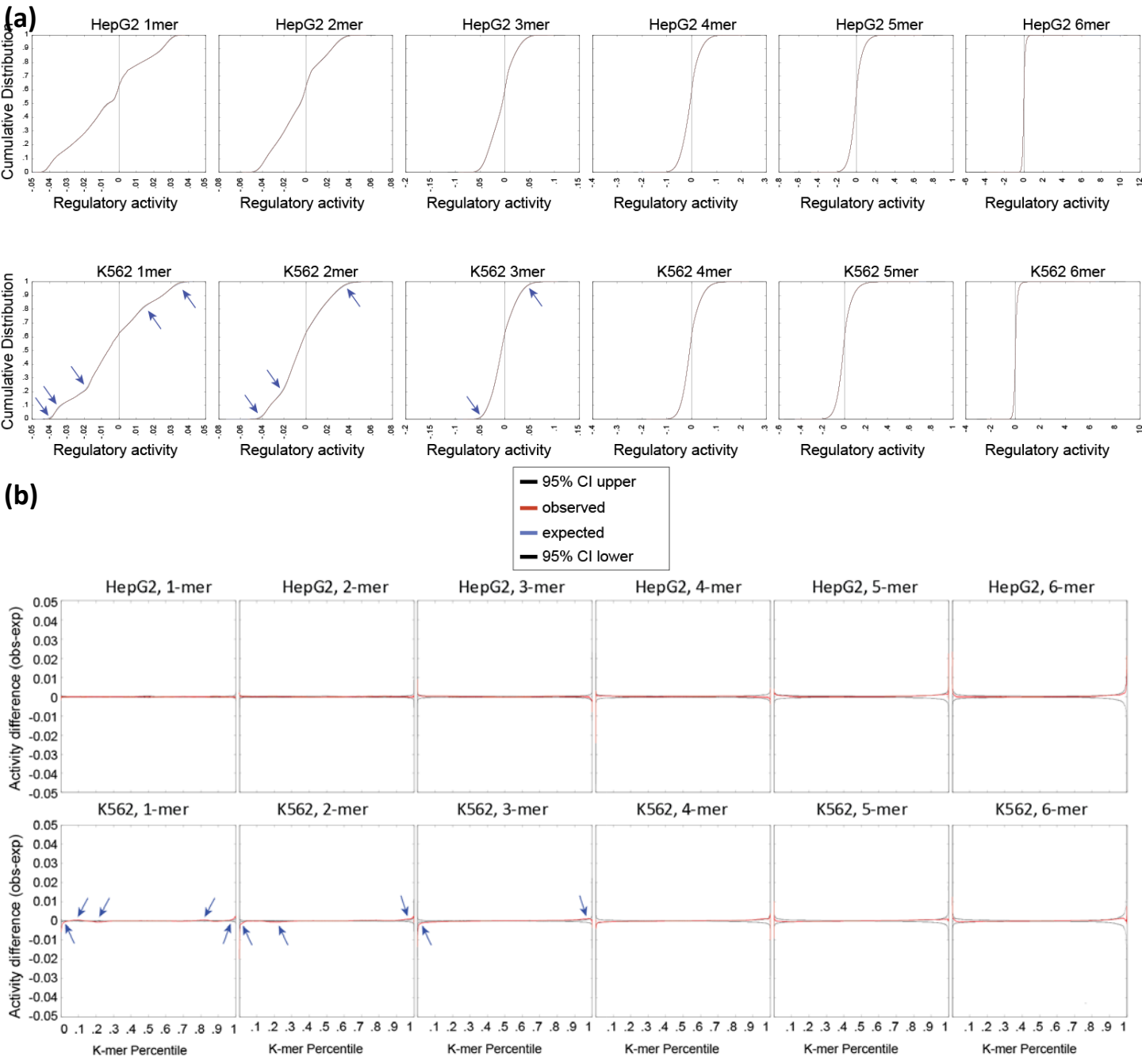
c. minP vs. SV40P (combine replicates, same barcode set)



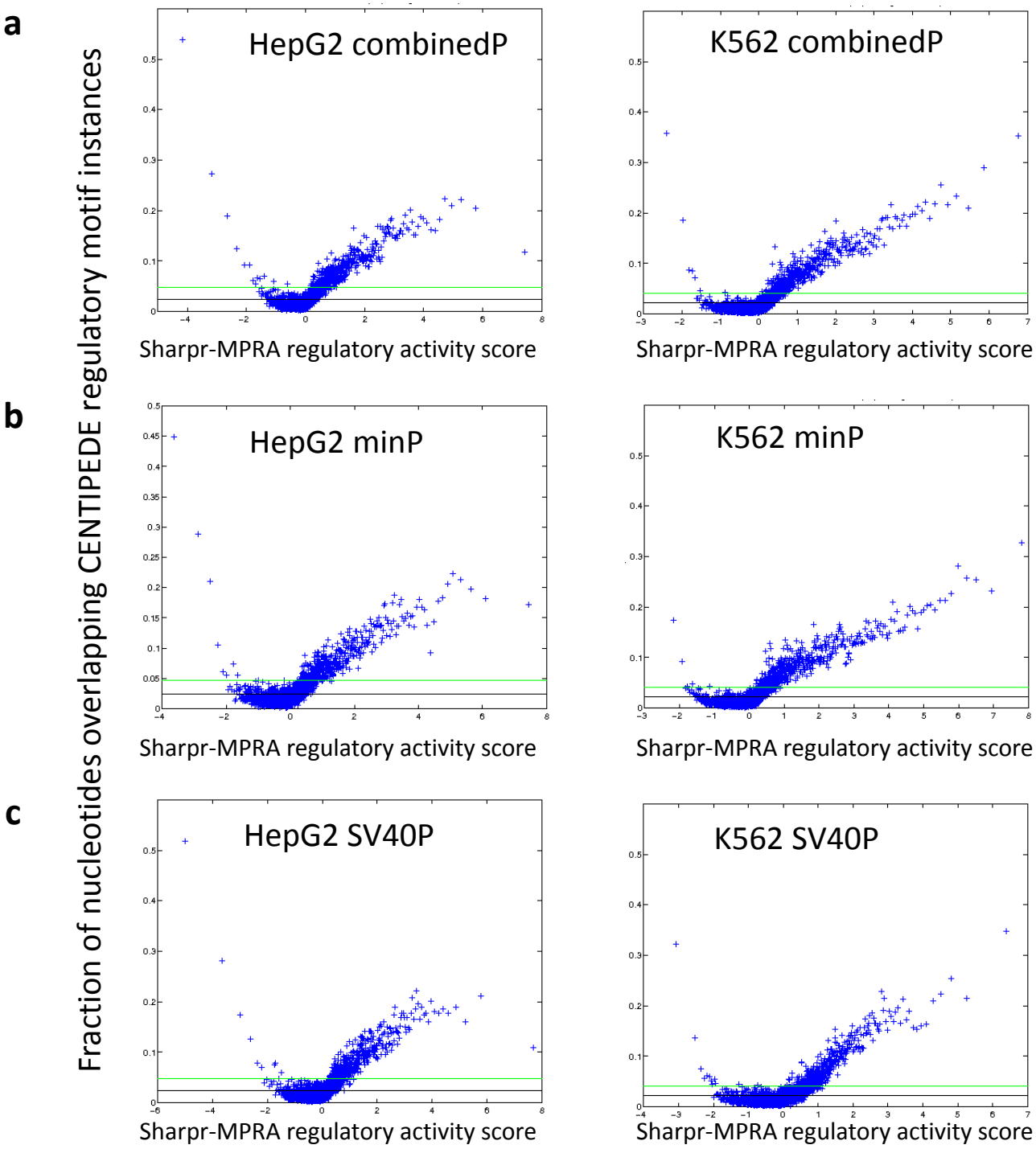
d. Barcode set #1 vs. barcode set #2, in 44 exact-overlap regions (combinedP score)



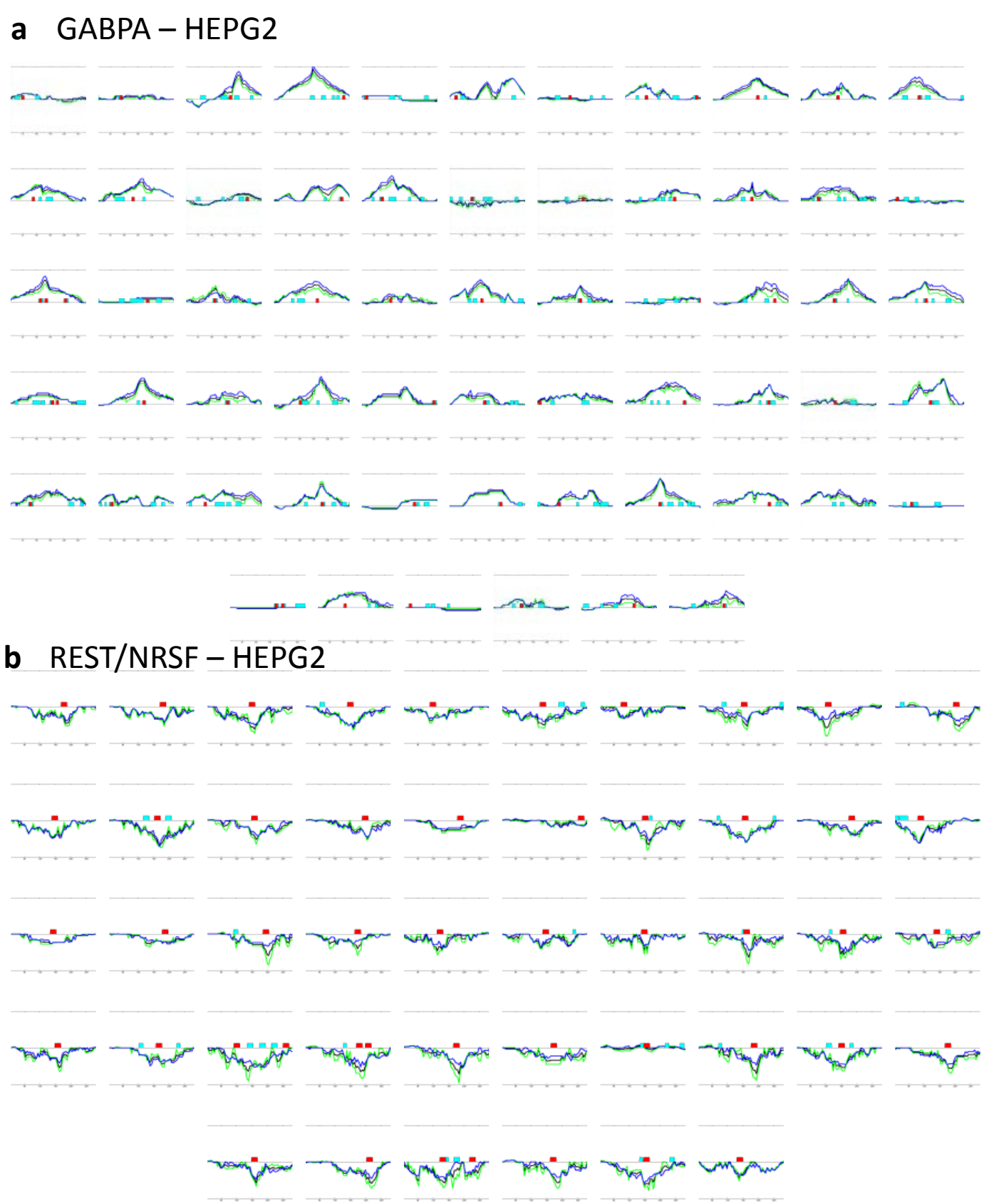
Supplementary Figure 11c,d. Agreement in position of maximum absolute activity (MaxPos) between replicates, promoter types, and barcodes. **(a,b)** Previous page. **(c)** Same comparison as (a), but the comparisons are between SV40P and minP. **(d)** Same comparisons as (a), but comparing the combined minP and SV40P data for two different barcode sets for the 44 regions tiled with exact overlap from two barcode sets.



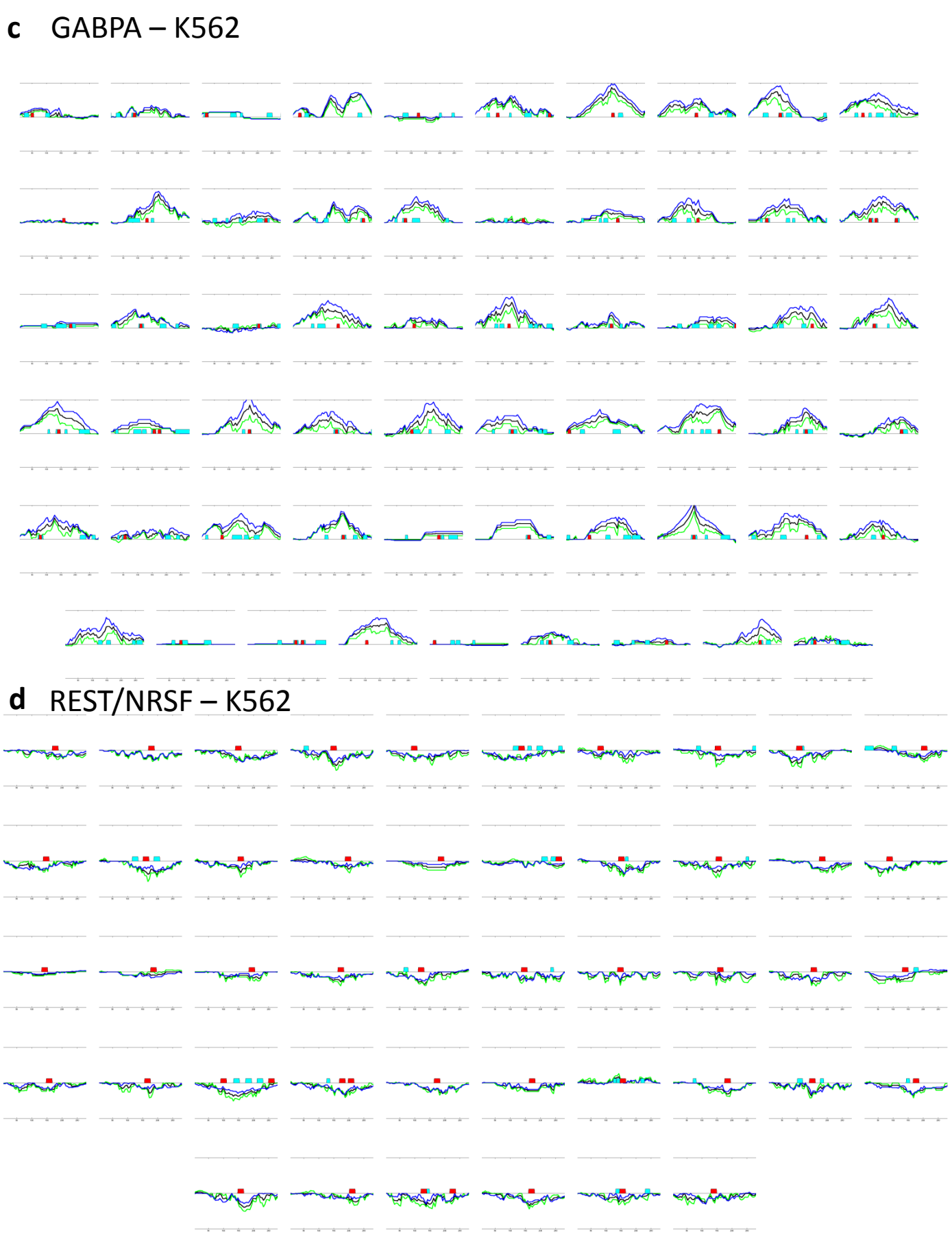
Supplementary Figure 12 –Effect of the k-mer sequence within the 10-nucleotide barcode sequence on the Sharpr-MPRA regulatory activity scores. (a) Cumulative distribution of the average inferred activity (red) over each combination of k-mer sequence in the barcode, presence in each of the 31 tile positions, and the 295 inferred activity positions. Also shown is the expected (blue) and 95% confidence intervals (black) based on 400 randomizations in which all the barcodes were randomly reassigned to reporter sequences. In all HepG2 plots, there is no visible difference between the observed, expected, and 95% confidence intervals. Arrows denote the visible differences where the observed distribution (red) deviates from the expected distribution (blue) and the 95% confidence intervals (black). **(b)** Same as (a), but showing the difference between observed (red curve in panel a) and expected (blue curve in panel (a)) for each k-mer length and each cell type. Arrows denote the difference between the distribution of observed and expected activity for short k-mers in K562. In particular, AA and AT di-nucleotides were over-represented in the last two positions of barcodes with low reporter activity, suggesting a technical bias in reverse transcription initiation, rather than a biological role of 3'UTR motifs, as the last two barcode positions are the first to be reverse-transcribed. Note that the highlighted differences are on the order of 0.01, compared to activity scores ranging between 0.5-8 for positions that show motif enrichments (see for example **Fig. 3c**, or **Supplementary Fig. S13**).



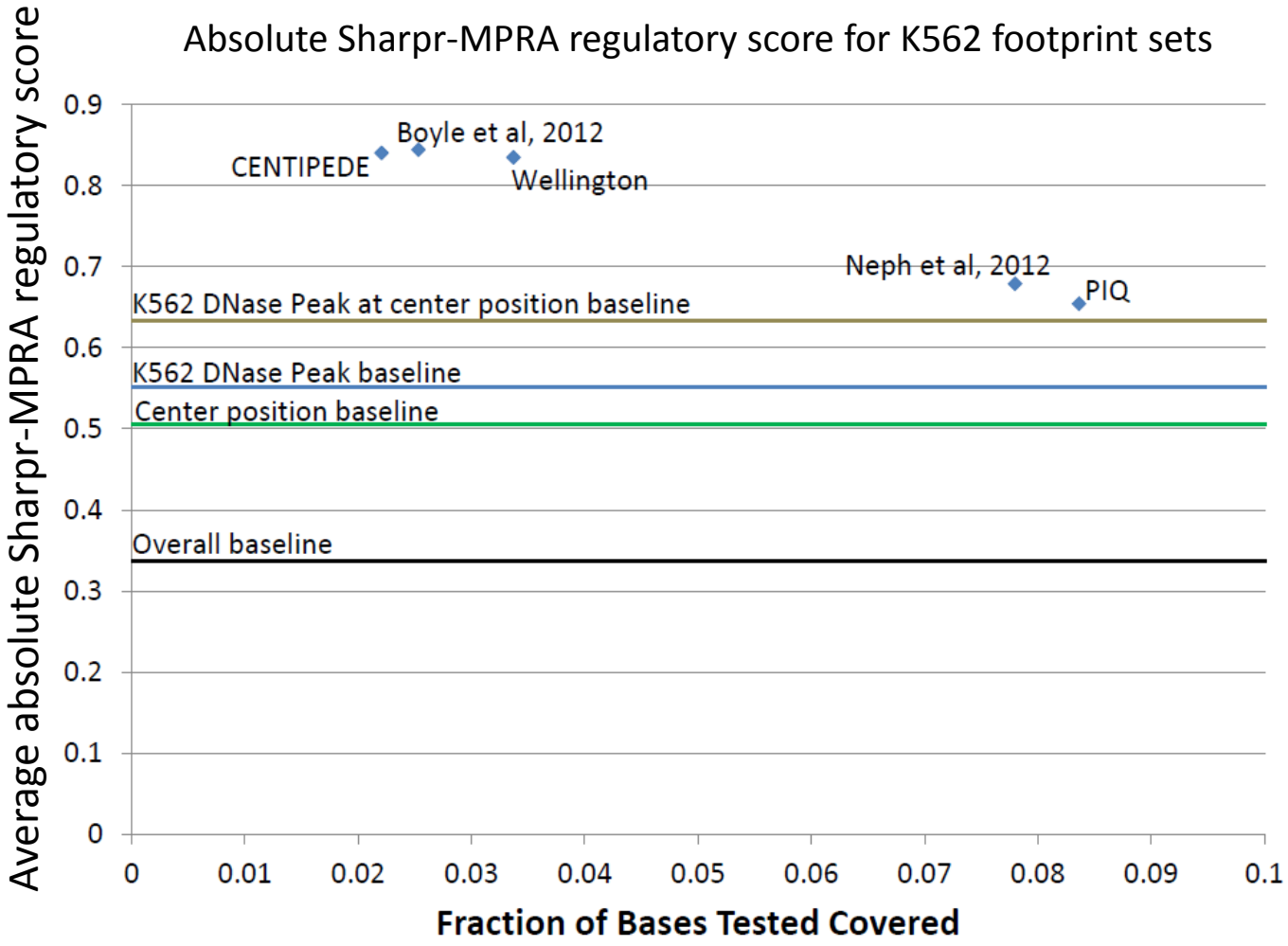
Supplementary Figure 13 – Overlap with CENTIPEDE predicted transcription factor binding sites as a function of Sharpr-MPRA regulatory activity score. Extended version of **Fig. 3c** left, but showing the results for both K562 and HepG2 and the combined data, minP only, and SV40P only data. Each point represents the average of 927 nucleotide positions in each of 5,000 quantiles. Horizontal black line shows the expected overlap averaged across all 295 nucleotide positions of each region, and the green line shows the expected overlap fraction at the center nucleotide position. These are shown for: **(a)** the combined minP and SV40P data for HepG2 (left) and K562 (right); **(b)** the minP data only for HepG2 (left) and K562 (right); **(c)** the SV40P data only for HepG2 (left) and K562 (right).



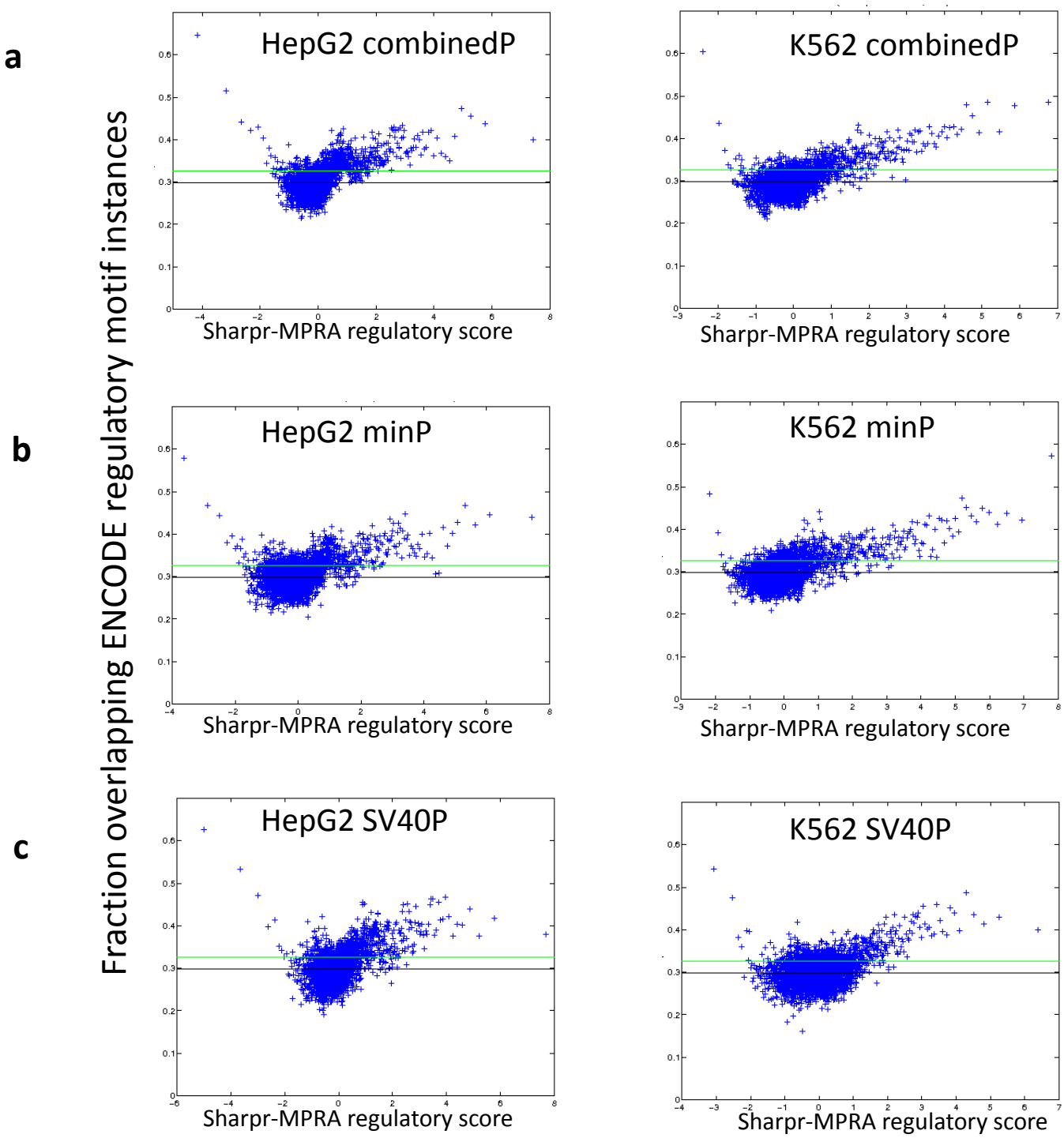
Supplementary Figure 14 – Individual Examples. (a) The figure shows relative Sharpr-MPRA regulatory activity scores overlapping all cell type specific binding sites for GABPA in HepG2 cells predicted based on the CENTIPEDE method[8] contained in regions tested. In red are the GABPA motifs, while in cyan are predicted binding sites in HepG2 for other regulators. The Sharpr-MPRA score based on the SV40P data is shown in green, the minP data in blue, and the combination of SV40P and minP in black. **(b)** Same as (a) except for the predicted binding sites of the repressor REST in HepG2 cells. **(c,d)** Next page.



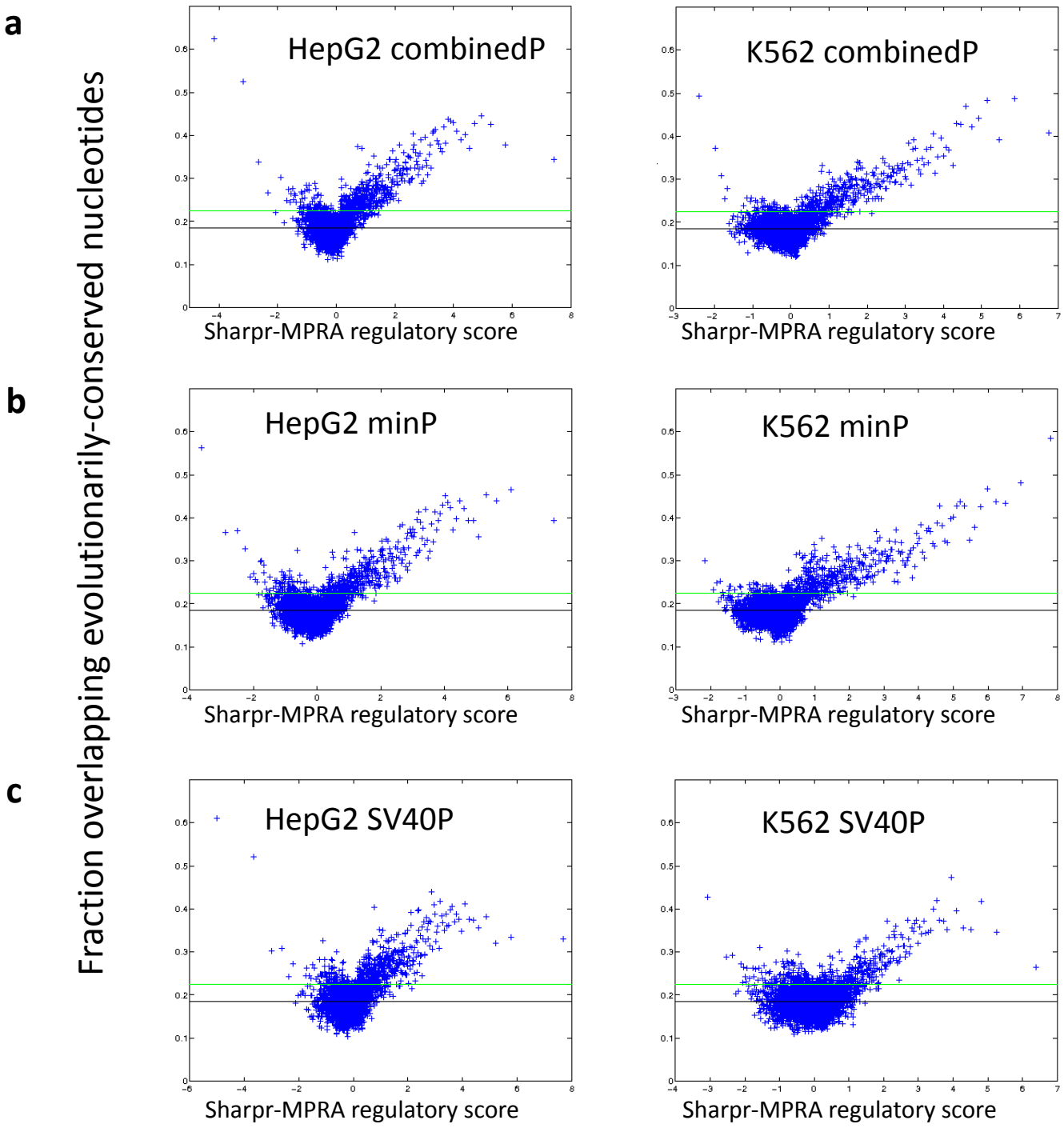
Supplementary Figure 14 – Individual Examples. (a,b) Previous page. **(c,d)** The same as (a,b) except for K562 cells.



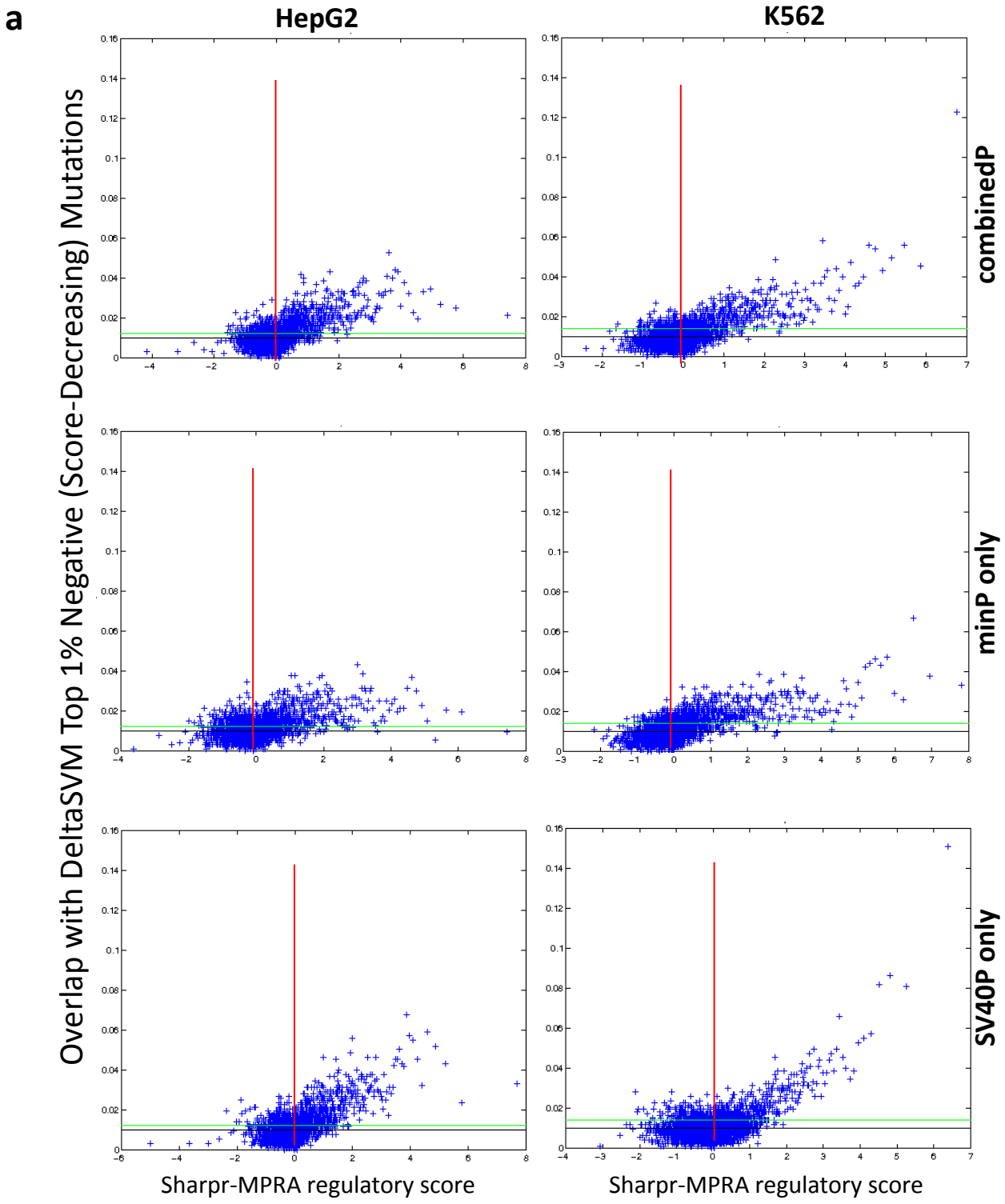
Supplementary Figure 15 – Comparison of Different Footprint Sets. The plot evaluates in K562 cells the average absolute regulatory activity (y-axis) for positions overlapping predictions of locations of transcription factor binding sites previously predicted based on DNase footprint information in K562 cells by five different methods. Two of the five methods also use motif information, CENTIPEDE[8] and PIQ[41], while the other three methods are motif independent, Wellington[40] and the methods of Boyle et al[5] and Neph et al[6]. The x-axis shows the fraction of nucleotides each of these footprint sets overlap, showing the two footprint sets that overlap more nucleotides tested[6,41] had a relatively lower average absolute activity compared to the other three. All five sets had greater absolute activity than four different baselines: (1) at the center position restricted to regions overlapping a K562 DNase peak, (2) at any position overlapping a K562 DNase peak, (3) at the center position over all regions, and (4) over all positions in all regions tested.



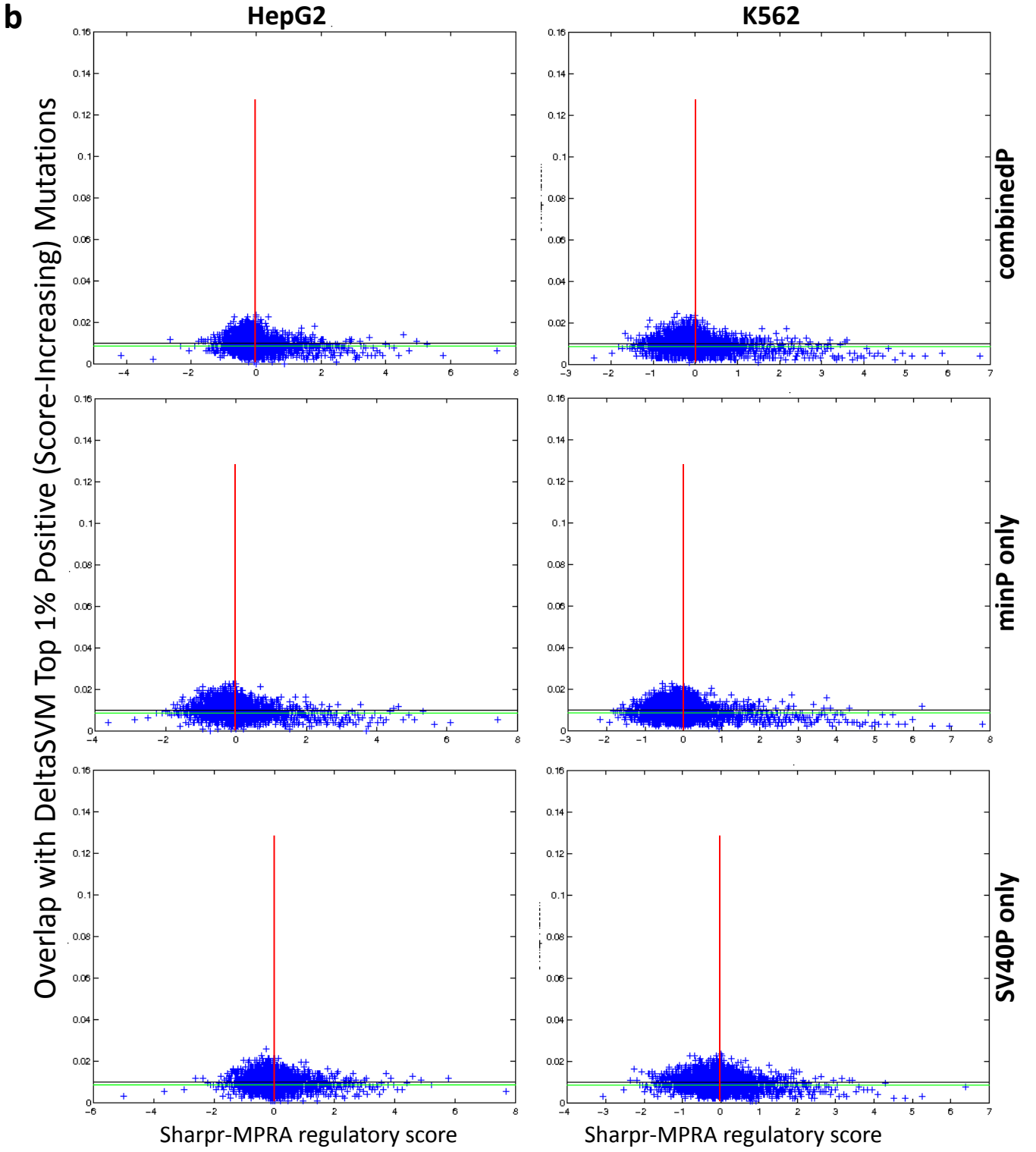
Supplementary Figure 16 – Overlap with motif instance predictions as a function of Sharpr-MPRA regulatory activity score. These are analogous plots to **Supplementary Fig. 13** except they are for the set of nucleotides covered by a motif instance prediction from Ref. 13, which does not use conservation or make cell type specific predictions. The plot is for Sharpr-MPRA regulatory activity scores in HepG2 (left) and K562 (right) cells based on: (**a**) minP and SV40P combined data; (**b**) minP data only; and (**c**) SV40P data only.



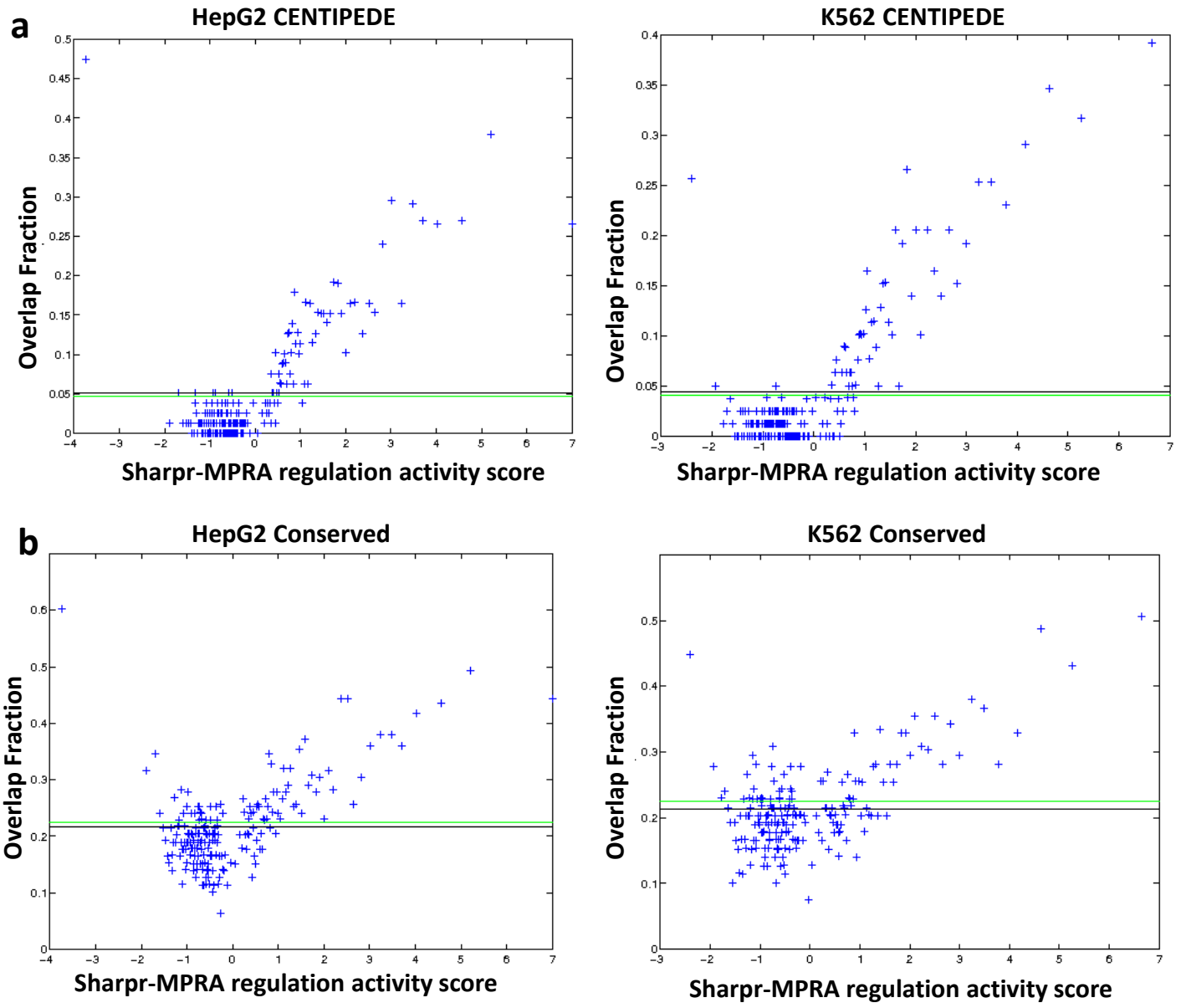
Supplementary Figure 17 – Overlap with Conserved Elements as a function of Sharpr-MPRA regulatory activity score. Extended version of **Fig. 3c** right and analogous to **Supplementary Fig. 13**, but showing the results for bases overlapping conserved elements from the SiPhy-PI method[42,65] for HepG2 cells (left) and K562 cells (right), based on: **(a)** the combined minP and SV40P data; **(b)** minP data only; and **(c)** SV40P data only.



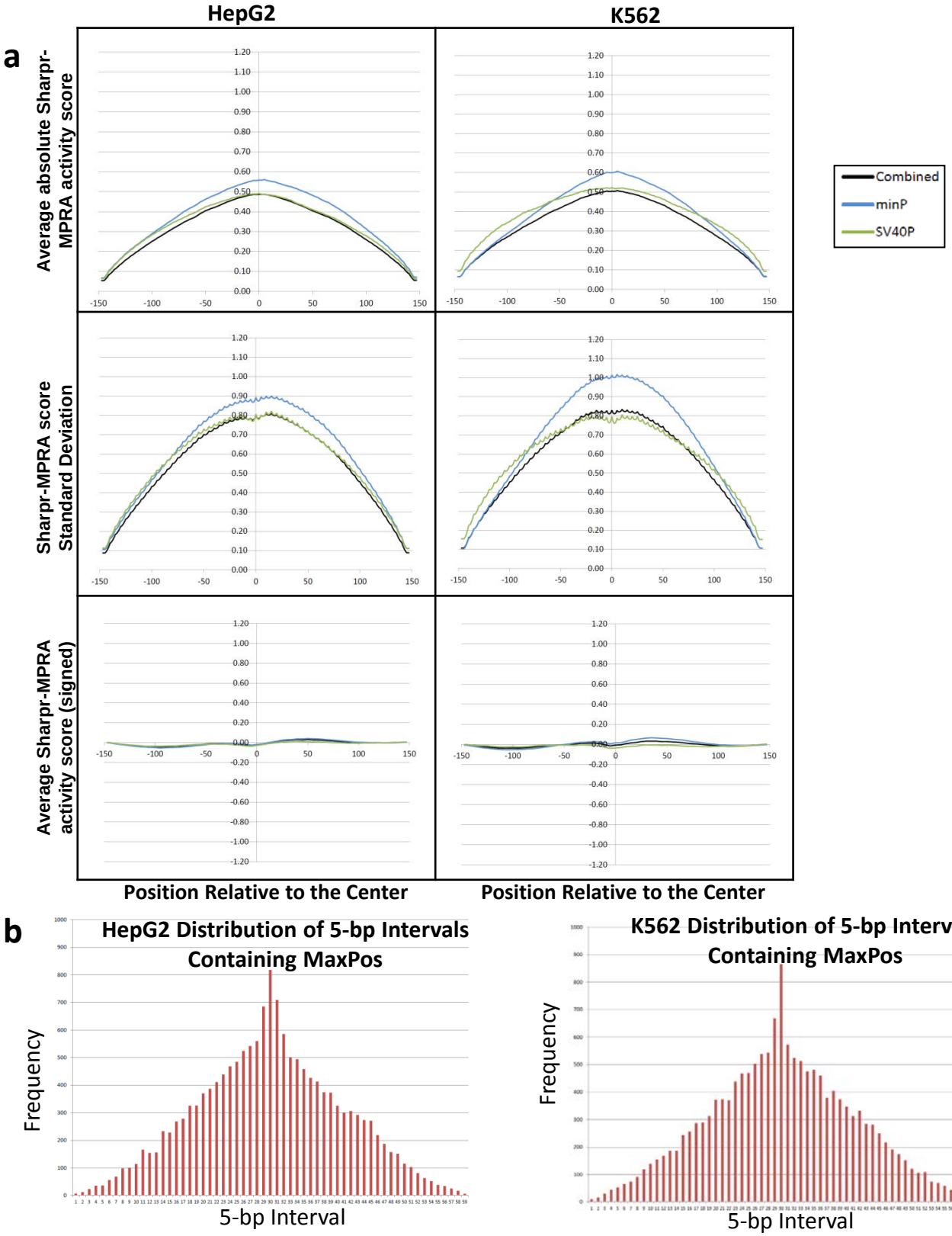
Supplementary Figure 18(a). Comparison with DeltaSVM predictions. (a) Overlap with Nucleotide Positions predicted by DeltaSVM to contain Top 1% Negative Mutations to Reference Sequence. Analogous plots to **Fig. 3c, Supplementary Figs. 13, 16-17**, but the enrichment is based on the 1% nucleotide positions tested that are predicted by DeltaSVM[14] to have a possible mutation to the reference sequence that would cause the greatest decrease in being regulatory (see Methods). Overlaps are shown for HepG2 (left) and K562 (right) for the combined SV40P and minP data (top row), minP only (middle row), and SV40P only (bottom row). An enrichment for DeltaSVM predictions is seen for nucleotides among our most activating predicted bases, but not our most repressive. **(b)** Next Page.



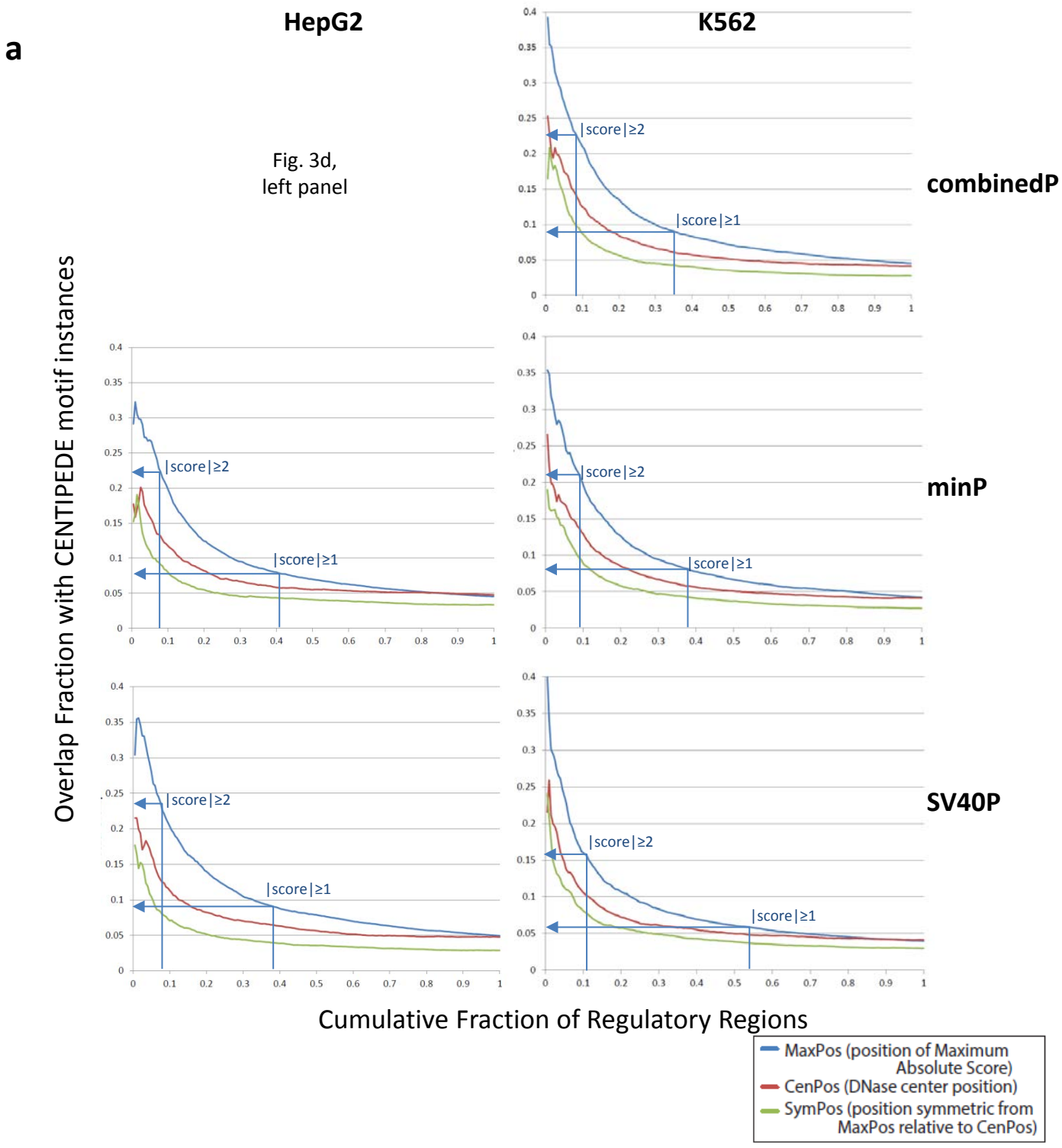
Supplementary Figure 18(b). Comparison with DeltaSVM predictions. (a) Previous page. **(b) Overlap with Nucleotides Predicted by DeltaSVM to contain Top 1% Positive Mutations to the Reference Sequence**. Analogous to (a), but based on the top 1% nucleotides having a possible mutation to the reference sequence leading to the greatest predicted increase in the sequence being regulatory (see Methods). Overlap is shown for HepG2 (left) and K562 (right), for the combined minP and SV40P data (top row), the minP data only (middle row), and the SV40P data only (bottom row). A depletion is seen in our most activating predicted bases.



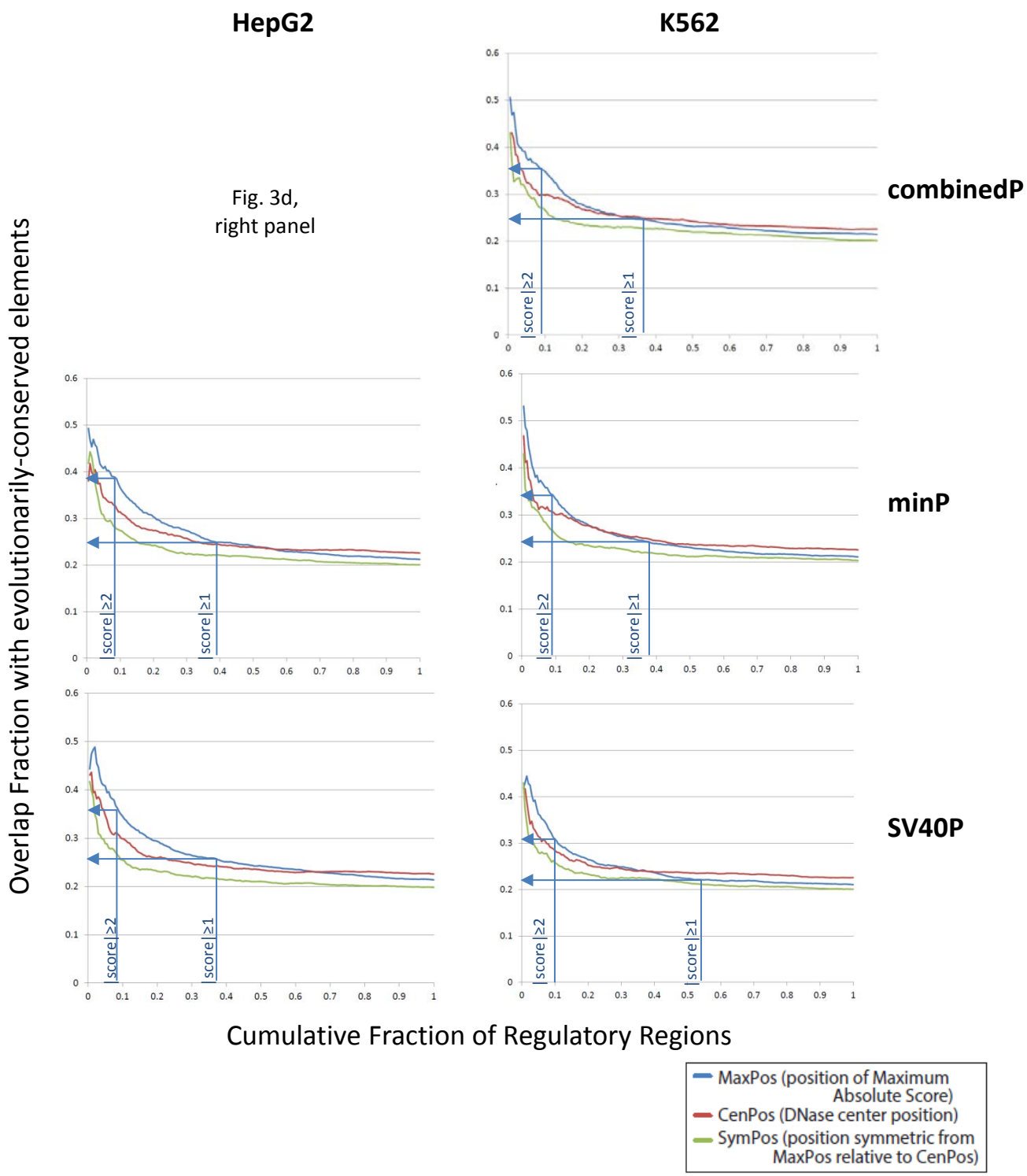
Supplementary Figure 19 – Enrichment Based on Maximum Absolute Activity Position. These are similar plots to those shown in **Fig. 3c** and **Supplementary Figs. 13 and 17** except the scatter plots are now based on only the single position which had the highest absolute Sharpr-MPRA regulatory activity score in each region tested (MaxPos nucleotides). The sign of the score is preserved for the analysis. In these plots there are 200 quantiles, so each point corresponds to around 79 unique MaxPos nucleotides. The green horizontal line represents the overlap fraction based on the center position and the black horizontal line overall overlap fraction of MaxPos nucleotides. The plots are **(a)** for transcription factor binding sites predicted by the CENTIPEDE method[8] for HepG2 (left) and K562 (right) and **(b)** for conserved elements detected by the SiPhy-PI method[42,65] for HepG2 (left) and K562 (right).



Supplementary Figure 20 – Sharpr-MPRA activity score and standard deviation by position. (a) Top row: Central positions show higher average absolute activity score for HepG2 (left) and K562 (right) for the minP data (blue), SV40P data (green), and combinedP data (black). Middle row: Central positions also show higher standard deviation of activity score. Bottom row: Average signed activity does not show a bias for central positions, as expected when averaging both positive and negative values. **(b)** Distribution of the location of maximum absolute regulatory activity (MaxPos) in each region among the 59 five-nucleotide intervals, for HepG2 (left) and in K562 (right).

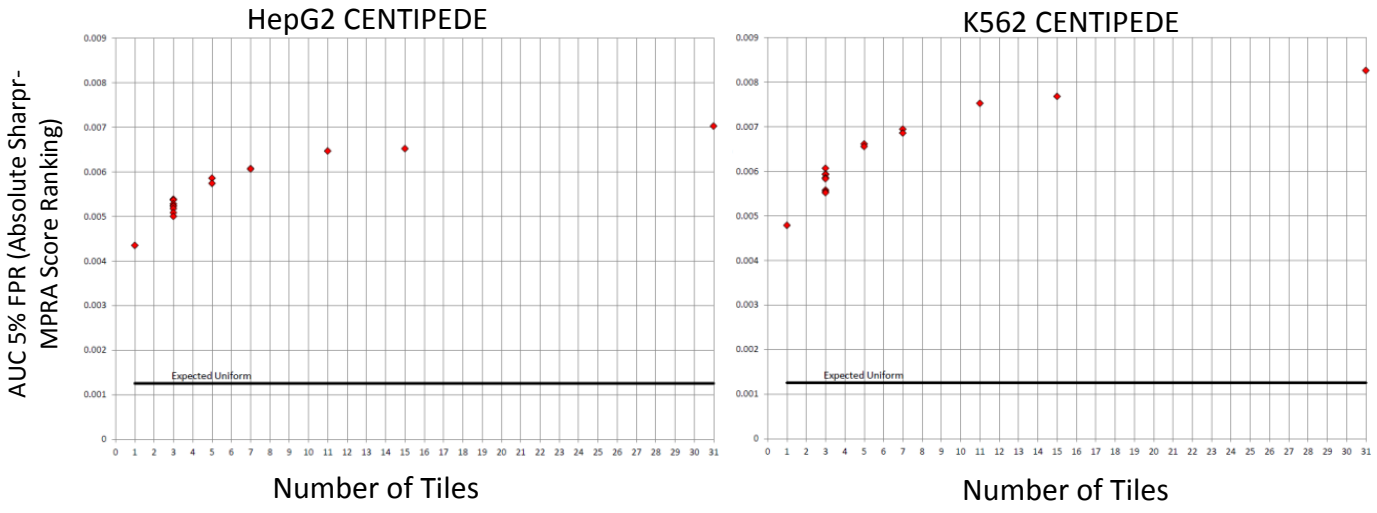


Supplementary Figure 21 – Ranked MaxPos overlap with CENTIPEDE motifs and evolutionarily-conserved nucleotides. (a) The three-way comparison of the cumulative overlap (y-axis) with CENTIPEDE[8] predicted transcription factor binding sites analogous to **Fig. 3d** (left panel) for HepG2, and shown here also for K562 for center nucleotide positions (CenPos, red), maximum-absolutely-score nucleotide positions (MaxPos, blue), and their symmetric nucleotide positions (SymPos, green), each ranked from highest to lowest based on the absolute Sharpr-MPRA scores (x-axis) for the (top) combinedP, (middle) minP, and (bottom) SV40P data indicating that our inference strategy captures regulatory nucleotides at high resolution. MaxPos, CenPos, and SymPos nucleotide positions are illustrated in the example of **Fig. 3b**. **(b)** next page.

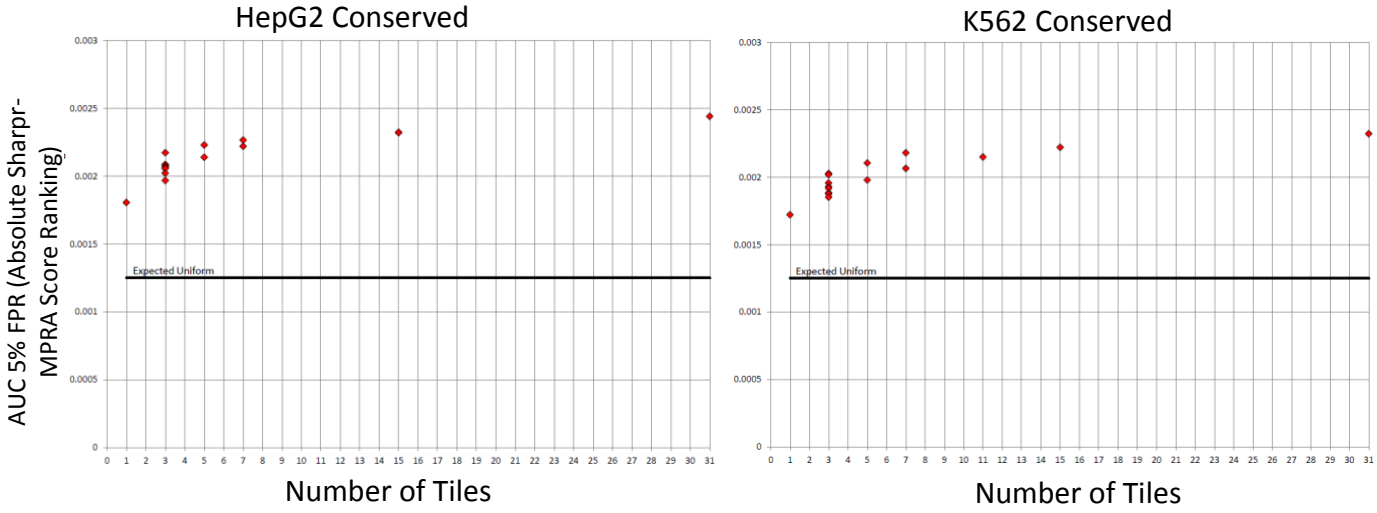


Supplementary Figure 21 – Ranked MaxPos overlap with CENTIPEDE motifs and evolutionarily-conserved nucleotides. (a) Previous page. **(b)** Overlap with conserved elements called by the SiPhy-PI method[42,65] analogous to **Fig. 3d** (right panel) for HepG2, and shown here also for K562 for center nucleotide positions (CenPos, red), maximum-absolutely-score nucleotide positions (MaxPos, blue), and their symmetric nucleotide positions (SymPos, green), each ranked from highest to lowest based on the absolute Sharpr-MPRA scores (x-axis) for the (top) combinedP, (middle) minP, and (bottom) SV40P data. MaxPos, CenPos, and SymPos nucleotide positions are illustrated in the example of **Fig. 3b**. The plots indicate that our inference strategy captures functional nucleotides at high resolution. The MaxPos nucleotides have higher overlap with conserved elements than CenPos nucleotides except at low absolute activity scores (**Supplementary Fig. 7c**).

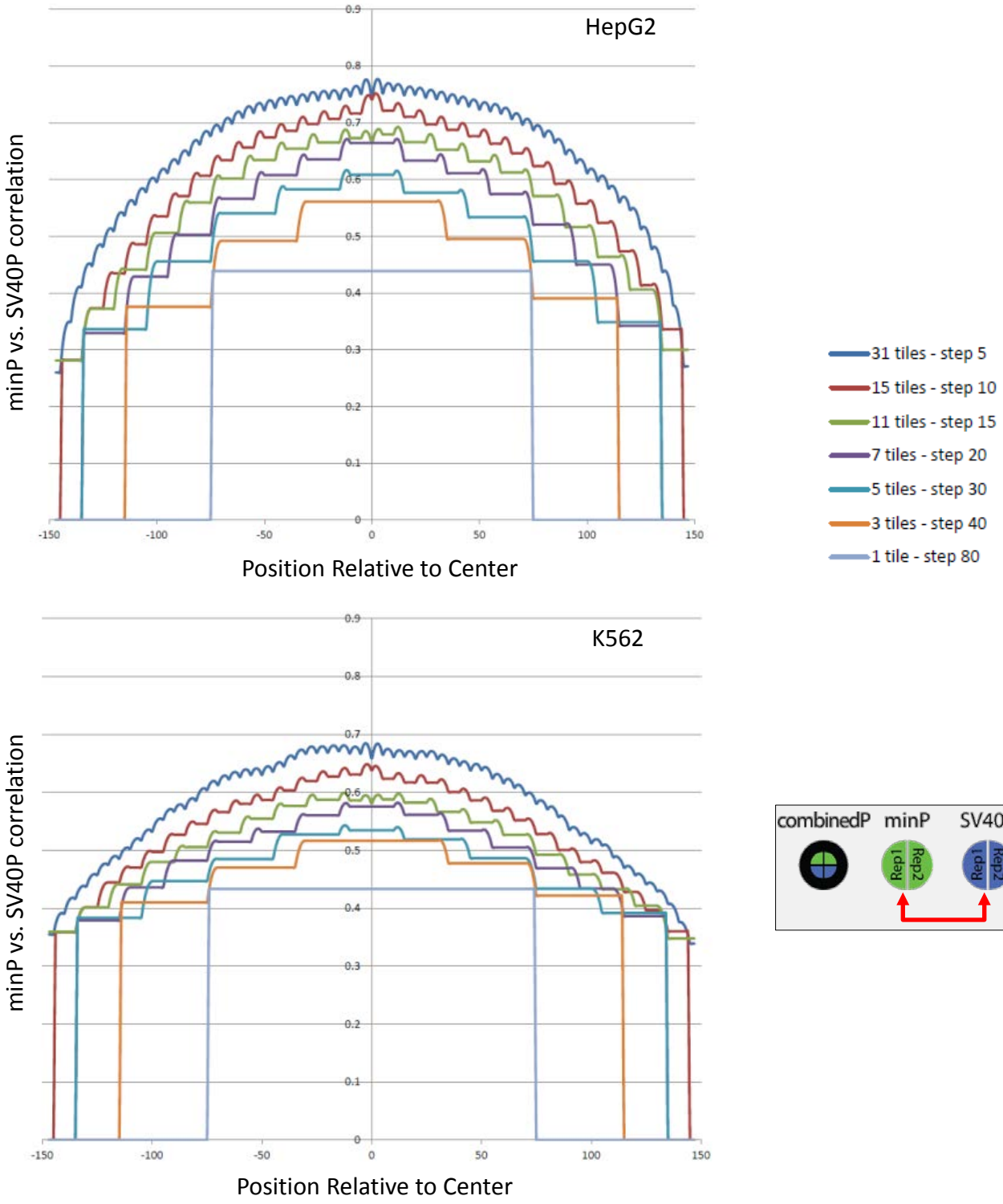
a. Motif recovery with varying numbers of tiles



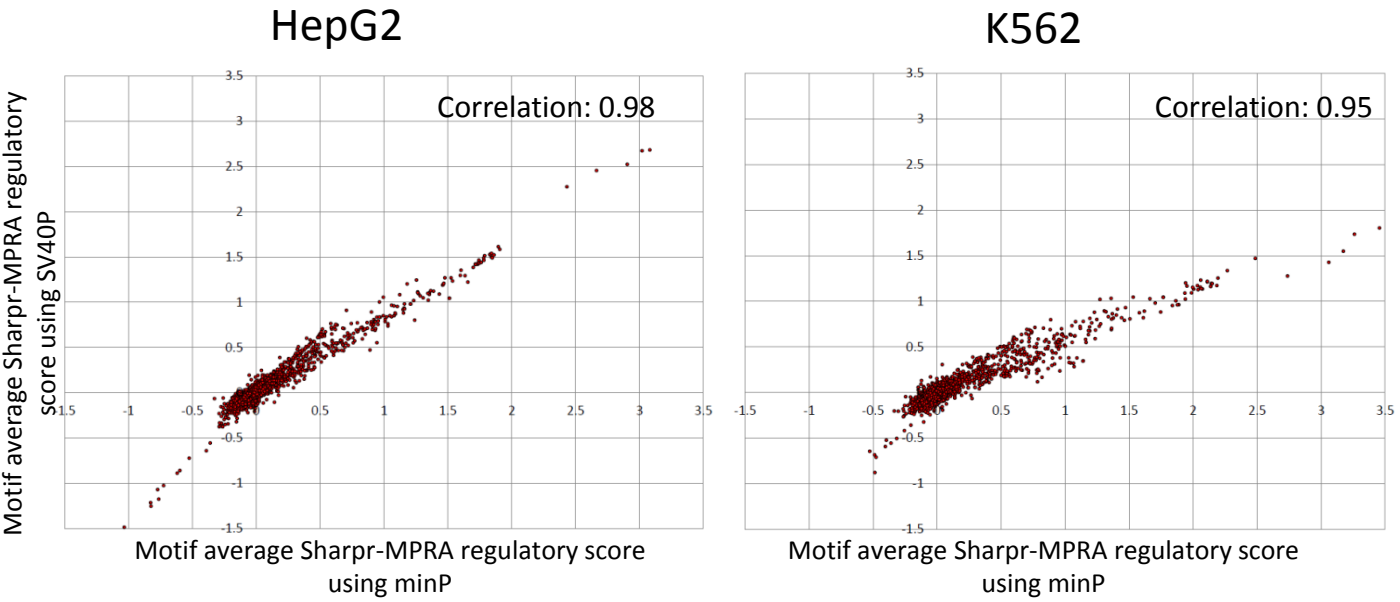
b. Evolutionarily-conserved element recovery with varying numbers of tiles



Supplementary Figure 22 – Impact of Number of Tiles and Tiling Interval on Motif and Conserved Element Recovery. Recovery of **(a)** CENTIPEDE[8] motif instances and **(b)** evolutionarily-conserved elements from the SiPhy-PI method[42,65] based on the Area under the ROC curve (AUC) up to a false positive rate (FPR) of 5% (y-axis), as a function of the number of tiles (x-axis), when ranking nucleotides by absolute Sharpr-MPRA regulatory activity score in HepG2 (left) and K562 (right). Multiple points for the same vertical line correspond to different step sizes leading to the same number of tiles within the 295bp region tested.



Supplementary Figure 23 – Impact of Number of Tiles and Tiling Interval on Correlation. Correlation between the minP and SV40P experiments (y-axis) at different positions relative to the center (x-axis) for regulatory activity inferred using only a subset of tiles selected by varying the step size (colors) for HepG2 (top) and K562 (bottom). In all cases the center tile is retained. If two or more step sizes led to the same number of tiles within the 295bp region tested, then only the correlations based on the smallest step size is plotted. We find that increasing the number of tiles (i.e. decreasing the step size) leads to increased correlation levels supporting the high-density tiling approach.



Supplementary Figure 24 – Motif average Sharpr-MPRA regulatory score concordance between minP and SV40P data. Scatter plot of the motif average Sharpr-MPRA regulatory obtained by averaging the activity at the central motif position for all motif instances using the minP scores (x-axis) or the SV40P scores (y-axis), for HepG2 (left) and K562 (right). Correlation between minP and SV40P motif scores is 0.98 in HepG2 and 0.95 in K562.

a. Motifs with strongest difference in activation or in repression for HepG2 vs. K562

b	c	d	e	f	g	h	i	k	m	-log10(j)	n	o	p	q	r	s	t	u	v	w
	Summary			Average Score Analysis				Pdiff	-log10 Pdiff		Activating/Repressive Enrichment Analysis						Avg Activity in Matched			
	AvgScore	Enriched in |score|≥1		Avg Score							Act Enr (-logP)		Repr Enr (-logP)		Enr HepG2-K562		HepG2match		K562match	
Motif	SigHighe	HepG2	K562	N	HepG2	K562	Diff	corrP	corr	uncorr	HepG2	K562	HepG2	K562	ActDiff	ReprDiff	N	Avg	N	Avg
TP53_7	HepG2	Activating	Neither	83	1.26	-0.10	1.36	4.7E-04	3.3	6.6	17.7	0.4	0.5	1.2	17.3	-0.7	21	2.113	17	-0.17
HNF4_known9	HepG2	Activating	Neither	393	0.60	-0.07	0.67	1.8E-25	24.7	28.0	41.3	0.0	0.0	1.0	41.3	-1.0	245	0.754	44	0.12
RXRA_disc1	HepG2	Activating	Neither	281	0.38	-0.03	0.41	2.8E-08	7.5	10.8	14.5	0.2	0.4	0.4	14.2	0.0	148	0.559	44	0.17
HNF1B_3	HepG2	Activating	Neither	172	0.27	-0.11	0.38	2.0E-06	5.7	9.0	4.9	0.0	0.0	0.1	4.9	-0.1	97	0.357	20	-0.05
PPARA_4	HepG2	Activating	Neither	323	0.37	0.00	0.37	4.1E-07	6.4	9.7	12.7	0.2	0.1	0.4	12.5	-0.4	157	0.589	46	0.15
NFKB_known2	HepG2	Activating	Neither	177	0.33	-0.03	0.36	4.2E-04	3.4	6.7	6.8	0.5	0.1	0.1	6.4	0.1	46	0.757	24	0.13
HNF1A_4	HepG2	Activating	Neither	204	0.26	-0.09	0.35	7.6E-07	6.1	9.4	5.1	0.0	0.0	0.1	5.1	0.0	111	0.363	22	-0.05
RXRG_3	HepG2	Activating	Neither	259	0.41	0.08	0.33	1.1E-04	4.0	7.3	12.6	1.2	0.1	0.5	11.5	-0.3	134	0.558	34	0.27
NFAT5_1	HepG2	Activating	Neither	236	0.27	-0.05	0.31	4.5E-05	4.3	7.6	5.9	0.1	0.3	0.2	5.8	0.1	57	0.755	47	0.04
RXRB_1	HepG2	Activating	Neither	252	0.40	0.09	0.31	2.3E-03	2.6	5.9	11.2	1.0	0.3	0.8	10.1	-0.5	133	0.559	37	0.25
HNF1_3	HepG2	Activating	Neither	271	0.16	-0.14	0.30	1.8E-05	4.7	8.0	2.5	0.0	0.3	0.1	2.5	0.1	126	0.361	41	-0.15
NR2C2_known1	HepG2	Activating	Neither	263	0.34	0.07	0.27	9.2E-03	2.0	5.3	8.3	0.9	0.7	1.0	7.4	-0.2	133	0.493	41	0.27
NR2F6_3	HepG2	Activating	Neither	253	0.33	0.07	0.26	6.2E-03	2.2	5.5	8.2	1.0	0.3	0.3	7.2	0.0	135	0.459	39	0.27
TCF7L1_2	HepG2	Activating	Neither	330	0.10	-0.11	0.20	8.6E-04	3.1	6.4	3.6	0.0	0.1	0.6	3.5	-0.5	152	0.319	51	0.01
TCF7_2	HepG2	Activating	Neither	315	0.11	-0.09	0.19	1.2E-03	2.9	6.2	4.3	0.0	0.1	0.5	4.2	-0.4	142	0.322	58	0.14
EP300_disc3	HepG2	Activating	Neither	541	0.06	-0.10	0.16	4.8E-05	4.3	7.6	3.2	0.0	0.3	0.3	3.2	0.0	259	0.162	66	-0.02
RFX5_known6	HepG2	Repressive	Neither	147	-0.34	-0.03	0.31	6.0E-05	4.2	7.5	0.3	0.8	18.6	0.9	-0.5	17.7	42	-0.32	18	0.04
RFX2_1	HepG2	Repressive	Neither	140	-0.33	-0.03	0.30	2.1E-04	3.7	7.0	0.3	0.9	16.8	1.0	-0.5	15.9	37	-0.341	17	0.1
RFX3_2	HepG2	Repressive	Neither	158	-0.31	-0.02	0.29	6.2E-05	4.2	7.5	0.5	1.1	17.8	0.8	-0.6	17.0	43	-0.288	20	0.12
BCL6B_1	HepG2	Repressive	Neither	131	-0.17	0.09	0.26	1.7E-03	2.8	6.1	0.0	0.6	2.5	0.4	-0.5	2.1	22	-0.095	22	0.28
MXI1_disc1	HepG2	Repressive	Neither	280	-0.28	-0.02	0.26	3.1E-07	6.5	9.8	0.4	0.8	30.5	1.8	-0.3	28.7	67	-0.252	38	0.17
BCL_disc3	HepG2	Repressive	Repressive	136	-1.02	-0.58	0.43	3.9E-04	3.4	6.7	0.2	0.1	43.8	39.9	0.2	3.9	42	-0.618	28	-0.78
REST_disc1	HepG2	Repressive	Repressive	216	-0.92	-0.49	0.43	1.2E-08	7.9	11.2	0.0	0.0	59.7	47.6	0.0	12.1	45	-0.892	49	-0.48
SIN3A_disc1	HepG2	Repressive	Repressive	217	-0.88	-0.46	0.42	1.0E-08	8.0	11.3	0.2	0.2	59.6	44.9	0.0	14.7	46	-0.772	54	-0.41
SREBP_disc1	HepG2	Repressive	Repressive	254	-0.22	0.00	0.22	1.6E-02	1.8	5.1	1.1	1.7	27.9	2.1	-0.6	25.7	64	-0.203	40	0.13
MYC_disc4	HepG2	Repressive	Repressive	234	-0.31	-0.12	0.19	7.5E-03	2.1	5.4	0.0	0.1	22.6	2.9	0.0	19.8	56	-0.196	35	-0.01
FOXA_known2	HepG2	Neither	Neither	603	0.03	-0.12	0.15	1.2E-05	4.9	8.2	1.4	0.0	0.1	0.3	1.4	-0.2	260	0.128	79	-0.06
GATA_known14	K562	Neither	Activating	284	-0.12	0.36	0.48	3.6E-11	10.4	13.7	0.1	13.6	1.2	0.2	-13.5	1.0	39	0.041	152	0.39
TAL1_disc1	K562	Neither	Activating	293	-0.14	0.34	0.47	7.7E-12	11.1	14.4	0.2	12.5	0.9	0.2	-12.3	0.6	41	0.003	163	0.37
CCNT2_disc1	K562	Neither	Activating	313	-0.11	0.33	0.43	3.0E-10	9.5	12.8	0.3	9.8	1.0	0.5	-9.6	0.6	53	0.103	158	0.36
STAT_disc1	K562	Neither	Activating	279	0.01	0.36	0.35	4.9E-07	6.3	9.6	1.1	9.8	0.6	0.1	-8.7	0.6	57	0.143	96	0.69
KLF13_1	K562	Neither	Activating	142	0.18	0.43	0.26	2.6E-03	2.6	5.9	1.5	7.2	0.5	0.0	-5.7	0.4	36	0.239	33	0.48
KLF12_2	K562	Activating	Activating	230	0.19	0.52	0.32	3.3E-09	8.5	11.8	4.3	19.6	0.9	0.1	-15.4	0.8	52	0.035	65	0.64
SP4_2	K562	Activating	Activating	534	0.36	0.67	0.31	7.3E-21	20.1	23.4	24.0	68.7	1.0	0.0	-44.7	1.0	134	0.345	137	0.82
KLF14_1	K562	Activating	Activating	326	0.29	0.60	0.30	7.6E-12	11.1	14.4	10.3	37.7	1.3	0.2	-27.4	1.1	87	0.181	86	0.82
IRF_disc4	K562	Activating	Activating	879	0.35	0.64	0.30	1.6E-33	32.8	36.1	33.4	104.0	0.7	0.0	-70.6	0.7	228	0.237	231	0.71
KLF7_1	K562	Activating	Activating	534	0.24	0.52	0.28	2.6E-19	18.6	21.9	15.0	47.8	1.9	0.1	-32.9	1.7	131	0.162	142	0.68
KLF4_1	K562	Activating	Activating	638	0.22	0.48	0.26	5.6E-22	21.3	24.5	12.9	50.1	0.8	0.1	-37.1	0.7	171	0.122	170	0.63
SP1_known4	K562	Activating	Activating	904	0.32	0.56	0.24	5.0E-26	25.3	28.6	31.5	84.7	0.9	0.0	-53.2	0.9	244	0.193	231	0.65
TATA_disc4	K562	Activating	Activating	747	0.37	0.59	0.22	1.1E-15	15.0	18.3	30.9	73.6	0.9	0.1	-42.6	0.8	204	0.257	194	0.71
PAX5_disc5	K562	Activating	Activating	152	0.47	0.67	0.20	1.6E-02	1.8	5.1	16.4	22.0	0.7	0.3	-5.5	0.5	43	0.296	35	0.81
KLF16_1	K562	Activating	Activating	648	0.21	0.41	0.20	6.4E-14	13.2	16.5	9.6	39.3	0.6	0.3	-29.7	0.3	161	0.114	155	0.56
TRIM28_disc1	K562	Activating	Activating	652	0.51	0.71	0.20	3.4E-04	3.5	6.8	66.8	88.4	0.4	0.0	-21.6	0.4	167	0.9	152	0.95
BACH2_1	K562	Activating	Activating	418	0.51	0.71	0.20	2.2E-02	1.6	4.9	35.0	51.5	0.0	0.0	-16.6	0.0	117	0.78	94	0.86
NRF1_disc3	K562	Activating	Activating	453	0.18	0.37	0.19	8.9E-10	9.0	12.3	5.6	26.0	1.1	0.2	-20.4	0.9	137	0.069	104	0.41
ATF3_disc4	K562	Activating	Activating	433	0.27	0.45	0.18	1.8E-07	6.8	10.0	12.5	29.5	1.2	0.0	-17.0	1.2	117	0.096	102	0.54
AP1_disc3	K562	Activating	Activating	773	0.42	0.60	0.17	2.1E-03	2.7	6.0	56.0	83.2	0.0	0.0	-27.2	0.0	194	0.787	150	0.84
TCF7L2_disc1	K562	Activating	Activating	819	0.58	0.75	0.17	1.7E-03	2.8	6.0	90.8	111.1	0.0	0.0	-20.3	0.0	213	0.944	169	0.99
PRDM1_disc2	K562	Activating	Activating	816	0.44	0.61	0.17	1.1E-03	3.0	6.2	65.2	98.1	0.0	0.0	-32.9	0.0	204	0.776	161	0.92
SP8_1	K562	Activating	Activating	461	0.22	0.38	0.16	8.5E-05	4.1	7.4	9.0	26.1	0.4	0.2	-17.0	0.2	120	0.12	108	0.54
TFAP2_disc1	K562	Activating	Activating	830	0.69	0.85	0.16	1.1E-02	1.9	5.2	113.5	128.2	0.0	0.0	-14.8	0.0	208	1.108	199	0.99
E2F_disc7	K562	Activating	Activating	504	0.27	0.42	0.15	3.9E-04	3.4	6.7	13.1	35.4	1.4	0.1	-22.2	1.3	132	0.112	124	0.55
CHD2_disc3	K562	Activating	Activating	366	0.21	0.35	0.14	4.5E-03	2.3	5.6	6.4	15.0	1.0	0.2	-8.5	0.9	97	0.062	78	0.65
CTCF_known2	K562	Activating	Activating	939	0.16	0.28	0.12	1.0E-08	8.0	11.3	5.9	20.3	0.5	0.0	-14.4	0.5	265	0.184	231	0.37
ETS_disc9	K562	Activating	Activating	742	0.20	0.32	0.12	2.2E-05	4.7	8.0	11.8	25.1	0.7	0.0	-13.3	0.7	214	0.121	153	0.48
EGR1_disc6	K562	Activating	Activating	668	0.20	0.31	0.12	3.5E-04	3.5	6.7	11.1	19.0	0.9	0.0	-7.9	0.9	196	0.147	151	0.37
SRF_disc2	K562	Activating	Activating	502	0.15	0.26	0.11	1.1E-02	1.9	5.2	4.8	14.4	0.7	0.1	-9.6	0.6	147	0.063	117	0.34
YY1_disc3	K562	Activating	Activating	527	0.23	0.34	0.11	3.0E-02	1.5	4.8	7.1	18.6	0.8	0.1	-11.5	0.7	158	0.181	118	0.45
SMC3_disc1	K562	Activating	Activating	1472	0.14	0.22	0.09	1.3E-05	4.9	8.2	7.4	18.8	0.7	0.0	-11.4	0.7	402	0.15	335	0.3
RAD21_disc1	K562	Activating	Activating	1539	0.12	0.21	0.08	5.3E-06	5.3	8.6	5.5	18.5	0.7	0.0	-13.0	0.7	431	0.142	370	0.3
HDAC2_disc6	K562	Neither	Neither	3205	-0.07	-0.03	0.04	4.4E-05	4.4	7.6	0.0	0.0	0.0	0.0	0.0	0.0	736	0.014	681	0.03

Supplementary Figure 25 – Top Differential Motifs and Motifs with Most Significant Activation or Repression Enrichment. (a,b) For both panels the column headers (see also Supplementary Table 2) are: (b) Motif logo and name (TF family and id within family); motif names with a 'disc' were based on de novo discovery in ENCODE ChIP-seq data[13] (c) cell type of most significant enrichment (Act and Rep Enr columns); (d-e) classification as Activating (Act, -$\log_{10}$P≥2), Repressive (Repr, -$\log_{10}$P≥2), Dual (both), Neither (neither) based on Act and Rep enrichment (Enr); (f) total number of motif instances; (g-i) average combinedP score at center motif position in HepG2 and K562 and their difference; (k) t-test corrected P-value of difference in activity; (m) -$\log_{10}$P of t-test corrected (for 1934 motifs); (n-q) -$\log_{10}$P of enrichment in activating (Act, activity≥1, blue) or repressive (Repr, activity≤-1, red) nucleotides for HepG2 and K562; (r-s) difference in activating (Act) or repressive (Repr) -$\log_{10}$P of enrichment (Orange=more activating/repressive in HepG2, Green=more activating/repressive in K562); (t-w) number of instances selected based on chromatin data in the cell type tested and the average activity of those instances for HepG2 and then K562. Only top-ranked member of each 'TF group' are shown. **(a)** Motifs with most significant activity difference between HepG2 and K562 (Pdiff). Top-ranked motifs determined by t-test of difference in activity. Display Cutoff: t-test corrected p-value≤0.05. **(b)** (next 3 pages) Motifs with most significant (P<0.01) enrichment in activating (score≥1) or repressive (≤-1) bases in K562 or in HepG2. Rows sorted by max absolute activity enrichment of (n-q) (af). Columns reordered to match the information highlighted. N/S denotes difference between HepG2 and K562 scores not significant (corr P>0.05). Full list in **Supplementary Table 2**.

b. Motifs with strongest activation or repression enrichments in either HepG2 or in K562 (page 1 of 3)

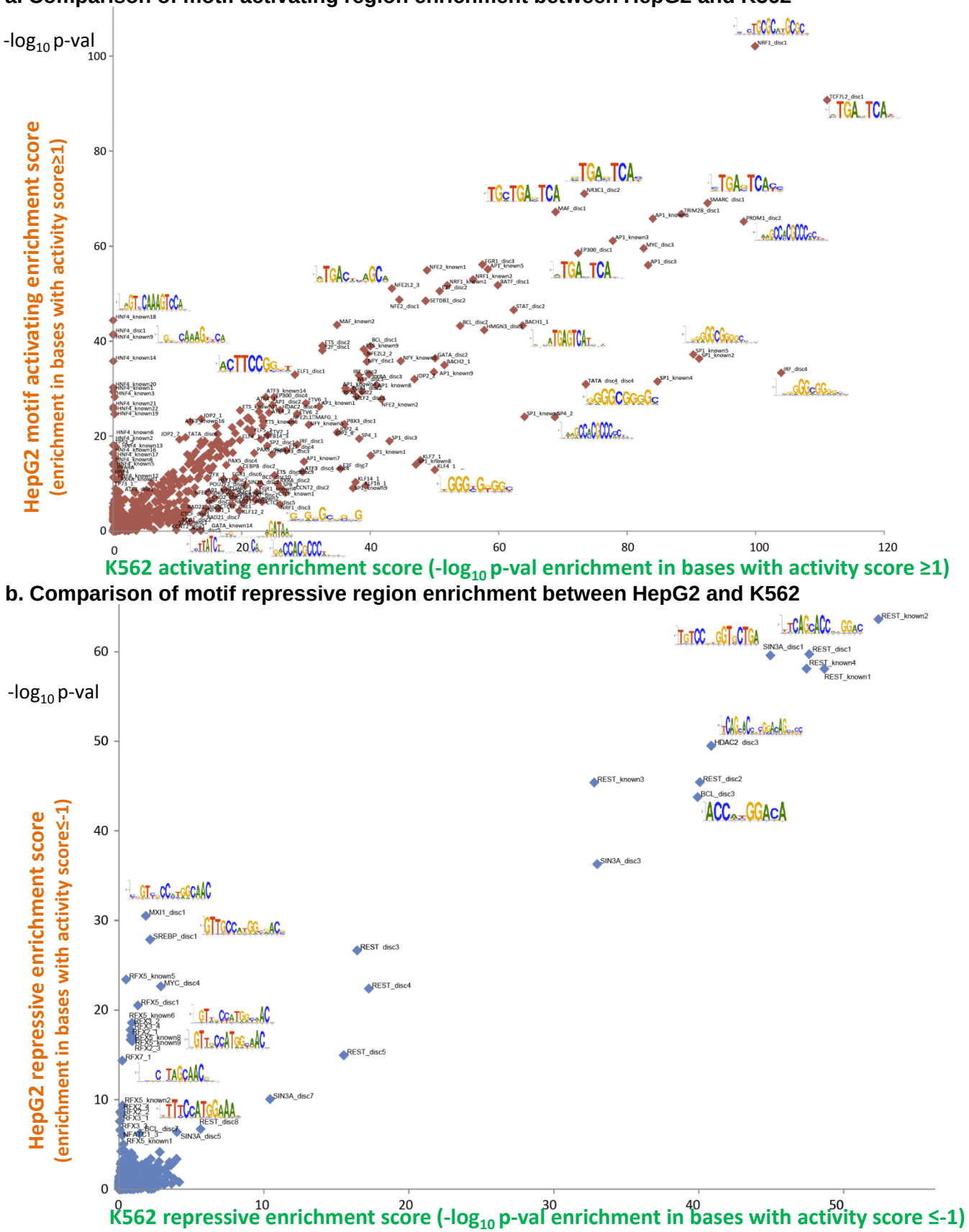
	b	af	d	e	n	o	p	q	r	s	f	g	h	i	m	c	t	u	v	w
		For	Summary		Activating/Repressive Enrichment Analysis						Average Score Analysis						Avg Activity in Matched			
		Sorting	Enrichment in |score|≥1		Act Enr (-logP)		Repr Enr (-logP)		Enr HepG2-K562		Avg Score				Pdiff		HepG2match		K562match	
	Motif	MaxAct	HepG2	K562	HepG2	K562	HepG2	K562	ActDiff	ReprDiff	N	HepG2	K562	Diff	-lg10Cor	Higher	N	Avg	N	Avg
	TFAP2_disc1	128.24	Activating	Activating	113.5	128.2	0.0	0.0	-14.8	0.0	830	0.69	0.85	0.16	1.9	K562	208	1.108	199	0.99
	TCF7L2_disc1	111.11	Activating	Activating	90.8	111.1	0.0	0.0	-20.3	0.0	819	0.58	0.75	0.17	2.8	K562	213	0.944	169	0.99
	IRF_disc4	103.96	Activating	Activating	33.4	104.0	0.7	0.0	-70.6	0.7	879	0.35	0.64	0.30	32.8	K562	228	0.237	231	0.71
	NRF1_disc1	102.12	Activating	Activating	102.1	99.9	0.2	0.0	2.2	0.1	249	1.35	1.44	0.09	0.0	N/S	46	1.196	53	1.46
	PRDM1_disc2	98.14	Activating	Activating	65.2	98.1	0.0	0.0	-32.9	0.0	816	0.44	0.61	0.17	3.0	K562	204	0.776	161	0.92
	SMARC_disc1	92.50	Activating	Activating	69.1	92.5	0.0	0.0	-23.4	0.0	729	0.58	0.71	0.14	0.6	N/S	175	0.88	164	0.89
	SP1_known2	91.19	Activating	Activating	36.3	91.2	0.8	0.0	-54.9	0.8	1070	0.32	0.54	0.22	23.8	K562	273	0.233	270	0.65
	TRIM28_disc1	88.44	Activating	Activating	66.8	88.4	0.4	0.0	-21.6	0.4	652	0.51	0.71	0.20	3.5	K562	167	0.9	152	0.95
	AP1_known6	83.98	Activating	Activating	65.8	84.0	0.1	0.0	-18.2	0.1	637	0.61	0.68	0.06	0.0	N/S	150	1.063	140	0.82
	MYC_disc3	82.56	Activating	Activating	59.6	82.6	0.3	0.0	-23.0	0.3	748	0.45	0.58	0.13	0.4	N/S	199	0.878	162	0.76
	TATA_disc4	73.56	Activating	Activating	30.9	73.6	0.9	0.1	-42.6	0.8	747	0.37	0.59	0.22	15.0	K562	204	0.257	194	0.71
	NR3C1_disc2	73.32	Activating	Activating	71.1	73.3	0.0	0.0	-2.3	0.0	717	0.56	0.66	0.10	0.0	N/S	145	0.748	174	0.77
	EP300_disc1	72.36	Activating	Activating	58.5	72.4	0.3	0.0	-13.8	0.3	751	0.45	0.56	0.11	0.0	N/S	205	0.888	152	0.83
	MAF_disc1	68.86	Activating	Activating	67.2	68.9	0.1	0.0	-1.6	0.0	447	0.68	0.78	0.10	0.0	N/S	112	1.094	119	0.96
	SP4_2	68.75	Activating	Activating	24.0	68.7	1.0	0.0	-44.7	1.0	534	0.36	0.67	0.31	20.1	K562	134	0.345	137	0.82
	BACH1_1	63.72	Activating	Activating	43.3	63.7	0.1	0.0	-20.4	0.1	618	0.47	0.57	0.10	0.0	N/S	161	0.846	131	0.75
	REST_known2	63.63	Repressive	Repressive	0.3	0.2	63.6	52.4	0.0	11.2	132	-1.26	-0.68	0.58	6.2	HepG2	36	-0.894	32	-0.73
	STAT_disc2	62.34	Activating	Activating	46.6	62.3	0.2	0.0	-15.8	0.2	698	0.36	0.50	0.15	1.2	N/S	189	0.727	146	0.84
	BATF_disc1	59.87	Activating	Activating	51.8	59.9	0.1	0.0	-8.1	0.1	624	0.50	0.61	0.11	0.0	N/S	135	0.954	122	0.87
	SIN3A_disc1	59.59	Repressive	Repressive	0.2	0.2	59.6	44.9	0.0	14.7	217	-0.88	-0.46	0.42	8.0	HepG2	46	-0.772	54	-0.41
	HMGN3_disc1	57.75	Activating	Activating	42.4	57.8	0.1	0.3	-15.4	-0.1	493	0.59	0.70	0.11	0.0	N/S	124	0.911	114	0.85
	EGR1_disc3	57.47	Activating	Activating	56.2	57.5	0.2	0.0	-1.3	0.2	153	1.26	1.40	0.14	0.0	N/S	30	0.914	40	1.45
	NFE2_known1	54.94	Activating	Activating	54.9	48.8	0.6	0.0	6.1	0.6	458	0.56	0.62	0.06	0.0	N/S	103	1.01	118	0.76
	BCL_disc2	54.00	Activating	Activating	43.2	54.0	0.5	0.0	-10.8	0.5	715	0.34	0.43	0.09	0.0	N/S	189	0.728	147	0.73
	BACH2_1	51.54	Activating	Activating	35.0	51.5	0.0	0.0	-16.6	0.0	418	0.51	0.71	0.20	1.6	K562	117	0.78	94	0.86
	NFE2L2_3	51.11	Activating	Activating	51.1	43.4	0.1	0.1	7.7	-0.1	391	0.65	0.64	0.01	0.0	N/S	83	1.187	125	0.61
	E2F_disc2	50.81	Activating	Activating	50.6	50.8	0.3	0.1	-0.3	0.2	121	1.33	1.43	0.10	0.0	N/S	23	1.137	27	1.55
	GATA_disc2	50.12	Activating	Activating	36.4	50.1	0.1	0.0	-13.7	0.0	510	0.38	0.49	0.10	0.0	N/S	106	0.782	109	0.73
	KLF4_1	50.07	Activating	Activating	12.9	50.1	0.8	0.1	-37.1	0.7	638	0.22	0.48	0.26	21.3	K562	171	0.122	170	0.63
	HDAC2_disc3	49.50	Repressive	Repressive	0.5	0.6	49.5	40.9	-0.1	8.6	243	-0.63	-0.33	0.29	2.5	HepG2	68	-0.31	60	-0.43
	SETDB1_disc2	48.63	Activating	Activating	48.5	48.6	0.2	0.1	-0.1	0.1	143	1.19	1.32	0.13	0.0	N/S	32	1.018	30	1.27
	KLF7_1	47.82	Activating	Activating	15.0	47.8	1.9	0.1	-32.9	1.7	534	0.24	0.52	0.28	18.6	K562	131	0.162	142	0.68
	JDP2_2	46.89	Activating	Activating	32.0	46.9	0.1	0.0	-14.9	0.1	502	0.41	0.50	0.10	0.0	N/S	126	0.78	102	0.83
	NFY_known3	44.75	Activating	Activating	35.8	44.7	0.7	0.0	-8.9	0.6	218	0.61	0.75	0.14	0.0	N/S	44	0.592	45	0.66
	HNF4_known18	44.41	Activating	Neither	44.4	0.0	0.1	0.8	44.4	-0.8	241	0.82	-0.14	0.97	20.6	HepG2	166	0.971	25	-0.05
	RXRA_disc3	39.83	Activating	Activating	33.0	39.8	0.5	0.0	-6.8	0.5	566	0.32	0.41	0.09	0.0	N/S	153	0.708	112	0.72
	KLF16_1	39.29	Activating	Activating	9.6	39.3	0.6	0.3	-29.7	0.3	648	0.21	0.41	0.20	13.2	K562	161	0.114	155	0.56
	ETS_disc2	39.03	Activating	Activating	39.0	32.6	0.2	0.1	6.5	0.0	74	1.70	1.81	0.11	0.0	N/S	15	2.229	11	2.02
	MEF2_disc3	37.88	Activating	Activating	28.3	37.9	0.1	0.0	-9.5	0.1	620	0.38	0.42	0.04	0.0	N/S	154	0.789	125	0.45
	KLF14_1	37.72	Activating	Activating	10.3	37.7	1.3	0.2	-27.4	1.1	326	0.29	0.60	0.30	11.1	K562	87	0.181	86	0.82
	RFX5_disc2	36.24	Activating	Activating	29.5	36.2	0.5	0.2	-6.7	0.3	201	0.54	0.68	0.15	0.0	N/S	45	0.342	46	0.69
	PBX3_disc1	35.99	Activating	Activating	22.5	36.0	0.4	0.0	-13.5	0.4	216	0.47	0.62	0.15	0.5	N/S	45	0.475	50	0.48
	ELF1_disc1	32.96	Activating	Activating	33.0	28.3	0.2	0.1	4.7	0.0	75	1.48	1.57	0.10	0.0	N/S	15	1.859	13	1.34
	MXI1_disc1	30.52	Repressive	Neither	0.4	0.8	30.5	1.8	-0.3	28.7	280	-0.28	-0.02	0.26	6.5	HepG2	67	-0.252	38	0.17
	ETV6_1	29.98	Activating	Activating	26.9	30.0	0.1	0.2	-3.0	-0.1	170	0.78	0.97	0.19	0.0	N/S	26	1.192	51	0.91
	ATF3_disc4	29.47	Activating	Activating	12.5	29.5	1.2	0.0	-17.0	1.2	433	0.27	0.45	0.18	6.8	K562	117	0.096	102	0.54
	SREBP_disc1	27.87	Repressive	Repressive	1.1	1.7	27.9	2.1	-0.6	25.7	254	-0.22	0.00	0.22	1.8	HepG2	64	-0.203	40	0.13
	NFE2L1::MAFG_1	27.73	Activating	Activating	23.0	27.7	0.1	0.1	-4.8	0.0	254	0.50	0.58	0.08	0.0	N/S	59	0.784	93	0.76
	CCNT2_disc2	26.81	Activating	Activating	8.1	26.8	1.3	0.0	-18.7	1.3	1248	0.13	0.23	0.10	5.9	K562	320	0.116	284	0.42
	ATF4_3	26.73	Activating	Activating	26.7	22.9	0.3	0.2	3.9	0.1	112	1.10	1.16	0.06	0.0	N/S	37	1.319	32	1.22
	SP8_1	26.09	Activating	Activating	9.0	26.1	0.4	0.2	-17.0	0.2	461	0.22	0.38	0.16	4.1	K562	120	0.12	108	0.54
	ZNF143_disc3	24.70	Activating	Activating	16.2	24.7	0.9	1.3	-8.5	-0.4	554	0.30	0.34	0.04	0.0	N/S	146	0.381	105	0.42
	ETV7_1	24.04	Activating	Activating	20.2	24.0	0.0	0.0	-3.8	0.0	250	0.59	0.57	0.03	0.0	N/S	46	0.629	61	0.73
	SP2_disc1	23.99	Activating	Activating	18.0	24.0	0.5	0.1	-6.0	0.5	187	0.46	0.56	0.10	0.0	N/S	45	0.507	42	0.57
	NR2C2_disc1	23.79	Activating	Activating	23.8	20.8	0.2	0.4	3.0	-0.2	70	1.25	1.34	0.10	0.0	N/S	19	1.209	13	1.25
	ELF5_3	23.36	Activating	Activating	23.2	23.4	0.0	0.1	-0.2	-0.1	170	0.76	0.81	0.05	0.0	N/S	37	0.848	35	0.7
	CTCF_disc5	23.19	Activating	Activating	6.2	23.2	0.6	0.4	-17.0	0.2	1084	0.16	0.22	0.06	0.0	N/S	304	0.097	213	0.33
	ZBTB14_3	22.98	Activating	Activating	19.3	23.0	0.8	0.1	-3.6	0.7	99	0.66	0.90	0.24	0.2	N/S	33	0.494	13	0.73
	PAX5_disc5	21.96	Activating	Activating	16.4	22.0	0.7	0.3	-5.5	0.5	152	0.47	0.67	0.20	1.8	K562	43	0.296	35	0.81
	HEY1_disc1	21.86	Activating	Activating	21.9	18.5	0.1	0.5	3.4	-0.5	151	0.68	0.61	0.07	0.0	N/S	35	0.614	35	0.62
	ERF_1	21.49	Activating	Activating	21.5	20.9	0.3	0.1	0.6	0.3	105	1.19	1.12	0.07	0.0	N/S	18	0.613	26	1.32
	EHF_1	20.64	Activating	Activating	20.6	19.0	0.2	0.1	1.6	0.0	79	1.08	1.20	0.12	0.0	N/S	15	1.628	19	1.44
	CEBPB_disc2	19.80	Activating	Activating	13.0	19.8	0.7	0.3	-6.8	0.4	155	0.40	0.52	0.12	0.0	N/S	36	0.239	29	0.64
	ELF4_1	19.73	Activating	Activating	19.2	19.7	0.1	0.3	-0.6	-0.2	87	1.00	1.10	0.10	0.0	N/S	17	1.304	19	1.13
	KLF12_2	19.63	Activating	Activating	4.3	19.6	0.9	0.1	-15.4	0.8	230	0.19	0.52	0.32	8.5	K562	52	0.035	65	0.64
	ELK3_1	19.17	Activating	Activating	19.2	16.3	0.7	0.6	2.9	0.1	53	1.68	1.68	0.00	0.0	N/S	8	1.235	12	1.81
	CREB3_1	18.96	Activating	Activating	19.0	14.3	0.2	0.4	4.7	-0.2	76	1.12	0.94	0.19	0.0	N/S	19	1.064	15	0.69
	SMC3_disc1	18.83	Activating	Activating	7.4	18.8	0.7	0.0	-11.4	0.7	1472	0.14	0.22	0.09	4.9	K562	402	0.15	335	0.3
	BRCA1_disc1	18.64	Activating	Activating	18.6	18.3	0.5	1.1	0.3	-0.6	25	2.84	2.49	0.35	0.0	N/S	4	3.917	9	3.16
	YY1_disc3	18.59	Activating	Activating	7.1	18.6	0.8	0.1	-11.5	0.7	527	0.23	0.34	0.11	1.5	K562	158	0.181	118	0.45
	RAD21_disc1	18.53	Activating	Activating	5.5	18.5	0.7	0.0	-13.0	0.7	1539	0.12	0.21	0.08	5.3	K562	431	0.142	370	0.3
	ELK4_2	18.41	Activating	Activating	18.4	16.8	0.8	0.6	1.7	0.1	51	1.68	1.66	0.01	0.0	N/S	7	1.101	13	1.45
	FLI1_3	18.41	Activating	Activating	18.4	16.8	0.8	0.6	1.7	0.1	51	1.68	1.66	0.01	0.0	N/S	7	1.101	13	1.45
	HF1H3B_1	18.19	Activating	Activating	7.9	18.2	0.1	0.0	-10.3	0.0	1102	0.14	0.19	0.04	0.0	N/S	290	0.159	243	0.34
	RFX3_2	17.79	Repressive	Neither	0.5	1.1	17.8	0.8	-0.6	17.0	158	-0.31	-0.02	0.29	4.2	HepG2	43	-0.288	20	0.12
	TP53_7	17.73	Activating	Neither	17.7	0.4	0.5	1.2	17.3	-0.7	83	1.26	-0.10	1.36	3.3	HepG2	21	2.113	17	-0.17
	ERG_3	16.91	Activating	Activating	16.9	15.3	0.8	0.6	1.6	0.1	52	1.60	1.58	0.01	0.0	N/S	7	1.101	14	1.37
	RFX2_1	16.81	Repressive	Neither	0.3	0.9	16.8	1.0	-0.5	15.9	140	-0.33	-0.03	0.30	3.7	HepG2	37	-0.341	17	0.1
	ETV3_1	16.68	Activating	Activating	16.7	15.1	0.7	0.2	1.6	0.5	53	1.49	1.52	0.03	0.0	N/S	9	1.341	11	1.36
	ETV1_1	16.38	Activating	Activating	16.4	14.8	0.8	0.7	1.6	0.1	49	1.63	1.62	0.01	0.0	N/S	8	1.026	13	1.69
	CTCFL_disc1	16.33	Activating	Activating	4.8	16.3	0.1	0.2	-11.5	-0.1	472	0.24	0.34	0.10	0.1	N/S	129	0.219	95	0.54
	SIX5_disc1	16.16	Activating	Activating	15.6	16.2	0.2	0.1	-0.6	0.0	77	0.84	0.86	0.01	0.0	N/S	23	0.841	18	0.46
	FEV_2	15.92	Activating	Activating	15.9	14.4	0.8	0.2	1.5	0.6	51	1.57	1.60	0.02	0.0	N/S	8	1.026	13	1.71
	ATF7_1	15.85	Activating	Activating	15.9	12.0	0.3	0.2	3.9	0.1	46	1.25	1.10	0.14	0.0	N/S	11	1.017	10	0.79
	ETV4_2	15.61	Activating	Activating	15.6	14.1	0.8	0.2	1.6	0.6	47	1.58	1.62	0.04	0.0	N/S	8	1.026	12	1.83
	TFEB_1	15.50	Activating	Activating	15.5	15.1	0.1	0.7	0.4	-0.6	84	0.92	0.75	0.17	0.0	N/S	27	0.767	9	2.02
	CHD2_disc3	14.99	Activating	Activating	6.4	15.0	1.0	0.2	-8.5	0.9	366	0.21	0.35	0.14	2.3	K562	97	0.062	78	0.65

b. Motifs with strongest activation or repression enrichments in either HepG2 or in K562 (page 2 of 3)

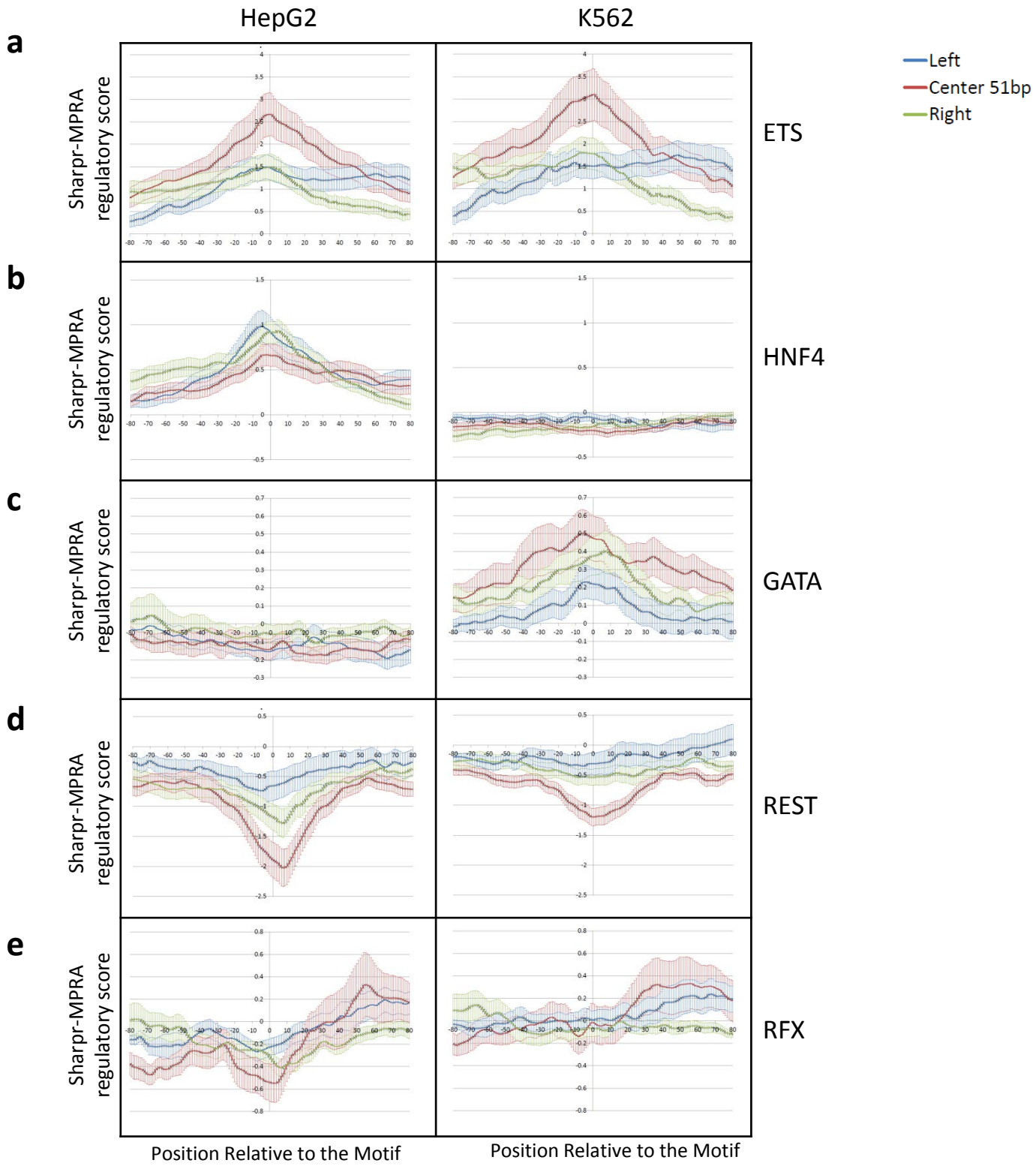
Motif	For Sorting	Summary		Activating/Repressive Enrichment Analysis						Average Score Analysis						Avg Activity in Matched			
		Enrichment in |score|≥1		Act Enr (-logP)		Repr Enr (-logP)		Enr HepG2-K562		Avg Score			Pdiff			HepG2match		K562match	
	MaxAct	HepG2	K562	HepG2	K562	HepG2	K562	ActDiff	ReprDiff	N	HepG2	K562	Diff	-lg10Cor	Higher	N	Avg	N	Avg
XBP1_3	14.81	Activating	Activating	14.8	12.0	0.4	0.3	2.8	0.1	40	1.24	1.12	0.12	0.0	N/S	6	0.922	11	0.56
TFE3_1	14.75	Activating	Activating	14.8	13.3	0.2	0.1	1.4	0.1	69	0.87	0.78	0.09	0.0	N/S	22	0.943	6	1.73
POU2F2_disc2	14.73	Activating	Activating	8.1	14.7	0.5	0.3	-6.6	0.2	962	0.13	0.18	0.06	0.0	N/S	239	0.119	195	0.21
ZFX_1	14.67	Activating	Activating	11.2	14.7	0.0	0.5	-3.5	-0.5	416	0.31	0.28	0.03	0.0	N/S	114	0.286	82	0.46
ELF3_3	14.48	Activating	Activating	14.5	12.3	0.1	0.5	2.2	-0.4	114	0.73	0.75	0.02	0.0	N/S	28	0.812	21	0.77
SRF_disc2	14.42	Activating	Activating	4.8	14.4	0.7	0.1	-9.6	0.6	502	0.15	0.26	0.11	1.9	K562	147	0.063	117	0.34
RFX7_1	14.36	Repressive	Neither	1.2	1.1	14.4	0.2	0.1	14.1	111	-0.21	0.06	0.28	1.2	N/S	26	-0.379	18	0.17
CREB5_1	14.09	Activating	Activating	14.1	11.8	0.2	0.1	2.3	0.0	80	0.85	0.76	0.09	0.0	N/S	21	0.608	17	0.5
ZNF281_1	13.92	Activating	Activating	3.7	13.9	0.0	0.0	-10.2	0.0	1198	0.12	0.19	0.06	0.9	N/S	317	0.172	242	0.31
ETV5_1	13.64	Activating	Activating	13.6	13.3	0.3	0.7	0.3	-0.4	45	1.55	1.56	0.01	0.0	N/S	6	1.455	11	1.19
ETV2_1	12.96	Activating	Activating	13.0	10.9	0.3	0.2	2.1	0.1	112	0.78	0.66	0.11	0.0	N/S	15	0.732	16	1.08
PPARA_4	12.72	Activating	Neither	12.7	0.2	0.1	0.4	12.5	-0.4	323	0.37	0.00	0.37	6.4	HepG2	157	0.589	46	0.15
RXRG_3	12.62	Activating	Neither	12.6	1.2	0.1	0.5	11.5	-0.3	259	0.41	0.08	0.33	4.0	HepG2	134	0.558	34	0.27
TAL1_disc1	12.51	Neither	Activating	0.2	12.5	0.9	0.2	-12.3	0.6	293	-0.14	0.34	0.47	11.1	K562	41	0.003	163	0.37
SIRT6_disc1	11.93	Activating	Activating	11.9	8.6	0.3	0.6	3.3	-0.3	54	0.98	0.69	0.29	0.0	N/S	16	0.969	7	1.15
SPI1_disc1	11.67	Activating	Activating	8.3	11.7	0.1	0.1	-3.4	0.0	276	0.26	0.32	0.07	0.0	N/S	43	0.141	94	0.37
ARNTL_1	11.61	Activating	Activating	11.6	9.3	0.3	0.2	2.3	0.1	56	0.91	0.73	0.19	0.0	N/S	17	0.804	9	1.43
YY2_2	11.52	Activating	Activating	11.5	9.1	0.3	0.7	2.4	-0.4	44	0.87	1.06	0.19	0.0	N/S	10	0.696	9	0.96
RXRB_1	11.17	Activating	Neither	11.2	1.0	0.3	0.8	10.1	-0.5	252	0.40	0.09	0.31	2.6	HepG2	133	0.559	37	0.25
BHLHE40_disc1	11.14	Activating	Activating	11.1	10.8	0.2	0.2	0.3	0.1	66	0.78	0.73	0.05	0.0	N/S	20	0.65	8	1.41
TP73_1	10.87	Activating	Neither	10.9	0.4	0.6	0.9	10.5	-0.3	68	1.03	-0.08	1.11	1.2	N/S	18	2.158	14	-0.18
TFEC_1	10.78	Activating	Activating	10.8	8.8	0.1	1.2	2.0	-1.0	84	0.69	0.53	0.16	0.0	N/S	26	0.833	10	0.67
ATF2_2	10.40	Activating	Activating	10.4	9.0	0.4	0.3	1.4	0.1	38	0.75	1.03	0.28	0.0	N/S	6	0.458	13	0.77
ZBTB7A_disc2	9.82	Activating	Activating	4.4	9.8	0.4	0.3	-5.4	0.1	98	0.29	0.43	0.14	0.0	N/S	16	0.24	25	0.57
BHLHE41_2	9.49	Activating	Activating	9.5	9.3	0.4	0.4	0.2	0.1	31	1.06	0.93	0.13	0.0	N/S	10	0.921	2	2.78
ZBTB33_disc1	9.31	Activating	Activating	9.3	9.2	0.8	1.7	0.2	-0.9	13	2.71	2.35	0.36	0.0	N/S	4	1.922	3	2.62
WT1_1	8.46	Activating	Activating	2.4	8.5	0.0	0.0	-6.1	0.0	513	0.13	0.23	0.09	1.0	N/S	137	0.188	108	0.26
ELF2_1	8.40	Activating	Activating	5.1	8.4	0.3	0.4	-3.3	-0.1	372	0.17	0.16	0.01	0.0	N/S	73	0.227	65	0.59
SPDEF_5	8.33	Activating	Activating	8.3	8.1	1.1	0.4	0.2	0.8	31	1.02	1.00	0.03	0.0	N/S	6	1.169	7	1.39
NR2F6_3	8.25	Activating	Neither	8.2	1.0	0.3	0.3	7.2	0.0	253	0.33	0.07	0.26	2.2	HepG2	135	0.459	39	0.27
NFKB_known9	8.13	Activating	Neither	8.1	0.9	0.1	0.1	7.2	0.1	175	0.36	0.02	0.35	2.6	HepG2	47	0.814	24	0.16
CREB3L1_4	8.05	Activating	Activating	8.0	3.4	0.2	0.4	4.6	-0.3	69	0.60	0.37	0.23	0.0	N/S	20	0.649	22	0.51
TP63_1	7.95	Activating	Neither	7.9	0.3	0.3	0.6	7.6	-0.3	54	0.99	-0.08	1.07	0.0	N/S	19	1.718	11	-0.12
ARNT_2	7.83	Activating	Activating	7.0	7.8	0.4	0.3	-0.8	0.1	39	0.76	0.60	0.17	0.0	N/S	14	1.234	4	0.37
MAFF_1	7.81	Dual	Activating	7.8	5.2	2.0	1.2	2.6	0.9	159	0.30	0.33	0.03	0.0	N/S	40	0.821	47	0.43
CREB3L2_2	7.41	Activating	Activating	5.0	7.4	0.3	0.2	-2.4	0.1	49	0.48	0.54	0.06	0.0	N/S	10	0.919	12	0.48
MYCN_2	7.37	Neither	Activating	1.4	7.4	0.4	0.3	-6.0	0.1	100	0.15	0.39	0.24	0.5	N/S	22	0.118	34	0.65
RARA_1	7.23	Activating	Neither	7.2	0.6	0.1	0.4	6.6	-0.3	240	0.27	0.03	0.24	0.9	N/S	107	0.43	53	0.21
KLF13_1	7.21	Neither	Activating	1.5	7.2	0.5	0.0	-5.7	0.4	142	0.18	0.43	0.26	2.6	K562	36	0.239	33	0.48
MLX_2	7.05	Activating	Activating	7.1	4.9	0.4	0.4	2.1	0.1	32	0.76	0.58	0.18	0.0	N/S	9	1.287	5	0.65
ATF6_1	6.73	Activating	Neither	6.7	1.8	0.2	0.2	5.0	0.1	57	0.43	0.29	0.14	0.0	N/S	17	0.551	14	0.42
HINFP_2	6.60	Activating	Activating	6.6	4.8	0.3	0.6	1.9	-0.4	50	0.48	0.48	0.00	0.0	N/S	9	0.874	12	0.17
ZNF263_disc1	6.40	Activating	Activating	2.6	6.4	0.9	0.4	-3.8	0.5	1013	0.06	0.10	0.04	0.0	N/S	221	0.085	221	0.26
PATZ1_1	6.37	Activating	Activating	4.1	6.4	0.1	0.0	-2.3	0.1	367	0.16	0.21	0.04	0.0	N/S	99	0.222	93	0.28
PAX4_1	6.33	Activating	Activating	6.3	4.7	0.6	0.9	1.7	-0.3	70	0.47	0.55	0.08	0.0	N/S	21	0.394	16	1.42
NFATC1_3	6.02	Repressive	Neither	0.2	0.0	6.0	0.2	0.2	5.8	113	-0.11	-0.09	0.01	0.0	N/S	29	0.113	22	0.18
HNF1B_4	5.88	Activating	Neither	5.9	0.0	0.0	0.1	5.9	0.0	198	0.28	-0.10	0.38	5.6	HepG2	111	0.388	20	-0.06
NFAT5_1	5.88	Activating	Neither	5.9	0.1	0.3	0.2	5.8	0.1	236	0.27	-0.05	0.31	4.3	HepG2	57	0.755	47	0.04
SPIB_2	5.77	Activating	Activating	2.6	5.8	0.1	0.3	-3.1	-0.2	360	0.09	0.24	0.14	0.4	N/S	54	0.258	136	0.23
MAFG_1	5.76	Dual	Activating	5.8	3.4	2.9	1.8	2.3	1.1	141	0.21	0.26	0.05	0.0	N/S	28	0.992	43	0.19
ZNF219_1	5.65	Neither	Activating	1.0	5.6	0.1	0.1	-4.6	0.0	534	0.11	0.14	0.03	0.0	N/S	140	0.148	108	0.21
CACD_1	5.58	Neither	Activating	1.9	5.6	0.4	1.7	-3.7	-1.3	222	0.09	0.13	0.04	0.0	N/S	51	0.22	48	0.35
E4F1_1	5.50	Activating	Activating	5.5	3.2	0.7	0.6	2.3	0.1	16	1.18	1.09	0.10	0.0	N/S	5	0.488	2	2.66
HNF1A_1	5.18	Activating	Neither	5.2	0.0	0.4	0.5	5.2	0.0	216	0.22	-0.14	0.36	5.2	HepG2	113	0.378	32	-0.03
HNF1_1	4.82	Activating	Neither	4.8	0.0	0.4	0.5	4.8	0.0	213	0.21	-0.15	0.35	4.5	HepG2	107	0.376	36	-0.12
NR4A_known4	4.75	Activating	Neither	4.7	0.8	0.4	0.1	3.9	0.3	215	0.20	0.02	0.19	0.0	N/S	93	0.457	41	0.18
CEBPD_1	4.75	Activating	Activating	2.6	4.7	0.1	0.1	-2.1	0.0	89	0.23	0.34	0.11	0.0	N/S	28	0.189	20	0.51
RARB_1	4.52	Activating	Neither	4.5	0.0	0.3	0.1	4.5	0.2	195	0.17	-0.02	0.19	0.0	N/S	84	0.362	40	0.03
EGR3_1	4.35	Activating	Activating	2.9	4.3	0.8	0.1	-1.5	0.7	96	0.24	0.40	0.16	0.0	N/S	20	0.271	27	0.64
MYB_1	4.33	Activating	Activating	4.3	3.0	0.5	0.1	1.4	0.4	78	0.25	0.34	0.09	0.0	N/S	16	0.138	19	0.17
TCF7_2	4.27	Activating	Neither	4.3	0.0	0.1	0.5	4.2	-0.4	315	0.11	-0.09	0.19	2.9	HepG2	142	0.322	58	0.14
EVX2_1	4.24	Repressive	Neither	0.4	1.0	4.2	0.2	-0.6	4.1	67	-0.16	-0.01	0.15	0.0	N/S	17	0.065	12	0.12
ESRRA_disc2	4.23	Activating	Neither	4.2	0.5	0.1	0.1	3.7	0.0	273	0.17	0.01	0.16	0.0	N/S	134	0.315	46	0.17
BDP1_disc1	4.16	Activating	Activating	3.0	4.2	0.5	1.0	-1.2	-0.5	438	0.10	0.08	0.02	0.0	N/S	103	0.049	95	0.24
ZEB1_known2	4.15	Repressive	Repressive	0.0	0.0	4.2	2.8	0.0	1.3	93	-0.25	-0.15	0.09	0.0	N/S	30	-0.228	16	-0.11
TFAP2B_4	4.15	Neither	Repressive	0.9	1.4	0.8	4.2	-0.5	-3.4	144	0.00	-0.09	0.10	0.0	N/S	33	0.07	28	-0.19
RARG_4	4.12	Activating	Neither	4.1	0.1	0.1	0.2	4.0	-0.1	180	0.20	0.03	0.17	0.0	N/S	77	0.217	35	0.01
GBX2_1	4.04	Repressive	Neither	0.1	0.2	4.0	0.8	-0.1	3.2	72	-0.23	-0.08	0.15	0.0	N/S	15	-0.038	12	-0.01
CACBP_1	3.98	Activating	Dual	3.3	4.0	0.3	2.7	-0.7	-2.4	843	0.09	0.06	0.03	0.0	N/S	218	0.137	179	0.21
MYOD1_2	3.97	Repressive	Repressive	0.0	0.0	3.4	4.0	0.0	-0.6	151	-0.18	-0.21	0.03	0.0	N/S	48	-0.11	25	-0.19
HERPUD1_1	3.93	Repressive	Neither	0.5	0.5	3.9	0.6	0.0	3.4	100	-0.11	-0.06	0.05	0.0	N/S	23	-0.087	13	0.24
EGR4_1	3.88	Neither	Activating	1.8	3.9	1.0	0.2	-2.1	0.8	117	0.16	0.26	0.10	0.0	N/S	26	0.381	37	0.45
FOXA_known5	3.87	Activating	Neither	3.9	0.0	0.3	0.3	3.9	0.0	471	0.09	-0.09	0.18	4.0	HepG2	227	0.18	67	0.01
ESRRB_1	3.87	Activating	Neither	3.9	1.6	0.1	0.1	2.3	0.0	215	0.18	0.10	0.09	0.0	N/S	111	0.332	37	0.29
ZBTB7C_1	3.79	Neither	Repressive	0.1	0.8	1.2	3.8	-0.7	-2.6	29	-0.18	-0.07	0.11	0.0	N/S	5	-0.223	7	0.04
LMO2_1	3.78	Repressive	Repressive	0.1	0.1	3.0	3.8	0.1	-0.8	243	-0.11	-0.11	0.00	0.0	N/S	59	-0.005	43	0.07
NR2F2_2	3.75	Activating	Neither	3.7	0.5	0.2	0.3	3.3	0.0	219	0.18	0.00	0.18	0.0	N/S	98	0.366	43	0.2
NANOG_disc3	3.63	Neither	Repressive	0.9	1.7	0.6	3.6	-0.7	-3.0	585	0.06	-0.03	0.08	0.0	N/S	134	0.138	107	0.13
GCM2_1	3.61	Activating	Dual	2.8	3.6	1.5	2.3	-0.8	-0.8	20	0.43	0.36	0.08	0.0	N/S	4	-0.16	4	-0.26
LHX8_1	3.61	Repressive	Neither	0.3	0.5	3.6	1.0	-0.2	2.6	60	-0.18	0.00	0.18	0.0	N/S	12	0.138	11	0.18
NKX2-8_2	3.56	Repressive	Neither	0.4	0.1	3.6	0.5	0.3	3.1	113	-0.12	-0.08	0.04	0.0	N/S	25	0.026	28	-0.08
TCF7L1_2	3.55	Activating	Neither	3.6	0.0	0.1	0.6	3.5	-0.5	330	0.10	-0.11	0.20	3.1	HepG2	152	0.319	51	0.01
GCM_1	3.43	Activating	Activating	3.4	2.6	0.3	0.6	0.8	-0.4	50	0.30	0.28	0.02	0.0	N/S	13	0.244	9	0.17
TCF3_3	3.40	Neither	Repressive	0.0	0.5	1.6	3.4	-0.5	-1.8	234	-0.08	-0.04	0.04	0.0	N/S	67	-0.034	39	0.2
TFE_1	3.39	Activating	Dual	3.4	2.0	0.0	2.0	1.4	-2.0	161	0.21	0.10	0.11	0.0	N/S	57	0.276	31	0.21
NFE2L1_1	3.37	Neither	Activating	1.1	3.4	0.3	0.2	-2.3	0.1	117	0.05	0.19	0.14	0.0	N/S	27	-0.024	39	0.42
ZBTB6_1	3.28	Repressive	Repressive	0.2	0.5	2.4	3.3	-0.3	-0.9	101	-0.09	-0.15	0.06	0.0	N/S	25	-0.076	19	0.07

b. Motifs with strongest activation or repression enrichments in either HepG2 or in K562 (page 3 of 3)

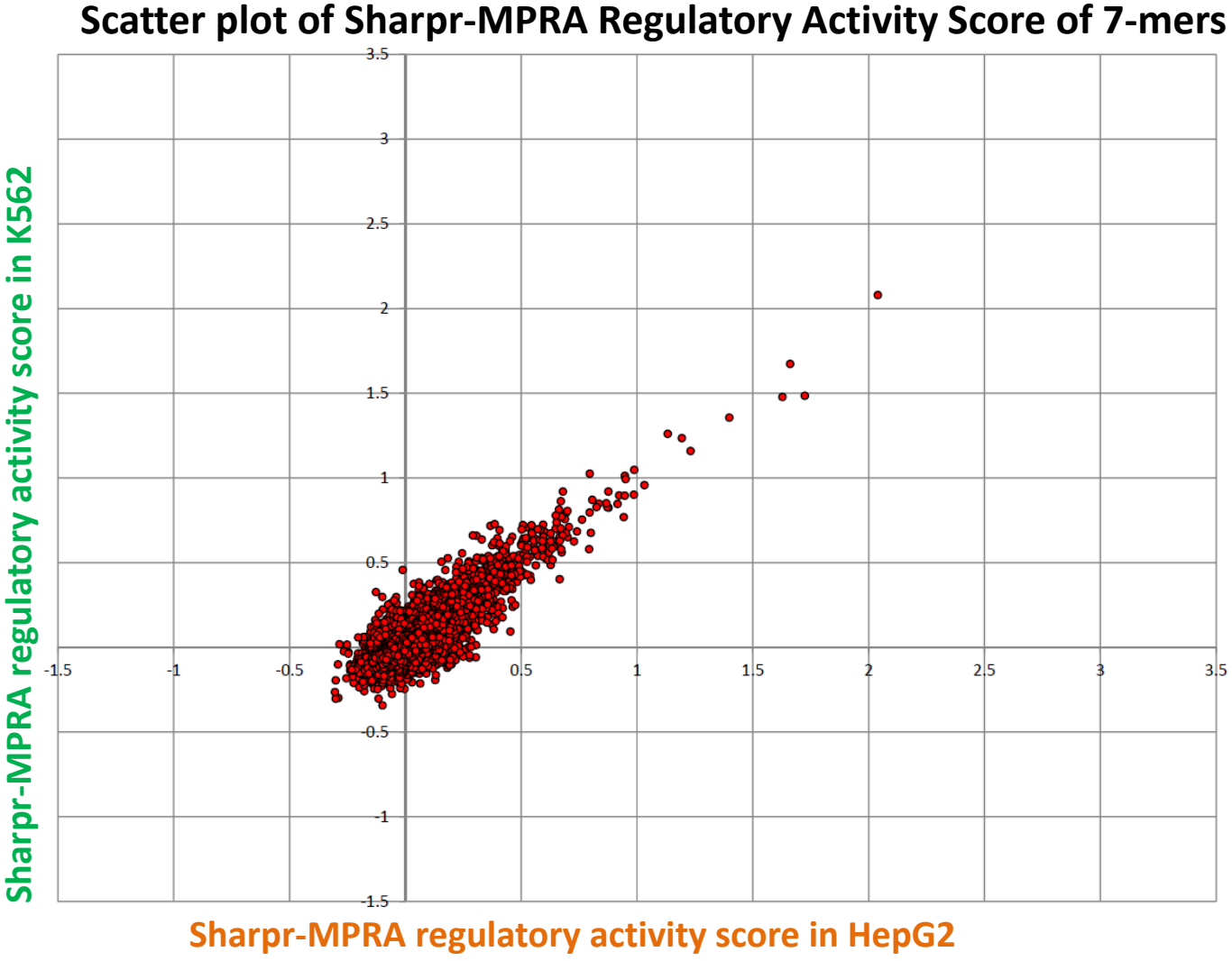
	b	af	d	e	n	o	p	q	r	s	f	g	h	i	m	c	t	u	v	w
		For Sorting	Summary		Activating/Repressive Enrichment Analysis						Average Score Analysis						Avg Activity in Matched			
			Enrichment in |score|≥1		Act Enr (-logP)		Repr Enr (-logP)		Enr HepG2-K562		Avg Score				Pdiff		HepG2match		K562match	
	Motif	MaxAct	HepG2	K562	HepG2	K562	HepG2	K562	ActDiff	ReprDiff	N	HepG2	K562	Diff	-lg10Cor	Higher	N	Avg	N	Avg
	DUX4_1	3.17	Activating	Activating	2.6	3.2	0.3	0.2	-0.5	0.1	52	0.32	0.37	0.06	0.0	N/S	13	0.324	10	0.46
	MSX1_1	3.15	Activating	Activating	3.1	2.0	0.9	0.8	1.1	0.1	10	1.33	0.81	0.52	0.0	N/S	4	0.528	2	0.7
	HIC1_1	3.14	Neither	Activating	1.8	3.1	0.3	0.4	-1.3	-0.1	176	0.11	0.11	0.00	0.0	N/S	47	0.133	26	0.44
	PAX6_2	3.11	Neither	Repressive	1.0	0.3	0.5	3.1	0.8	-2.6	190	0.08	0.03	0.05	0.0	N/S	50	0.1	56	0.05
	NKX2-2_1	3.09	Repressive	Neither	0.1	0.5	3.1	0.2	-0.5	2.9	133	-0.16	-0.03	0.14	0.0	N/S	26	-0.025	34	0.05
	RUNX3_1	3.06	Neither	Activating	1.7	3.1	0.5	0.1	-1.4	0.4	76	0.09	0.25	0.16	0.0	N/S	11	-0.052	22	0.33
	NKX2-3_3	3.03	Repressive	Neither	0.5	0.1	3.0	0.5	0.3	2.5	105	-0.14	-0.09	0.05	0.0	N/S	25	-0.187	33	-0.21
	TFCP2_3	2.98	Neither	Repressive	0.7	0.8	0.8	3.0	-0.1	-2.2	197	0.02	0.02	0.00	0.0	N/S	45	-0.052	38	0.08
	ZNF713_1	2.96	Repressive	Neither	0.2	0.1	3.0	0.2	0.0	2.7	173	-0.09	-0.02	0.07	0.0	N/S	33	-0.056	41	0.32
	MESP1_1	2.94	Repressive	Neither	0.1	0.0	2.9	1.8	0.0	1.1	109	-0.14	-0.14	0.00	0.0	N/S	28	-0.199	23	0.14
	HSFY2_3	2.90	Neither	Activating	1.1	2.9	0.8	0.7	-1.8	0.1	11	0.12	0.57	0.45	0.0	N/S	2	0.222	1	2.33
	TLX2_1	2.88	Activating	Activating	2.3	2.9	0.4	0.3	-0.6	0.1	36	0.36	0.20	0.17	0.0	N/S	6	0.546	8	-0.13
	SOX10_10	2.87	Repressive	Neither	0.2	0.3	2.9	0.3	-0.1	2.6	144	-0.07	-0.05	0.03	0.0	N/S	35	0.111	33	-0.02
	MYBL1_1	2.86	Neither	Activating	0.5	2.9	0.5	0.4	-2.4	0.1	27	0.10	0.32	0.22	0.0	N/S	7	0.12	9	0.45
	TCF4_2	2.84	Repressive	Repressive	0.0	0.0	2.1	2.8	0.0	-0.7	118	-0.18	-0.18	0.00	0.0	N/S	27	-0.084	31	-0.02
	VENTX_1	2.83	Neither	Repressive	0.3	0.3	0.8	2.8	0.0	-2.1	13	-0.21	-0.13	0.08	0.0	N/S	1	0.009	5	-0.28
	T_1	2.82	Repressive	Neither	0.0	0.0	2.8	0.1	0.0	2.7	84	-0.19	-0.10	0.09	0.0	N/S	23	-0.141	20	-0.01
	GLIS2_2	2.79	Neither	Activating	1.4	2.8	0.2	0.4	-1.4	-0.2	70	0.13	0.11	0.03	0.0	N/S	16	0.078	15	0.12
	THAP1_disc1	2.78	Activating	Neither	2.8	1.8	0.3	0.2	1.0	0.1	125	0.15	0.19	0.04	0.0	N/S	27	0.259	38	0.38
	ZBED1_1	2.78	Activating	Activating	2.8	2.7	1.1	1.0	0.1	0.1	6	1.53	1.45	0.09	0.0	N/S	3	-0.198	0	#N/A
	CLOCK::ARNTL_1	2.74	Activating	Activating	2.3	2.7	0.6	0.9	-0.4	-0.3	71	0.21	0.19	0.02	0.0	N/S	19	0.588	28	0.41
	MNT_1	2.74	Activating	Activating	2.3	2.7	0.2	0.4	-0.4	-0.2	71	0.11	0.18	0.07	0.0	N/S	13	0.195	22	0.46
	AHR_2	2.73	Repressive	Neither	0.7	1.1	2.7	1.6	-0.4	1.2	36	-0.15	-0.02	0.13	0.0	N/S	9	-0.374	2	-0.03
	PAX2_1	2.71	Activating	Neither	2.7	2.0	0.4	0.3	0.7	0.1	40	0.30	0.42	0.13	0.0	N/S	7	0.376	11	0.21
	SNAI2_1	2.68	Neither	Repressive	0.0	0.0	1.1	2.7	0.0	-1.6	154	-0.10	-0.23	0.14	0.0	N/S	41	0.005	25	-0.08
	GTF2I_1	2.67	Neither	Activating	0.8	2.7	0.1	0.1	-1.9	0.1	571	0.05	0.11	0.06	0.0	N/S	121	0.026	135	0.19
	HEY2_1	2.67	Activating	Activating	2.1	2.7	1.7	0.8	-0.6	0.9	39	0.31	0.36	0.06	0.0	N/S	10	0.199	11	0.75
	ID4_1	2.62	Repressive	Neither	0.0	0.1	2.6	1.6	0.0	1.0	124	-0.11	-0.17	0.06	0.0	N/S	43	-0.101	17	-0.06
	NFAT_2	2.61	Activating	Neither	2.6	0.2	0.2	0.2	2.4	0.0	234	0.14	-0.04	0.17	0.1	N/S	75	0.36	35	0.13
	HIF1A_1	2.60	Neither	Activating	1.7	2.6	0.8	0.7	-0.9	0.1	13	0.31	0.47	0.16	0.0	N/S	2	-0.058	2	-0.42
	FOXD1_1	2.60	Activating	Neither	2.6	0.0	0.1	0.1	2.6	0.0	330	0.08	-0.07	0.15	0.7	N/S	138	0.274	47	0.11
	PKNOX2_1	2.60	Neither	Repressive	1.2	0.6	0.5	2.6	0.6	-2.1	129	-0.03	-0.04	0.01	0.0	N/S	43	-0.019	25	0.07
	NKX3-2_1	2.58	Neither	Repressive	0.2	0.0	0.4	2.6	0.1	-2.2	102	-0.07	-0.15	0.08	0.0	N/S	21	0.146	29	0.07
	TBX5_2	2.58	Repressive	Neither	0.0	0.1	2.6	1.5	-0.1	1.1	65	-0.22	-0.14	0.08	0.0	N/S	14	-0.109	15	-0.16
	GCM1_1	2.56	Neither	Repressive	0.8	1.4	1.7	2.6	-0.6	-0.9	16	0.10	-0.14	0.24	0.0	N/S	4	-0.16	4	-0.26
	MYF6_1	2.50	Repressive	Neither	0.0	0.1	2.5	1.9	-0.1	0.6	202	-0.13	-0.13	0.01	0.0	N/S	57	-0.01	38	0.05
	ZNF423_1	2.50	Neither	Repressive	0.1	0.3	1.5	2.5	-0.2	-1.0	79	-0.03	-0.11	0.08	0.0	N/S	24	-0.052	14	-0.1
	CUX1_3	2.50	Neither	Activating	1.4	2.5	1.0	0.9	-1.1	0.1	7	0.24	0.34	0.10	0.0	N/S	3	0.455	2	0.08
	HOXA11_1	2.49	Repressive	Neither	0.2	0.0	2.5	0.5	0.2	2.0	68	-0.15	-0.02	0.13	0.0	N/S	20	-0.066	13	0.03
	BCL6B_1	2.48	Repressive	Neither	0.0	0.6	2.5	0.4	-0.5	2.1	131	-0.17	0.09	0.26	2.8	HepG2	22	-0.095	22	0.28
	NR5A1_2	2.46	Activating	Neither	2.5	0.3	0.8	0.8	2.1	0.0	241	0.10	-0.04	0.13	0.0	N/S	117	0.284	31	0.11
	HOXB9_1	2.45	Repressive	Neither	0.0	0.3	2.5	0.2	-0.3	2.2	168	-0.14	-0.03	0.11	0.0	N/S	33	-0.08	40	-0.01
	TBX4_2	2.45	Repressive	Neither	0.5	1.3	2.4	0.7	-0.8	1.7	43	-0.20	0.11	0.30	0.0	N/S	12	-0.072	7	0.73
	DLX3_1	2.44	Repressive	Neither	0.0	0.1	2.4	0.1	-0.1	2.3	70	-0.20	-0.07	0.13	0.0	N/S	16	-0.132	14	-0.07
	PAX8_2	2.43	Activating	Repressive	2.4	0.5	0.2	2.4	1.8	-2.2	57	0.15	0.00	0.16	0.0	N/S	16	0.12	9	0.22
	CEBPE_1	2.41	Activating	Neither	2.4	0.3	0.4	0.9	2.1	-0.5	34	0.33	-0.03	0.36	0.0	N/S	22	0.249	3	1.71
	TCF12_disc1	2.41	Neither	Repressive	0.4	0.2	1.0	2.4	0.3	-1.4	169	-0.01	-0.11	0.10	0.0	N/S	43	-0.048	39	0.09
	MYBL2_4	2.40	Neither	Repressive	0.2	0.7	0.6	2.4	-0.4	-1.8	18	-0.14	-0.16	0.02	0.0	N/S	4	-0.188	7	0.27
	ZIC3_4	2.40	Neither	Activating	0.4	2.4	0.2	0.4	-2.0	-0.2	234	0.08	0.09	0.01	0.0	N/S	47	0.116	55	0.21
	DPRX_1	2.40	Neither	Repressive	0.1	0.0	2.0	2.4	0.1	-0.4	58	-0.15	-0.24	0.09	0.0	N/S	16	-0.237	17	-0.26
	TFAP4_4	2.39	Repressive	Neither	0.8	1.2	2.4	0.2	-0.4	2.2	290	-0.04	0.04	0.07	0.0	N/S	55	-0.004	70	0.18
	GZF1_1	2.37	Repressive	Neither	1.4	0.5	2.4	0.9	0.9	1.4	7	0.17	-0.18	0.34	0.0	N/S	3	-0.459	0	#N/A
	NR6A1_1	2.36	Repressive	Neither	0.2	0.1	2.4	1.0	0.1	1.4	174	-0.08	-0.06	0.02	0.0	N/S	71	0.071	39	0.03
	TBX21_1	2.36	Neither	Repressive	0.3	0.6	0.4	2.4	-0.3	-2.0	37	-0.08	-0.16	0.08	0.0	N/S	11	0.04	6	-0.15
	SMAD_1	2.33	Neither	Repressive	0.1	0.0	0.1	2.3	0.1	-2.2	113	-0.05	-0.08	0.03	0.0	N/S	29	0.126	31	-0.1
	CEBPG_3	2.32	Activating	Neither	2.3	0.5	0.5	1.1	1.8	-0.6	25	0.36	0.07	0.29	0.0	N/S	16	0.172	3	1.71
	ZNF524_2	2.32	Activating	Neither	2.3	0.5	1.3	0.4	1.8	0.9	25	0.26	0.12	0.13	0.0	N/S	9	0.329	5	0.08
	MZF1_2	2.30	Neither	Activating	0.1	2.3	0.1	0.1	-2.2	0.0	349	0.03	0.06	0.03	0.0	N/S	76	0.12	80	0.16
	OSR2_1	2.30	Neither	Repressive	1.2	0.3	1.2	2.3	1.0	-1.1	61	-0.01	-0.12	0.11	0.0	N/S	18	0.402	11	-0.49
	ZBTB18_3	2.26	Repressive	Neither	0.1	0.0	2.3	0.7	0.1	1.6	220	-0.16	-0.10	0.06	0.0	N/S	52	-0.095	46	-0.01
	AHR::ARNT::HIF1A_1	2.26	Neither	Activating	1.1	2.3	0.4	1.6	-1.1	-1.2	35	0.18	0.19	0.01	0.0	N/S	10	-0.05	8	0.43
	AHR::ARNT_2	2.26	Neither	Activating	0.7	2.3	1.0	0.3	-1.6	0.7	35	0.20	0.40	0.21	0.0	N/S	8	0.327	4	0.38
	SPIC_2	2.25	Activating	Activating	2.2	2.3	0.0	0.1	0.0	-0.1	336	0.08	0.17	0.10	0.0	N/S	58	0.262	124	0.21
	CRX_2	2.24	Repressive	Neither	0.2	0.2	2.2	1.2	0.0	1.0	49	-0.16	-0.20	0.04	0.0	N/S	12	-0.047	12	-0.08
	ZIC4_1	2.24	Neither	Activating	0.4	2.2	0.3	0.5	-1.8	-0.3	196	0.06	0.08	0.03	0.0	N/S	39	0.073	42	0.19
	RAX_1	2.23	Neither	Repressive	0.1	0.1	1.7	2.2	0.0	-0.5	40	-0.25	-0.13	0.12	0.0	N/S	11	-0.291	8	-0.01
	NKX2-1_1	2.22	Neither	Repressive	0.4	0.1	0.8	2.2	0.3	-1.4	90	-0.07	-0.13	0.06	0.0	N/S	24	0.158	18	-0.21
	NFIL3_3	2.18	Activating	Neither	2.2	0.6	0.1	1.1	1.6	-1.0	88	0.21	0.01	0.20	0.0	N/S	32	0.351	17	0.08
	IRF4_1	2.17	Activating	Neither	2.2	0.9	1.2	1.1	1.3	0.2	27	0.12	0.16	0.04	0.0	N/S	5	0.131	6	0.77
	E2F7_1	2.17	Repressive	Neither	0.1	1.8	2.2	0.1	-1.8	2.1	81	-0.18	0.13	0.32	0.5	N/S	20	-0.148	17	-0
	MNX1_1	2.15	Repressive	Neither	0.4	0.3	2.2	1.2	0.0	1.0	52	-0.19	-0.14	0.05	0.0	N/S	10	-0.015	8	-0.03
	NKX3-1_1	2.15	Repressive	Neither	0.0	0.0	2.1	0.5	0.0	1.7	115	-0.19	-0.14	0.04	0.0	N/S	34	-0.106	21	-0.25
	OSR1_1	2.13	Neither	Repressive	0.7	0.1	1.1	2.1	0.6	-1.0	67	-0.04	-0.16	0.12	0.0	N/S	18	0.357	14	-0.46
	HESX1_1	2.12	Neither	Repressive	0.1	0.2	0.3	2.1	-0.2	-1.8	43	-0.17	-0.15	0.01	0.0	N/S	13	-0.12	4	0.26
	CPHX_1	2.11	Neither	Activating	1.2	2.1	0.9	0.2	-0.9	0.8	128	0.25	0.34	0.10	0.0	N/S	27	0.3	33	0.54
	NKX1-1_1	2.08	Neither	Activating	1.0	2.1	0.4	0.3	-1.0	0.1	38	0.07	0.19	0.12	0.0	N/S	6	0.523	8	0.04
	GSC_1	2.07	Repressive	Neither	0.3	0.0	2.1	1.8	0.3	0.3	55	-0.10	-0.19	0.09	0.0	N/S	14	0.102	10	-0.25
	HLF_2	2.05	Activating	Neither	2.1	0.5	0.4	0.3	1.5	0.1	40	0.20	0.06	0.14	0.0	N/S	17	0.222	7	0.71
	VAX2_1	2.05	Neither	Repressive	0.2	0.4	0.9	2.1	-0.2	-1.2	45	-0.22	-0.16	0.06	0.0	N/S	15	-0.156	5	-0.04
	E2F2_1	2.05	Neither	Activating	0.7	2.0	0.6	0.6	-1.3	0.1	18	0.26	0.44	0.18	0.0	N/S	2	0.004	6	0.68
	FOXO3_3	2.04	Activating	Neither	2.0	0.1	0.4	0.1	1.9	0.3	279	0.11	-0.03	0.14	0.0	N/S	111	0.296	47	0.08
	FOXK1_1	2.04	Activating	Neither	2.0	0.0	0.0	0.0	2.0	0.0	381	0.07	-0.02	0.09	0.0	N/S	140	0.302	60	0.03
	NR2E1_1	2.04	Neither	Repressive	0.0	0.0	0.0	2.0	0.0	-2.0	160	-0.12	-0.10	0.02	0.0	N/S	54	-0.061	45	0
	FOXO6_3	2.02	Neither	Repressive	0.4	1.4	0.7	2.0	-1.0	-1.3	161	0.03	0.05	0.02	0.0	N/S	33	0.117	39	0.27
	THRA_1	2.02	Activating	Neither	2.0	0.3	0.2	0.8	1.7	-0.6	79	0.10	-0.07	0.17	0.0	N/S	25	0.278	14	-0.18
	EMX2_2	2.01	Repressive	Neither	0.6	0.9	2.0	1.1	-0.3	0.9	57	-0.05	-0.03	0.02	0.0	N/S	12	0.22	11	0.12
	HMX3_3	2.00	Neither	Neither	0.0	0.4	2.0	0.1	-0.4	1.9	89	-0.19	-0.01	0.18	0.4	N/S	21	-0.075	22	0.06



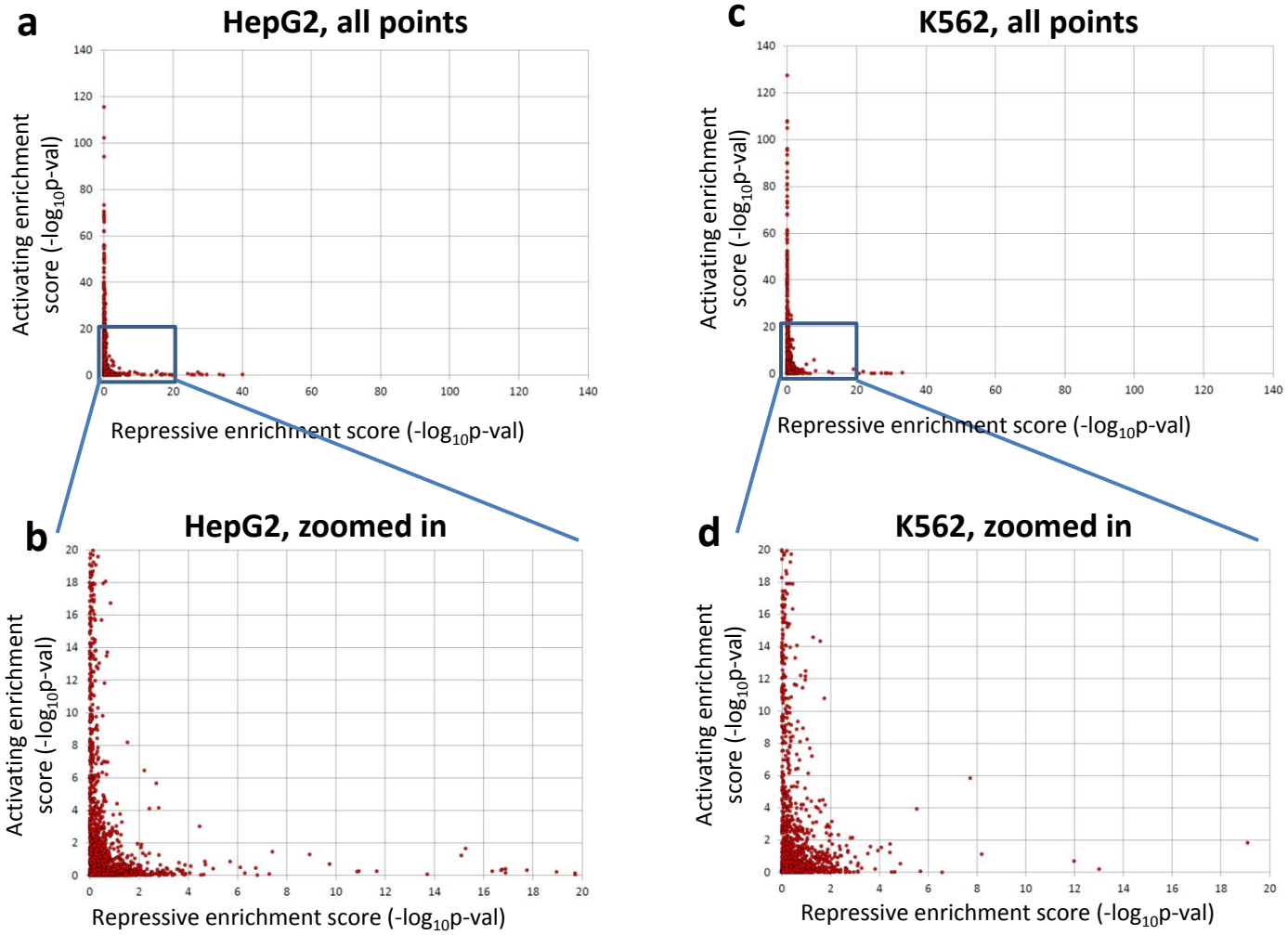
Supplementary Figure 26 – Scatterplot of regulatory motif enrichments (a) comparing the –log10 P-value of the enrichment for regulatory motif instances with activity ≥1 at the center position in HepG2 (y-axis) and in K562 (x-axis) for all regulatory motifs shown in **Supplementary Fig. 25** and in **Supplementary Table 2**. **(b)** Same as (a) except considering repression based on enrichment for regulatory motif instances with activity ≤-1.



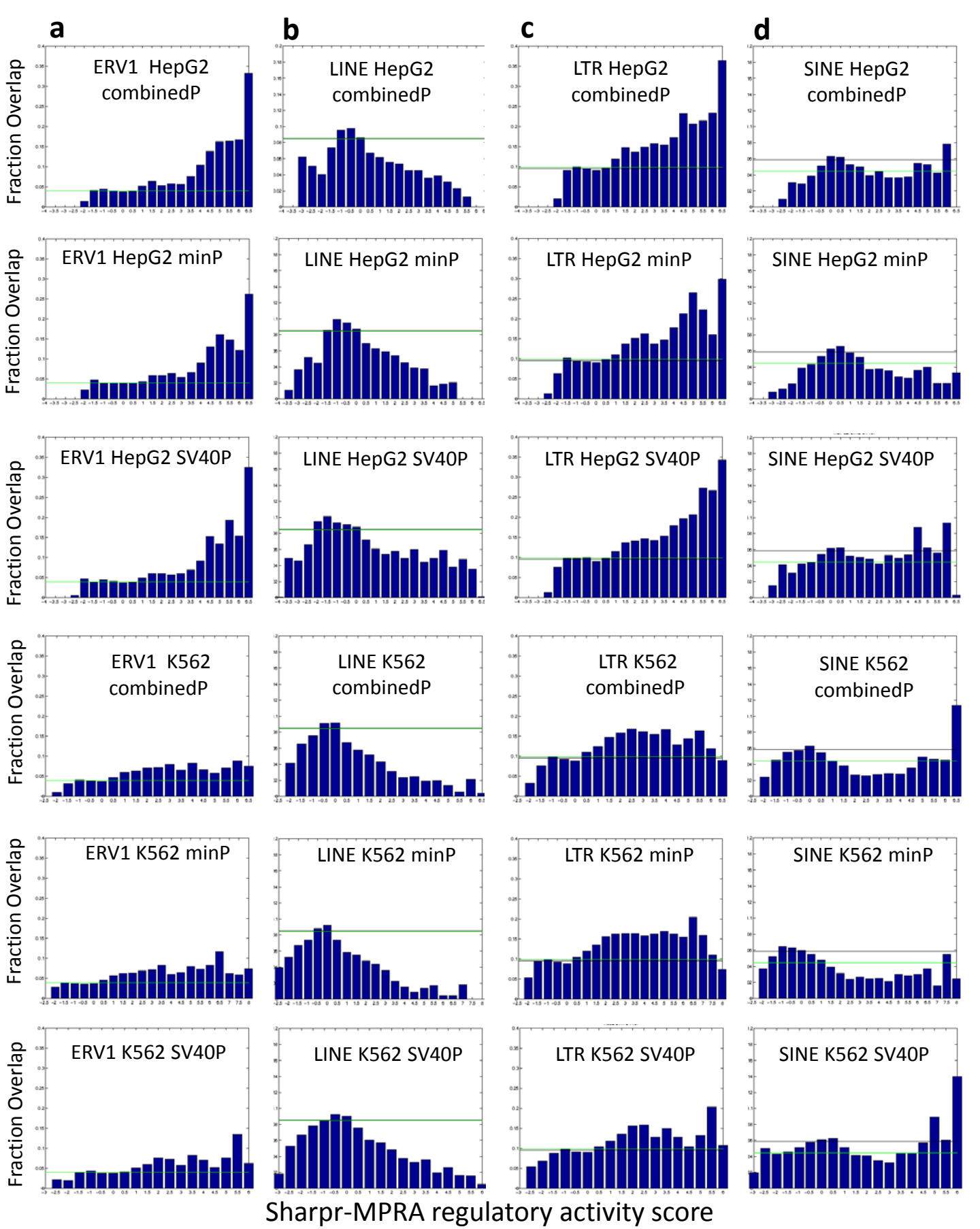
Supplementary Figure 27 – Sharpr-MPRA Regulatory Score Aggregation Plots Relative to Motif Position. Aggregation plots of Sharpr-MPRA regulatory scores relative to motif position from **Fig. 4b** shown separately for instances whose motif center fell within the central 51 bp (red), or to the left (blue) or right (green) of that in HepG2 (left) and K562 (right) cells, shown for: **(a)** a ETS motif, ETS_known9; **(b)** a HNF4 motif, HNF4_known18; **(c)** a GATA motif, GATA_known14; **(d)** a REST motif, REST_known2; **(e)** a RFX motif, RFX5_known6, as specified in Ref. 13, which predicted motif instances independent of cell type. Vertical error bars indicate standard error.



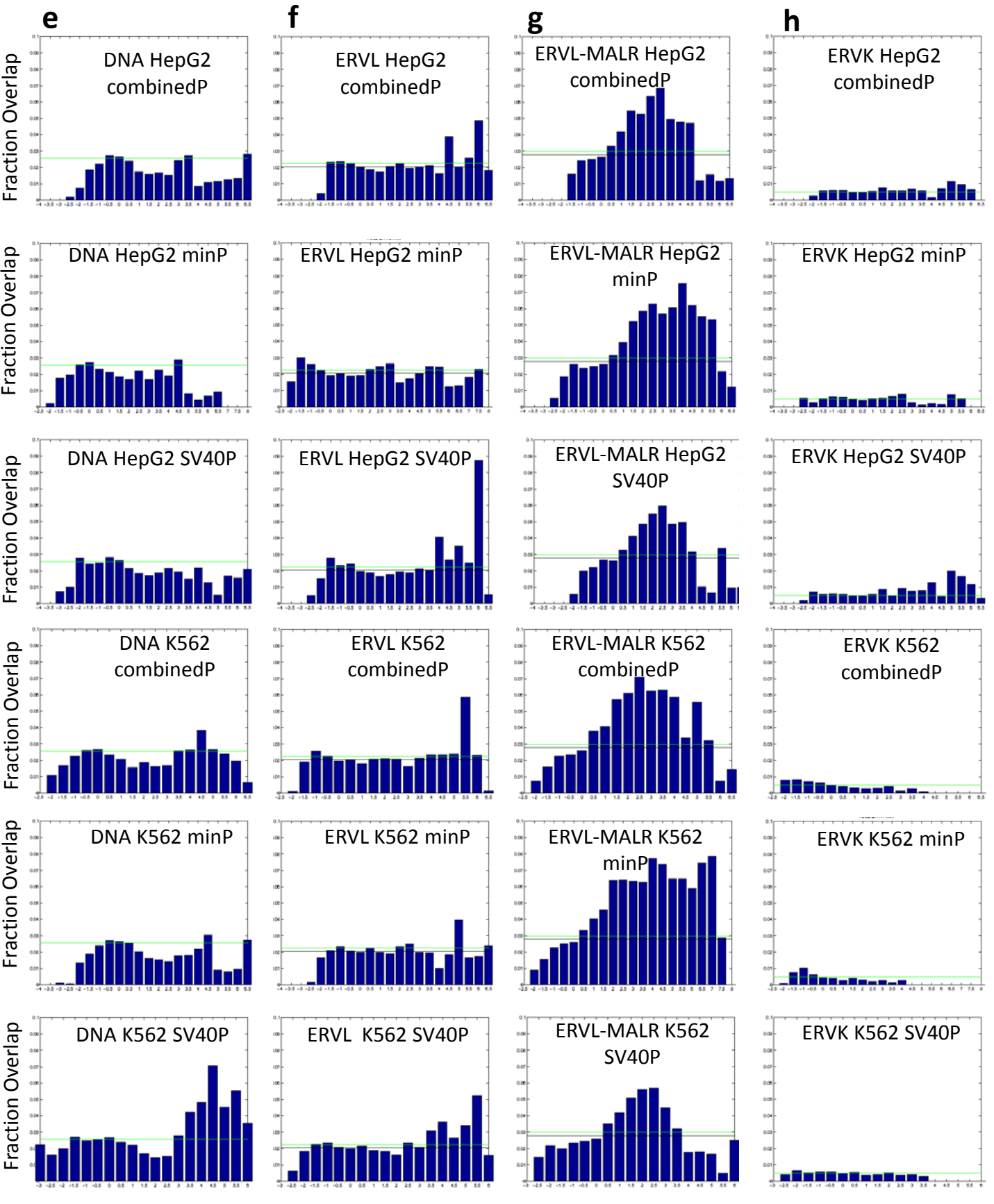
Supplementary Figure 28 – 7-mer Sharpr-MPRA regulatory activity score plot. The plot has a point for each 7-mer appearing more than ten times based on the forward strand showing the average regulatory activity score in HepG2 cells (x-axis) and the average regulatory activity score in K562 cells (y-axis).



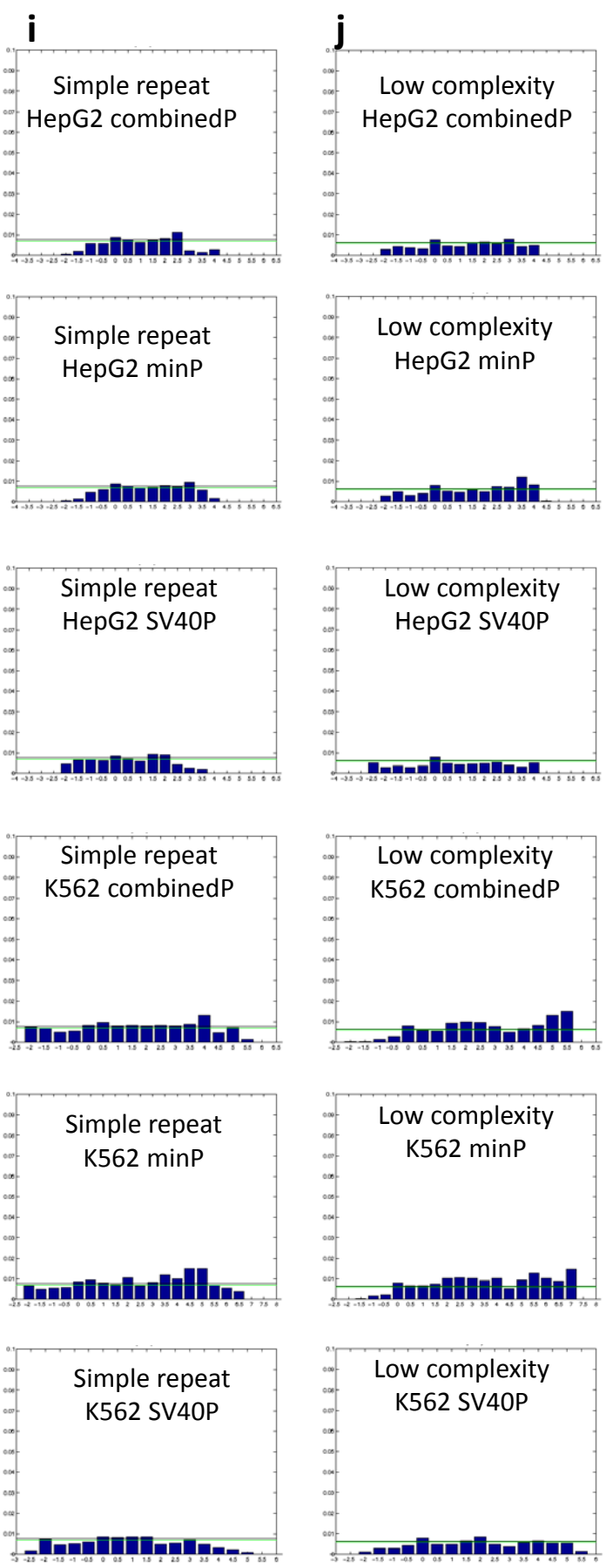
Supplementary Figure 29 – Enrichment for regulatory motif instances in top 5% activated and top 5% repressed nucleotides. These are similar to the plots shown in **Fig. 4c** except here the activating and repressive bases are defined as the 5% nucleotides that were given the most activating and repressive scores respectively instead of using the 1 and -1 thresholds. Activator score (y-axis) and repressor score (x-axis) for a compendium of regulatory motifs[13] (points) in HepG2 (left) and K562 (right), based on the statistical significance ($-\log_{10}P$) for the enrichment of the center position of motif instances in positions that had a regulatory score in the lowest 5% (most repressive) or highest 5% (most activating), computed based on one-side binomial tests where the probability of success for the binomial distribution is the fraction of total nucleotides tested that met the threshold. Motif instances that appeared on both strands at the same position are only counted once. **(a,b)** The scatter plots are for **(a)** all points in HepG2 and **(b)** a zoomed in view of the points in HepG2 that have values less than 20 on both axes**. (c,d)** are similar except for K562.



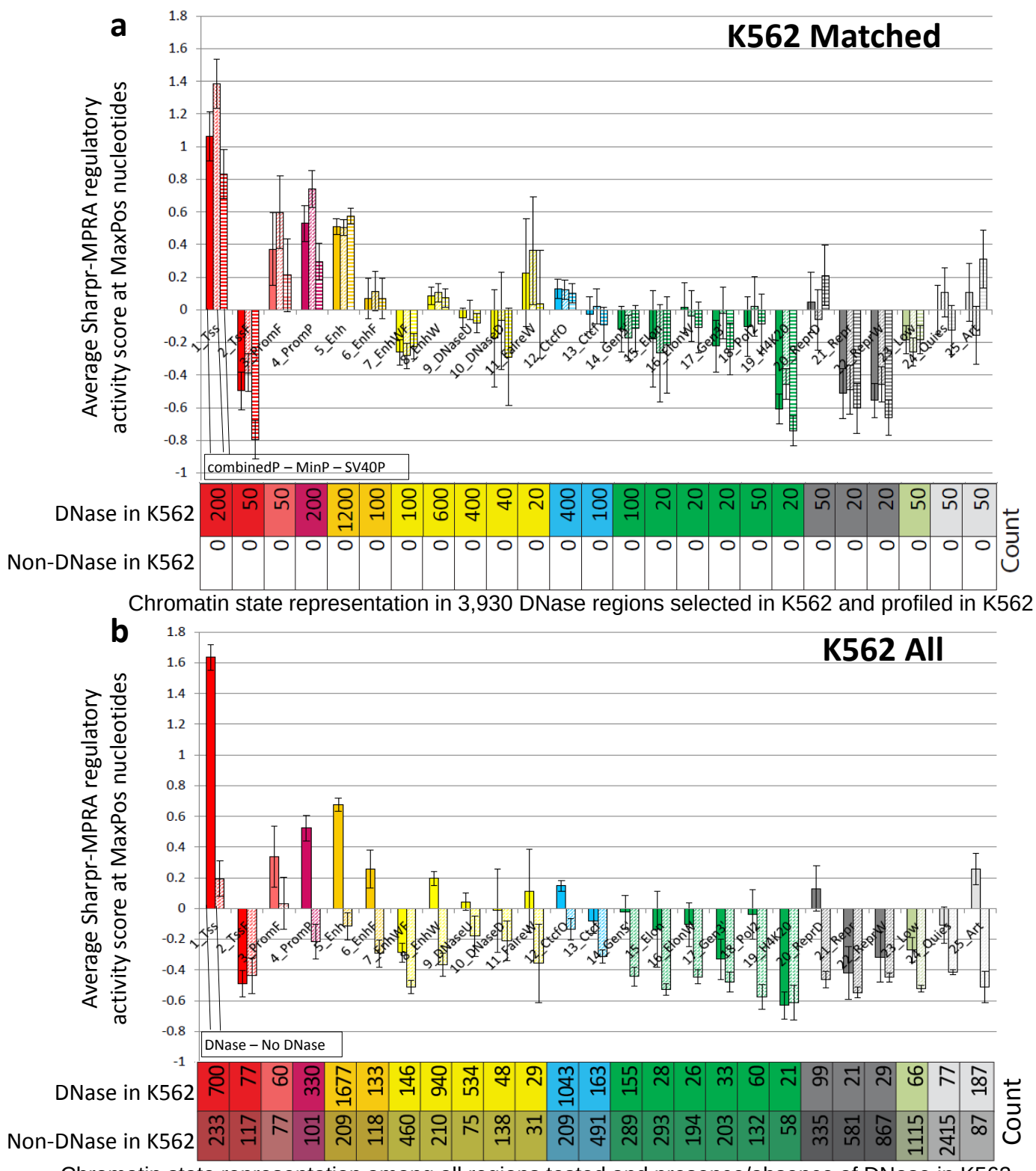
Supplementary Figure 30a-d – Regulatory Activity in Repeat Classes and Families (a-d) This page. **(e-h)** Next page. **(i,j)** Following page. For legend, see i,j.



Supplementary Figure 30e-h – Regulatory Activity in Repeat Classes and Families (a-d) Previous page. **(e-h)** This page. **(i,j)** Next page. For legend, see i,j.



Supplementary Figure 30 – Regulatory Activity in Repeat Classes and Families. Extended set of enrichments shown in **Fig. 5** now showing in each column for **(a)** ERV1, **(b)** LINE, **(c)** LTR, **(d)** SINE, **(e)** DNA, **(f)** ERVL, **(g)** ERVL-MaLR, **(h)** ERVK, **(i)** Simple Repeat, and **(j)** Low Complexity repeats as defined by RepeatMasker[70] for regulatory activity scores from top to bottom: HepG2 combinedP, HepG2 minP, HepG2 SV40P, K562 combinedP, K562 minP, and K562 SV40P data. Bins were formed by assigning each base to the nearest 0.5 value based on its regulatory score, and the two extreme bins contain all more extreme values. Green line denotes the expected fraction of overlap based on the center position (CenPos), and the black line the expected fraction based on all positions. Unlike in the corresponding main figure all bars are shown in the same blue color regardless if they are above or below expectation.



Supplementary Figure 31 – Activity by Chromatin State – K562 cells. Analogous figures as **Fig. 6 a** and **b** except for K562 cells. **(a)** For each chromatin state (x-axis) the average K562 regulatory score of all MaxPos nucleotides in K562, over all regions selected based on that state in that cell type. Results are shown for the combinedP, minP, and SV40P results (consecutive bars). The number of regions selected based on each state in the cell type is shown on the bottom. Since the regions were selected based on the indicated cell type they all have DNase sites. **(b)** For each chromatin state, the average score at MaxPos nucleotides for regions with the center position in the state based on all locations tested shown separately for locations overlapping DNase sites and those not overlapping DNase sites in K562. The number of regions in each set is indicated along the bottom. The vertical error bar indicates one standard error.

a. DNase 'matched' (selected in same cell type as tested)

DNase Matched (selected in same cell type)

HepG2 state			HepG2 Combined			HepG2 minP			HepG2 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	200	36%	9%	55%	38%	15%	48%	37%	11%	52%
2	TssF	50	8%	10%	82%	12%	20%	68%	12%	18%	70%
3	PromF	50	16%	16%	68%	18%	16%	66%	14%	24%	62%
4	PromP	200	23%	15%	62%	25%	20%	56%	21%	18%	61%
5	Enh	1200	29%	7%	64%	30%	14%	56%	32%	14%	54%
6	EnhF	100	15%	10%	75%	18%	11%	71%	16%	18%	66%
7	EnhWF	100	9%	11%	80%	12%	23%	65%	10%	21%	69%
8	EnhW	600	22%	11%	67%	21%	14%	65%	27%	19%	55%
9	DNaseU	400	18%	13%	69%	19%	17%	64%	22%	17%	62%
10	DNaseD	40	18%	15%	68%	20%	25%	55%	23%	23%	55%
11	FaireW	20	10%	0%	90%	10%	5%	85%	15%	5%	80%
12	CtcfO	400	15%	17%	69%	17%	21%	62%	19%	22%	59%
13	Ctcf	100	2%	11%	87%	5%	17%	78%	4%	16%	80%
14	Gen5'	100	24%	13%	63%	22%	22%	56%	21%	20%	59%
15	Elon	20	15%	10%	75%	15%	15%	70%	15%	10%	75%
16	ElonW	20	15%	10%	75%	20%	20%	60%	15%	20%	65%
17	Gen3'	20	5%	25%	70%	5%	25%	70%	5%	40%	55%
18	Pol2	50	20%	14%	66%	22%	24%	54%	18%	16%	66%
19	H4K20	20	0%	20%	80%	5%	25%	70%	5%	20%	75%
20	ReprD	50	12%	26%	62%	18%	34%	48%	16%	28%	56%
21	Repr	20	0%	35%	65%	5%	35%	60%	0%	25%	75%
22	ReprW	20	20%	25%	55%	15%	35%	50%	20%	35%	45%
23	Low	50	16%	14%	70%	16%	20%	64%	14%	26%	60%
24	Quies	50	6%	16%	78%	8%	20%	72%	6%	24%	70%
25	Art	50	12%	16%	72%	14%	28%	58%	22%	22%	56%
All states		3930	22%	11%	67%	23%	17%	60%	24%	17%	58%

DNase Matched (selected in same cell type)

K562 state			K562 Combined			K562 minP			K562 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	200	41%	13%	46%	44%	16%	41%	44%	19%	37%
2	TssF	50	6%	20%	74%	10%	26%	64%	6%	42%	52%
3	PromF	50	24%	8%	68%	30%	12%	58%	30%	20%	50%
4	PromP	200	34%	18%	49%	42%	17%	42%	36%	27%	38%
5	Enh	1200	32%	13%	55%	31%	16%	53%	40%	21%	39%
6	EnhF	100	14%	8%	78%	13%	9%	78%	21%	21%	58%
7	EnhWF	100	8%	13%	79%	9%	18%	73%	11%	28%	61%
8	EnhW	600	22%	18%	61%	21%	20%	59%	30%	30%	41%
9	DNaseU	400	11%	11%	78%	13%	11%	77%	18%	24%	58%
10	DNaseD	40	18%	40%	43%	18%	38%	45%	25%	50%	25%
11	FaireW	20	15%	0%	85%	15%	5%	80%	10%	5%	85%
12	CtcfO	400	20%	12%	68%	20%	12%	68%	31%	25%	44%
13	Ctcf	100	12%	13%	75%	9%	12%	79%	18%	22%	60%
14	Gen5'	100	13%	25%	62%	10%	21%	69%	20%	34%	46%
15	Elon	20	10%	20%	70%	5%	20%	75%	10%	30%	60%
16	ElonW	20	10%	5%	85%	10%	5%	85%	15%	20%	65%
17	Gen3'	20	5%	5%	90%	5%	5%	90%	15%	25%	60%
18	Pol2	50	14%	12%	74%	12%	12%	76%	30%	24%	46%
19	H4K20	20	0%	15%	85%	0%	5%	95%	0%	40%	60%
20	ReprD	50	24%	20%	56%	24%	24%	52%	34%	26%	40%
21	Repr	20	0%	25%	75%	5%	20%	75%	5%	40%	55%
22	ReprW	20	0%	10%	90%	5%	10%	85%	10%	40%	50%
23	Low	50	2%	10%	88%	2%	6%	92%	18%	24%	58%
24	Quies	50	14%	10%	76%	14%	12%	74%	16%	18%	66%
25	Art	50	18%	14%	68%	16%	22%	62%	36%	16%	48%
All states		3930	23%	14%	63%	23%	16%	61%	30%	25%	45%

b. DNase 'unmatched' (selected in different cell type from tested)

DNase Unmatched (selected in another cell type)

HepG2 state			HepG2 Combined			HepG2 minP			HepG2 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	599	53%	5%	42%	55%	9%	36%	53%	8%	39%
2	TssF	23	13%	22%	65%	13%	26%	61%	17%	35%	48%
3	PromF	4	25%	0%	75%	25%	25%	50%	25%	0%	75%
4	PromP	364	26%	10%	64%	31%	15%	54%	28%	14%	58%
5	Enh	191	43%	10%	47%	46%	17%	37%	47%	16%	37%
6	EnhF	13	23%	8%	69%	23%	15%	62%	15%	15%	69%
7	EnhWF	13	0%	15%	85%	8%	15%	77%	0%	23%	77%
8	EnhW	154	34%	12%	53%	37%	18%	45%	37%	16%	47%
9	DNaseU	45	38%	11%	51%	38%	27%	36%	33%	13%	53%
10	DNaseD	30	33%	20%	47%	37%	20%	43%	40%	23%	37%
11	FaireW	3	0%	33%	67%	33%	67%	0%	0%	67%	33%
12	CtcfO	319	14%	15%	70%	18%	24%	58%	18%	23%	60%
13	Ctcf	9	11%	0%	89%	11%	11%	78%	11%	11%	78%
14	Gen5'	27	26%	15%	59%	30%	15%	56%	30%	22%	48%
15	Elon	4	25%	25%	50%	25%	0%	75%	50%	25%	25%
16	ElonW	1	100%	0%	0%	100%	0%	0%	100%	0%	0%
17	Gen3'	11	45%	0%	55%	36%	0%	64%	45%	18%	36%
18	Pol2	8	50%	0%	50%	50%	13%	38%	75%	0%	25%
19	H4K20	0									
20	ReprD	25	28%	8%	64%	28%	16%	56%	28%	24%	48%
21	Repr	0									
22	ReprW	1	0%	0%	100%	0%	0%	100%	0%	0%	100%
23	Low	4	50%	0%	50%	25%	25%	50%	25%	0%	75%
24	Quies	1	0%	0%	100%	0%	0%	100%	0%	0%	100%
25	Art	61	20%	18%	62%	23%	20%	57%	21%	25%	54%
All states		1910	35%	10%	55%	38%	16%	46%	37%	15%	48%

DNase Unmatched (selected in another cell type)

K562 state			K562 Combined			K562 minP			K562 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	500	60%	9%	31%	64%	8%	28%	59%	15%	26%
2	TssF	27	0%	19%	81%	4%	11%	85%	11%	41%	48%
3	PromF	10	30%	0%	70%	20%	0%	80%	20%	30%	50%
4	PromP	130	38%	18%	45%	38%	17%	45%	41%	22%	38%
5	Enh	477	47%	11%	42%	47%	13%	40%	55%	18%	27%
6	EnhF	33	36%	6%	58%	33%	9%	58%	45%	18%	36%
7	EnhWF	46	2%	13%	85%	4%	17%	78%	11%	26%	63%
8	EnhW	340	32%	18%	50%	33%	20%	47%	41%	29%	30%
9	DNaseU	134	27%	12%	61%	26%	13%	61%	28%	22%	50%
10	DNaseD	8	38%	13%	50%	38%	13%	50%	63%	13%	25%
11	FaireW	9	22%	33%	44%	22%	33%	44%	11%	67%	22%
12	CtcfO	643	21%	13%	66%	21%	14%	65%	33%	23%	43%
13	Ctcf	63	6%	6%	87%	6%	5%	89%	10%	22%	68%
14	Gen5'	55	27%	13%	60%	25%	15%	60%	27%	35%	38%
15	Elon	8	38%	25%	38%	25%	25%	50%	25%	38%	38%
16	ElonW	6	0%	17%	83%	0%	17%	83%	17%	33%	50%
17	Gen3'	13	8%	23%	69%	8%	23%	69%	15%	31%	54%
18	Pol2	10	30%	10%	60%	20%	30%	50%	50%	10%	40%
19	H4K20	1	0%	100%	0%	0%	100%	0%	0%	100%	0%
20	ReprD	49	37%	27%	37%	33%	20%	47%	39%	41%	20%
21	Repr	1	100%	0%	0%	100%	0%	0%	100%	0%	0%
22	ReprW	9	44%	22%	33%	44%	33%	22%	44%	22%	33%
23	Low	16	0%	25%	75%	0%	31%	69%	19%	31%	50%
24	Quies	27	11%	19%	70%	11%	19%	70%	19%	37%	44%
25	Art	137	29%	18%	53%	26%	23%	51%	42%	24%	34%
All states		2752	35%	13%	52%	36%	14%	50%	42%	23%	35%

Supplementary Figure 32a,b – Fraction of regions showing above activating or below repressive threshold at the maximum absolute score position (MaxPos). (a) For each chromatin state (rows) in HepG2 (left panel) and K562 (right panel), the number of regions tested ('All') that were *selected based on the chromatin state annotation in the same cell type (DNase Matched)*, the percentage of those regions with score at MaxPos ≥1 ('Act' for 'Activating'), ≤-1 ('Rep' for 'Repressive'), or neither ('Non') using the combinedP score (first column group), minP score (second group), or SV40P score (third group). Each individual column is colored based on median (white), the 90th percentile (red), and the 10th percentile (blue), thus indicating chromatin states that are more often activating (e.g. Tss, Enh) or repressive (e.g. ReprD, Repr) than expected on average. Black boxes highlight the numbers discussed in the main text. **(b)** Similar to (a), but for DNase sites that were selected in a different cell type and are also DNase in the tested cell type. **(c-e)** Next pages.

c. DNase All ('matched' and 'unmatched')

DNase All

HepG2 state			HepG2 Combined			HepG2 minP			HepG2 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	799	49%	6%	45%	51%	11%	39%	49%	9%	42%
2	TssF	73	10%	14%	77%	12%	22%	66%	14%	23%	63%
3	PromF	54	17%	15%	69%	19%	17%	65%	15%	22%	63%
4	PromP	564	25%	12%	63%	29%	17%	54%	26%	15%	59%
5	Enh	1391	31%	8%	61%	32%	14%	53%	34%	14%	52%
6	EnhF	113	16%	10%	74%	19%	12%	70%	16%	18%	66%
7	EnhWF	113	8%	12%	81%	12%	22%	66%	9%	21%	70%
8	EnhW	754	25%	11%	64%	25%	14%	61%	29%	18%	53%
9	DNaseU	445	20%	12%	67%	21%	18%	61%	23%	16%	61%
10	DNaseD	70	24%	17%	59%	27%	23%	50%	30%	23%	47%
11	FaireW	23	9%	4%	87%	13%	13%	74%	13%	13%	74%
12	CtcfO	719	14%	16%	70%	18%	22%	61%	18%	22%	59%
13	Ctcf	109	3%	10%	87%	6%	17%	78%	5%	16%	80%
14	Gen5'	127	24%	13%	62%	24%	20%	56%	23%	20%	57%
15	Elon	24	17%	13%	71%	17%	13%	71%	21%	13%	67%
16	ElonW	21	19%	10%	71%	24%	19%	57%	19%	19%	62%
17	Gen3'	31	19%	16%	65%	16%	16%	68%	19%	32%	48%
18	Pol2	58	24%	12%	64%	26%	22%	52%	26%	14%	60%
19	H4K20	20	0%	20%	80%	5%	25%	70%	5%	20%	75%
20	ReprD	75	17%	20%	63%	21%	28%	51%	20%	27%	53%
21	Repr	20	0%	35%	65%	5%	35%	60%	0%	25%	75%
22	ReprW	21	19%	24%	57%	14%	33%	52%	19%	33%	48%
23	Low	54	19%	13%	69%	17%	20%	63%	15%	24%	61%
24	Quies	51	6%	16%	78%	8%	20%	73%	6%	24%	71%
25	Art	111	16%	17%	67%	19%	23%	58%	22%	23%	55%
All states		5840	26%	11%	63%	28%	16%	56%	28%	17%	55%

DNase All

K562 state			K562 Combined			K562 minP			K562 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	700	55%	10%	35%	58%	10%	31%	55%	16%	29%
2	TssF	77	4%	19%	77%	8%	21%	71%	8%	42%	51%
3	PromF	60	25%	7%	68%	28%	10%	62%	28%	22%	50%
4	PromP	330	35%	18%	47%	40%	17%	43%	38%	25%	38%
5	Enh	1677	36%	12%	51%	36%	15%	49%	44%	20%	35%
6	EnhF	133	20%	8%	73%	18%	9%	73%	27%	20%	53%
7	EnhWF	146	6%	13%	81%	8%	18%	75%	11%	27%	62%
8	EnhW	940	25%	18%	57%	25%	20%	55%	34%	30%	37%
9	DNaseU	534	15%	11%	74%	16%	11%	73%	21%	24%	56%
10	DNaseD	48	21%	35%	44%	21%	33%	46%	31%	44%	25%
11	FaireW	29	17%	10%	72%	17%	14%	69%	10%	24%	66%
12	CtcfO	1043	21%	13%	67%	20%	13%	66%	33%	24%	43%
13	Ctcf	163	10%	10%	80%	8%	9%	83%	15%	22%	63%
14	Gen5'	155	18%	21%	61%	15%	19%	66%	23%	34%	43%
15	Elon	28	18%	21%	61%	11%	21%	68%	14%	32%	54%
16	ElonW	26	8%	8%	85%	8%	8%	85%	15%	23%	62%
17	Gen3'	33	6%	12%	82%	6%	12%	82%	15%	27%	58%
18	Pol2	60	17%	12%	72%	13%	15%	72%	33%	22%	45%
19	H4K20	21	0%	19%	81%	0%	10%	90%	0%	43%	57%
20	ReprD	99	30%	23%	46%	28%	22%	49%	36%	33%	30%
21	Repr	21	5%	24%	71%	10%	19%	71%	10%	38%	52%
22	ReprW	29	14%	14%	72%	17%	17%	66%	21%	34%	45%
23	Low	66	2%	14%	85%	2%	12%	86%	18%	26%	56%
24	Quies	77	13%	13%	74%	13%	14%	73%	17%	25%	58%
25	Art	187	26%	17%	57%	24%	22%	54%	41%	22%	37%
All states		6682	28%	14%	58%	28%	15%	57%	35%	24%	41%

d. Non-DNase in tested cell type (selected as DNase in diff. cell type)

Non-DNase (unmatched) (DNase in a different cell type)

HepG2 state			HepG2 Combined			HepG2 minP			HepG2 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	352	18%	17%	65%	20%	26%	54%	20%	24%	57%
2	TssF	205	5%	17%	78%	5%	31%	63%	8%	27%	65%
3	PromF	233	9%	14%	77%	10%	29%	61%	7%	21%	72%
4	PromP	393	17%	18%	65%	20%	25%	55%	19%	20%	62%
5	Enh	145	17%	12%	70%	20%	22%	58%	23%	18%	59%
6	EnhF	142	14%	16%	70%	17%	27%	56%	13%	20%	66%
7	EnhWF	272	8%	21%	71%	11%	29%	60%	9%	22%	69%
8	EnhW	202	18%	13%	69%	20%	25%	55%	19%	19%	62%
9	DNaseU	60	22%	18%	60%	20%	22%	58%	20%	25%	55%
10	DNaseD	332	19%	23%	58%	22%	26%	52%	22%	30%	48%
11	FaireW	20	10%	15%	75%	10%	40%	50%	10%	30%	60%
12	CtcfO	532	14%	15%	72%	17%	23%	60%	13%	23%	64%
13	Ctcf	528	8%	16%	76%	12%	23%	64%	9%	21%	70%
14	Gen5'	253	9%	17%	73%	9%	26%	64%	9%	22%	69%
15	Elon	588	6%	21%	73%	8%	30%	61%	6%	26%	68%
16	ElonW	261	4%	15%	81%	6%	25%	69%	5%	24%	71%
17	Gen3'	184	7%	15%	78%	10%	22%	68%	7%	21%	72%
18	Pol2	124	15%	14%	72%	16%	23%	60%	14%	25%	61%
19	H4K20	147	5%	22%	73%	5%	25%	69%	7%	29%	64%
20	ReprD	329	13%	19%	67%	16%	25%	59%	13%	27%	60%
21	Repr	469	7%	20%	74%	9%	30%	62%	7%	27%	66%
22	ReprW	693	8%	19%	73%	11%	29%	60%	8%	26%	67%
23	Low	1310	5%	19%	76%	8%	32%	60%	5%	26%	68%
24	Quies	2043	6%	19%	75%	8%	29%	63%	7%	25%	69%
25	Art	63	6%	19%	75%	8%	27%	65%	8%	25%	67%
All states		9880	9%	18%	73%	11%	28%	61%	9%	25%	66%

Non-DNase (unmatched) (DNase in a different cell type)

K562 state			K562 Combined			K562 minP			K562 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	233	24%	27%	49%	27%	27%	46%	25%	36%	39%
2	TssF	117	8%	27%	65%	8%	42%	50%	15%	36%	50%
3	PromF	77	17%	22%	61%	19%	22%	58%	25%	25%	51%
4	PromP	101	14%	23%	63%	13%	20%	67%	17%	38%	46%
5	Enh	209	17%	26%	57%	17%	26%	57%	27%	33%	40%
6	EnhF	118	11%	19%	70%	8%	35%	57%	20%	31%	48%
7	EnhWF	460	8%	33%	59%	6%	39%	56%	17%	43%	40%
8	EnhW	210	12%	32%	56%	11%	39%	50%	19%	42%	39%
9	DNaseU	75	13%	20%	67%	15%	23%	63%	20%	29%	51%
10	DNaseD	138	20%	29%	51%	19%	32%	49%	25%	39%	36%
11	FaireW	31	3%	16%	81%	3%	23%	74%	6%	26%	68%
12	CtcfO	209	12%	20%	67%	12%	24%	64%	27%	29%	44%
13	Ctcf	491	10%	21%	69%	10%	24%	66%	18%	32%	50%
14	Gen5'	289	8%	32%	60%	10%	40%	49%	20%	43%	37%
15	Elon	293	1%	21%	77%	4%	22%	74%	9%	33%	58%
16	ElonW	194	2%	14%	84%	2%	18%	80%	8%	27%	64%
17	Gen3'	203	9%	23%	68%	10%	31%	59%	16%	38%	46%
18	Pol2	132	6%	34%	60%	5%	32%	64%	19%	42%	39%
19	H4K20	58	3%	28%	69%	5%	36%	59%	10%	47%	43%
20	ReprD	335	10%	33%	57%	10%	42%	48%	21%	41%	38%
21	Repr	581	6%	30%	64%	6%	41%	53%	18%	38%	44%
22	ReprW	867	5%	23%	72%	6%	26%	68%	13%	36%	51%
23	Low	1115	3%	23%	74%	3%	26%	71%	8%	35%	57%
24	Quies	2415	5%	18%	77%	5%	20%	75%	10%	31%	58%
25	Art	87	9%	37%	54%	7%	43%	51%	22%	44%	34%
All states		9038	7%	24%	69%	7%	27%	65%	15%	35%	50%

Supplementary Figure 32c,d – Fraction of regions showing above activating or below repressive threshold at the maximum absolute score position (MaxPos). (a,b) Previous page. **(c)** Same as (a), but showing total counts and activity for all DNase in sites, regardless of the cell type in which they were selected. Total region counts ('All') are simply the sum of panels (a) and (b). Black boxes highlight the numbers discussed in the main text. **(d)** Same percentages but shown for the converse set regions, namely all the regions that do not show DNase activity in the tested cell type (and thus "Unmatched", as all regions selected were DNase in the cell type in which they were selected). Differences in the fraction of active regions between (a,b,c) vs. (d) reflect that among all regions in a given chromatin state, those that also overlap DNase regions are more likely to be activating (see also **Fig. 6b** and **Supplementary Fig 31b**). **(e)** Next Page.

e. All regions (DNase and non-DNase, matched and unmatched)

All 15720 region selections (all states, both DNase and non-DNase)

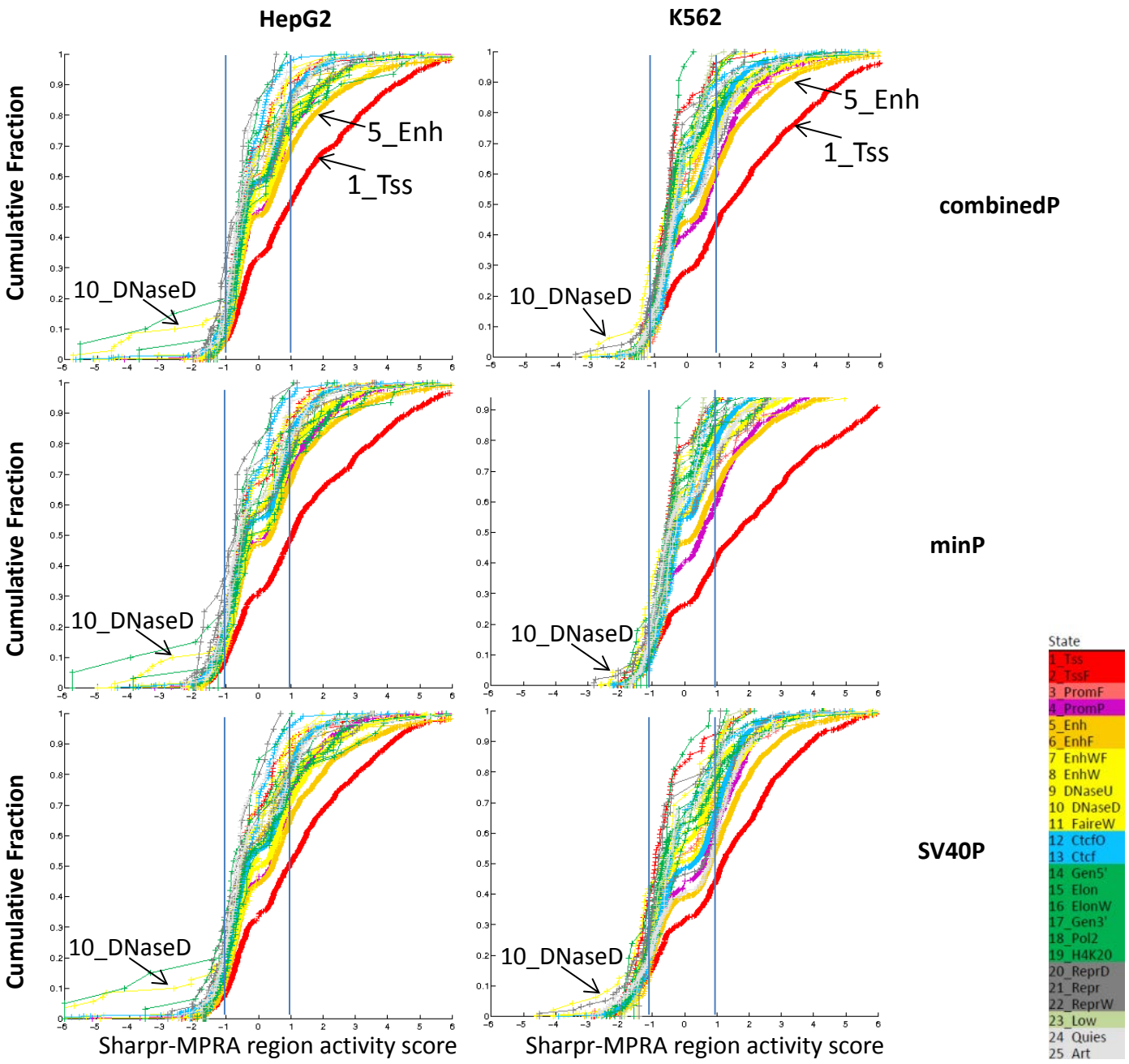
	HepG2 state		HepG2 Combined			HepG2 minP			HepG2 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	1151	39%	10%	51%	41%	15%	43%	40%	13%	47%
2	TssF	278	6%	16%	78%	7%	29%	64%	9%	26%	64%
3	PromF	287	10%	14%	76%	12%	26%	62%	9%	21%	70%
4	PromP	957	22%	14%	64%	25%	20%	55%	23%	17%	60%
5	Enh	1536	30%	8%	62%	31%	15%	54%	33%	14%	53%
6	EnhF	255	15%	13%	72%	18%	20%	62%	15%	19%	66%
7	EnhWF	385	8%	18%	74%	11%	27%	62%	9%	22%	70%
8	EnhW	956	23%	11%	65%	24%	17%	60%	27%	18%	55%
9	DNaseU	505	20%	13%	67%	21%	19%	60%	23%	17%	60%
10	DNaseD	402	20%	22%	58%	23%	26%	51%	23%	29%	48%
11	FaireW	43	9%	9%	81%	12%	26%	63%	12%	21%	67%
12	CtcfO	1251	14%	16%	70%	18%	22%	60%	16%	22%	62%
13	Ctcf	637	7%	15%	78%	11%	22%	67%	8%	20%	71%
14	Gen5'	380	14%	16%	69%	14%	24%	62%	14%	21%	65%
15	Elon	612	6%	20%	73%	9%	30%	62%	6%	25%	68%
16	ElonW	282	5%	15%	80%	7%	25%	68%	6%	23%	70%
17	Gen3'	215	9%	15%	76%	11%	21%	68%	9%	23%	68%
18	Pol2	182	18%	13%	69%	19%	23%	58%	18%	21%	61%
19	H4K20	167	5%	22%	74%	5%	25%	69%	7%	28%	65%
20	ReprD	404	14%	20%	66%	17%	25%	58%	14%	27%	59%
21	Repr	489	6%	20%	73%	8%	30%	62%	7%	27%	66%
22	ReprW	714	9%	19%	72%	11%	29%	59%	8%	26%	66%
23	Low	1364	6%	19%	75%	8%	32%	60%	5%	26%	68%
24	Quies	2094	6%	19%	75%	8%	29%	63%	7%	25%	69%
25	Art	174	13%	18%	70%	15%	25%	60%	17%	24%	59%
	All states	15720	15%	15%	69%	17%	24%	59%	16%	22%	62%

All 15720 region selections (all states, both DNase and non-DNase)

	K562 state		K562 Combined			K562 minP			K562 SV40P		
id	state	All	Act	Rep	Non	Act	Rep	Non	Act	Rep	Non
1	Tss	933	47%	14%	39%	50%	14%	35%	47%	21%	31%
2	TssF	194	6%	24%	70%	8%	34%	59%	12%	38%	50%
3	PromF	137	20%	15%	64%	23%	17%	60%	26%	23%	50%
4	PromP	431	30%	19%	51%	34%	18%	48%	33%	28%	39%
5	Enh	1886	34%	14%	52%	34%	16%	50%	42%	22%	36%
6	EnhF	251	16%	13%	72%	14%	21%	65%	24%	25%	51%
7	EnhWF	606	7%	28%	64%	6%	34%	60%	16%	39%	45%
8	EnhW	1150	23%	20%	57%	23%	23%	54%	31%	32%	37%
9	DNaseU	609	15%	12%	73%	16%	13%	71%	21%	24%	55%
10	DNaseD	186	20%	31%	49%	19%	32%	48%	27%	40%	33%
11	FaireW	60	10%	13%	77%	10%	18%	72%	8%	25%	67%
12	CtcfO	1252	19%	14%	67%	19%	15%	66%	32%	25%	44%
13	Ctcf	654	10%	18%	72%	9%	21%	70%	17%	30%	54%
14	Gen5'	444	11%	28%	60%	12%	33%	55%	21%	40%	39%
15	Elon	321	3%	21%	76%	4%	22%	74%	9%	33%	58%
16	ElonW	220	3%	13%	84%	3%	16%	81%	9%	27%	64%
17	Gen3'	236	8%	22%	70%	10%	28%	62%	16%	36%	47%
18	Pol2	192	9%	27%	64%	7%	27%	66%	23%	36%	41%
19	H4K20	79	3%	25%	72%	4%	29%	67%	8%	46%	47%
20	ReprD	434	15%	31%	55%	14%	38%	48%	25%	39%	36%
21	Repr	602	6%	30%	64%	6%	41%	53%	18%	38%	45%
22	ReprW	896	5%	23%	72%	6%	26%	68%	14%	36%	50%
23	Low	1181	3%	22%	75%	3%	25%	72%	9%	34%	57%
24	Quies	2492	5%	18%	77%	6%	19%	75%	11%	31%	58%
25	Art	274	21%	23%	56%	18%	29%	53%	35%	29%	36%
	All states	15720	16%	19%	65%	16%	22%	62%	23%	30%	46%

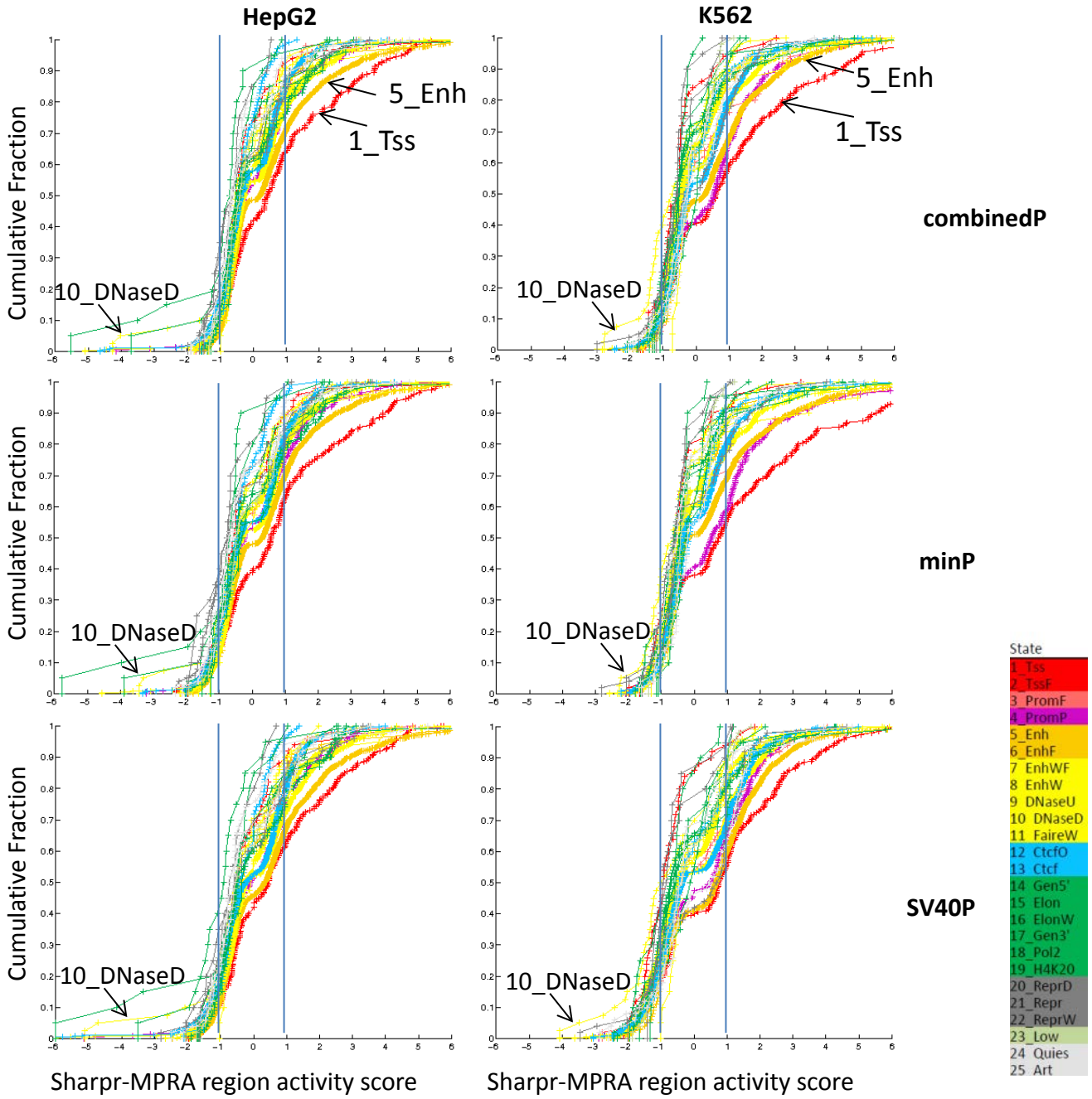
Supplementary Figure 32e – Fraction of regions showing above activating or below repressive threshold at the maximum absolute score position (MaxPos). (a-d) Previous pages. **(e)** Same as (a-d), but showing all tested regions, regardless of what cell type they were selected in, and regardless of whether they are found in a DNase region or a non-DNase region. This shows that active chromatin states (e.g. Tss, Enh) are more likely to show activating regulatory scores (≥1), that repressive chromatin states (e.g. Repr, ReprD, ReprW) are more likely to show repressive regulatory scores (≤-1).

a. DNase All (N=5840 in HepG2, N=6682 in K562)



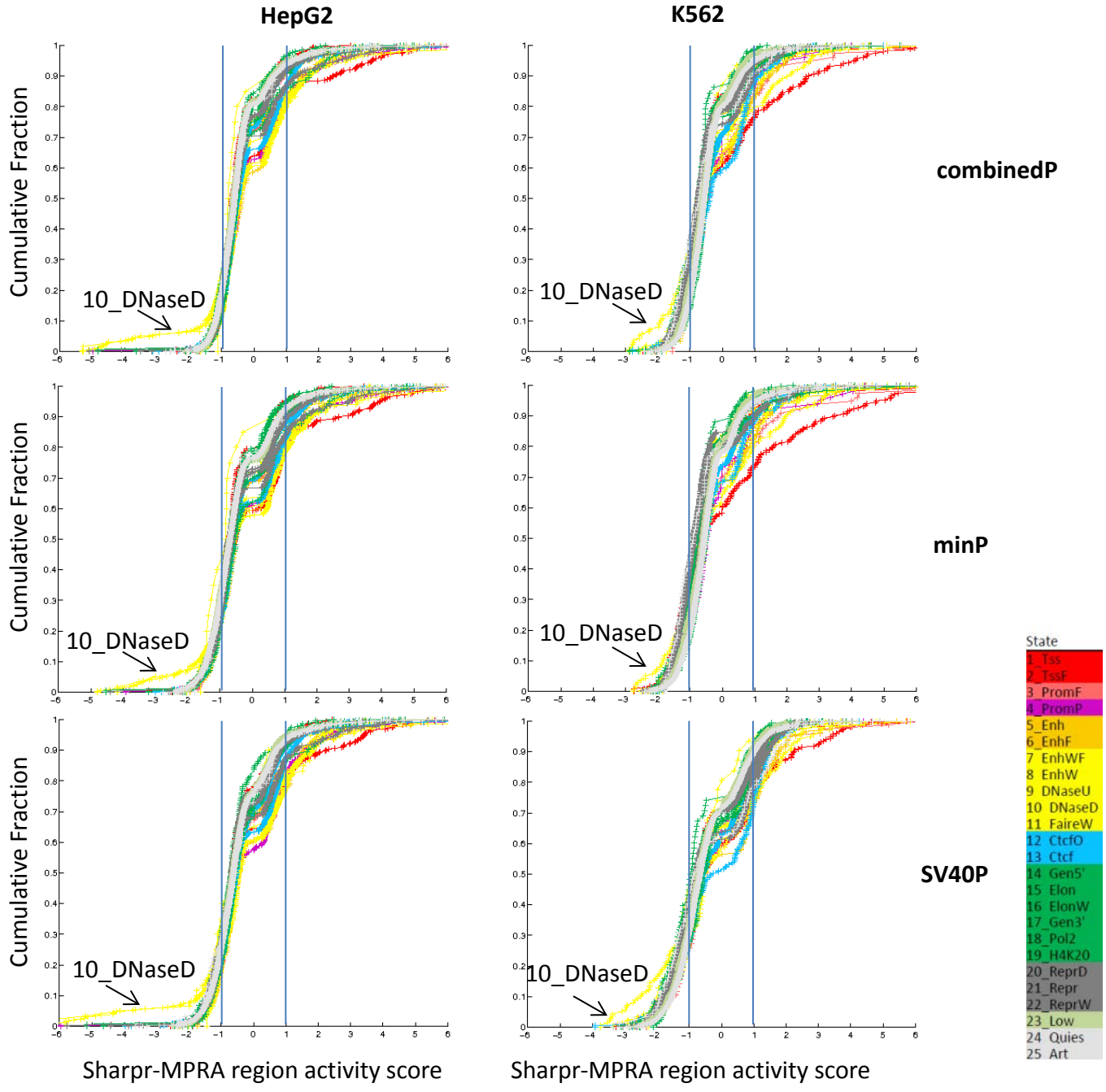
Supplementary Figure 33 – Cumulative distribution of region activity scores for each chromatin state. Cumulative fraction of regions (y-axis) showing a MaxPos activity score greater than indicated (on the x-axis) for each chromatin state (colors, see definitions in **Supplementary Fig. 5**) based on the combinedP score (top row), minP score (middle row) and SV40P score (bottom row) for HepG2 (left column) and K562 (right column), and shown for: **(a)** all selected regions from all four cell types that are DNase regions in the tested cell type (N=5840 in HepG2, N=6682 in K562); **(b**, next page**)** the subset of DNase regions in the tested cell type that were selected in that cell type; **(c**, following page**)** all DNase regions selected in another cell type, that were not in a DNase region in the tested cell type. Each region is shown as a separate point (plus signs). Arrows highlight the active promoter and enhancer states (1_Tss, 5_Enh) which had the greatest fraction of regions with strongly activating MaxPos scores, and the single-cut DNase state (10_DNaseD) which had an increased presence among regions with the most repressive MaxPos scores.

b. DNase Matched (N=3930 in HepG2, N=3930 in K562)

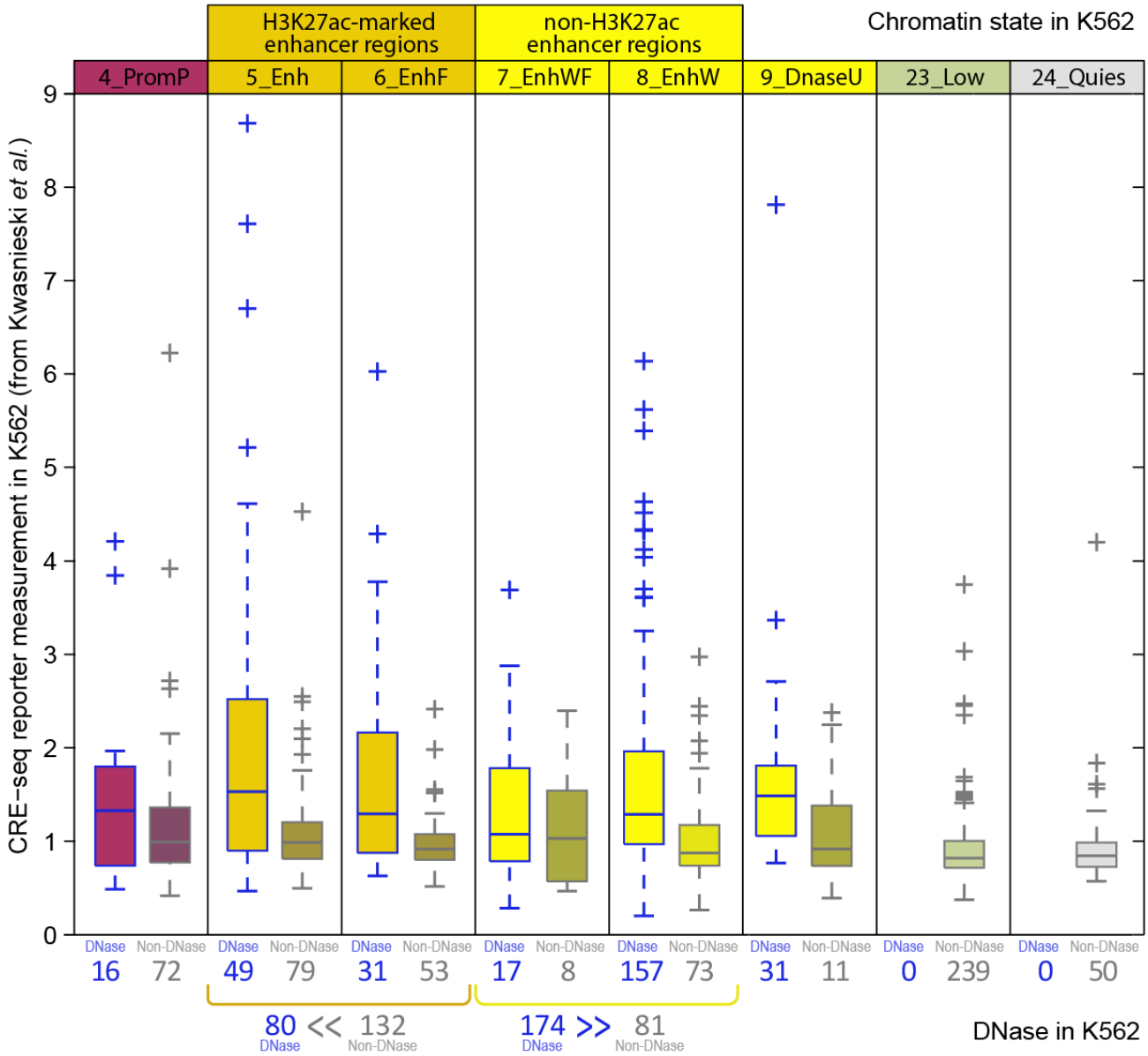


Supplementary Figure 33b – Cumulative distribution of region activity scores for each chromatin state. (a) Previous page. **(b)** Cumulative fraction of regions (y-axis) showing a MaxPos activity score greater than indicated (on the x-axis) for each chromatin state (colors, see definitions in **Supplementary Fig. 5**) based on the combinedP score (top row), minP score (middle row) and SV40P score (bottom row) for HepG2 (left column) and K562 (right column), and shown for the subset of DNase regions in the tested cell type that were selected in that cell type (N=3930 in each of HepG2 and K562); **(c)** Next page.

c. Non-DNase (N=9880 in HepG2, N=9038 in K562)

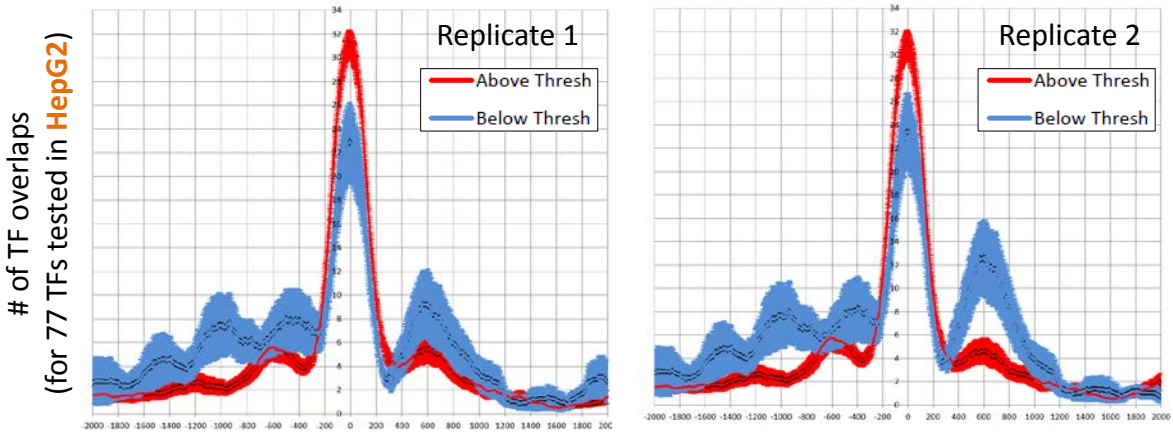


Supplementary Figure 33c – Cumulative distribution of region activity scores for each chromatin state. (a,b) Previous pages. **(c)** Cumulative fraction of regions (y-axis) showing a MaxPos activity score greater than indicated (on the x-axis) for each chromatin state (colors, see definitions in **Supplementary Fig. 5**) based on the combinedP score (top row), minP score (middle row) and SV40P score (bottom row) for HepG2 (left column) and K562 (right column), and shown for all DNase regions selected in another cell type, that were not in a DNase region in the tested cell type (N=9880 in HepG2, N=9038 in K562).

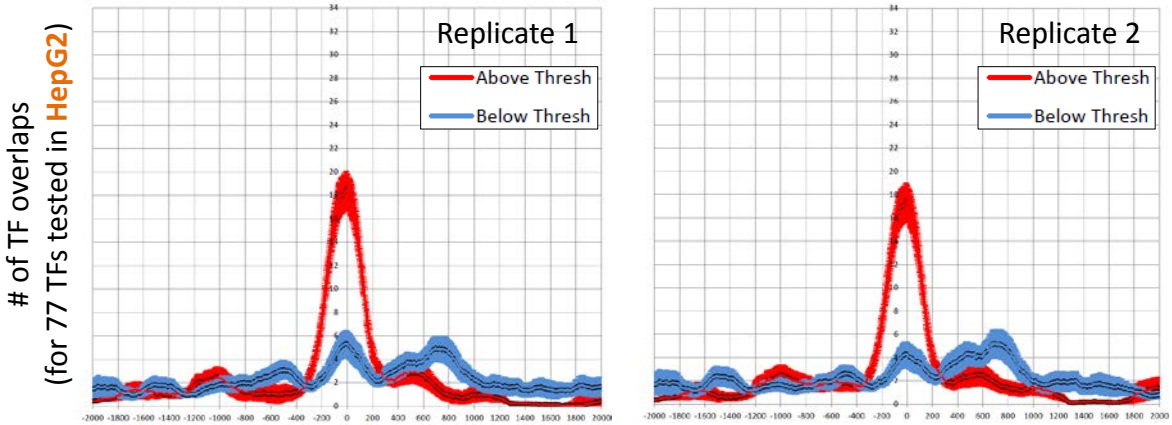


Supplementary Figure 34 – **CRE-seq activity distribution for ChromHMM chromatin states conditioned on presence or absence of DNase.** Average CRE-seq expression data[31] in K562 (y-axis) partitioned by ChromHMM state in K562 (fill color, see **Supplementary Fig. 5**) and by DNase (blue lines) vs. non-DNase (grey lines) based on the center position of tested segments in Ref. 31. The study reported lower CRE-seq activity for H3K27ac-marked enhancer regions, and higher activity for non-H3K27ac enhancer regions, contrary to the previous literature and to our results here and elsewhere[2,4,30,59]. As the figure shows, this reversal in effect is partly explained by the study mixing together two very different classes of regions (DNase vs. non-DNase), whose representation was highly imbalanced between the two groups compared, a statistical effect known as Simpson's Paradox. Only 80 of 212 H3K27ac-marked enhancer regions tested captured DNase sites, compared to 174 of 255 non-H3K27ac enhancer regions. When we separate the tested regions by both chromatin state and by DNase overlap, we find that regions overlapping DNase sites show higher median CRE-seq activity than non-DNase regions within the same chromatin state (as we saw in **Fig. 6b** and **Supplementary Fig. 31b**), and that H3K27ac-marked enhancer regions show higher median activity than non-H3K27ac enhancer regions within DNase regions (as we saw in **Fig. 6a** and **Supplementary Fig. 31a**). This emphasizes the importance of centering tested elements on candidate driver nucleotides (e.g. regulatory motifs, H3K27ac dips, DNase peaks), as we did here and elsewhere[30]. This is especially true when testing short elements without tiling (130-bp in the case of CRE-seq), as positioning the elements without such evidence may exclude driver regulatory nucleotides from tested regions, which can have a strong effect on reporter activity, as we show here. All states with >5 regions tested shown. Boxplots generated with the Matlab boxplot command.

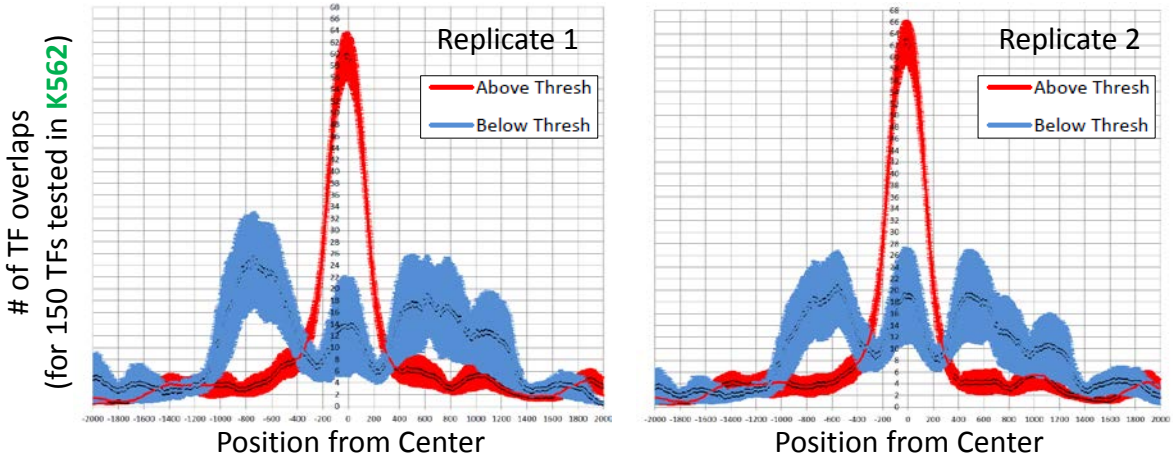
a. 100 pilot regions with high HepG2 dip scores tested in HepG2 vs. TF binding in HepG2.



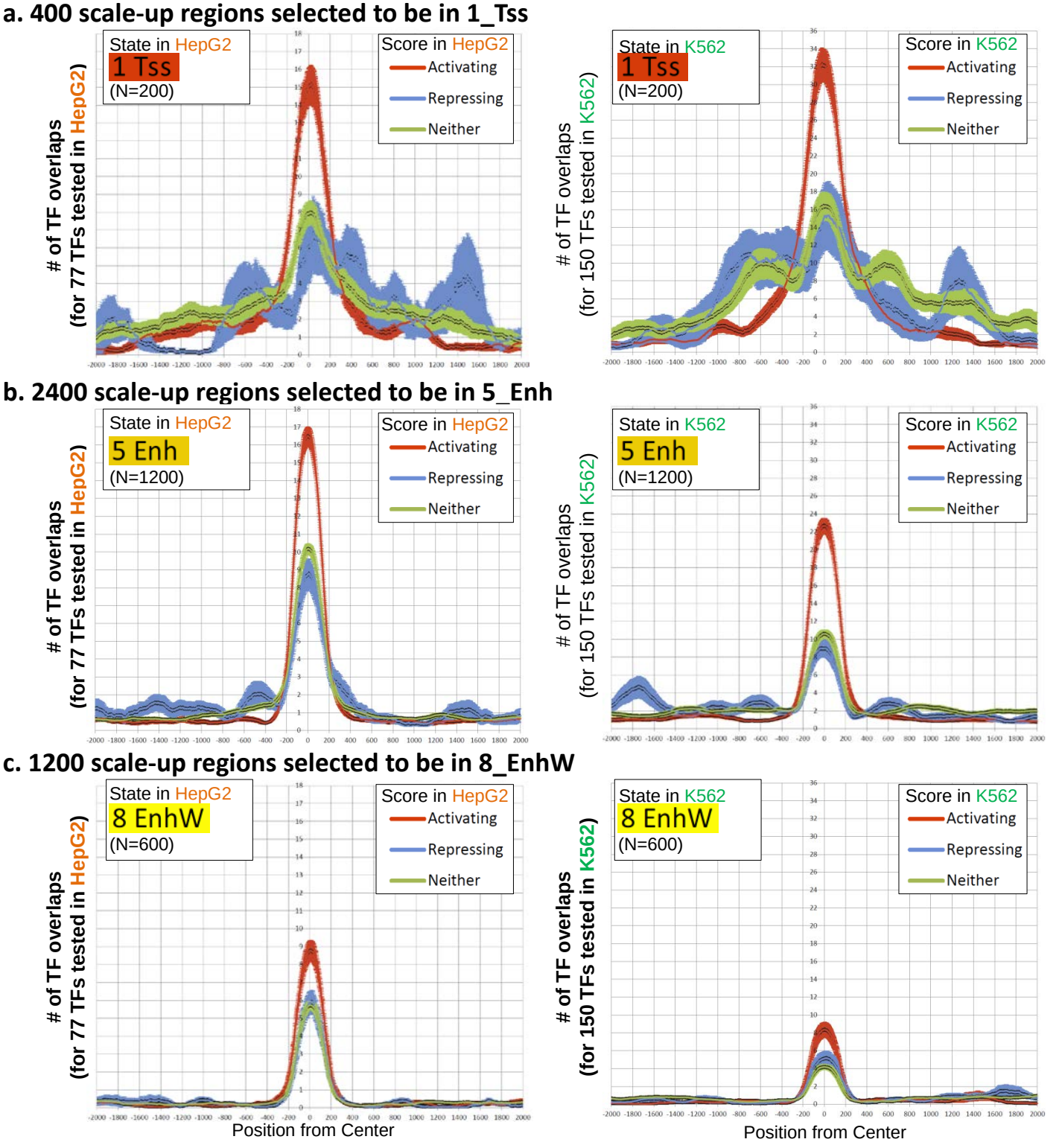
b. 100 pilot regions at a range of HepG2 dip scores tested in HepG2 vs. TF binding in HepG2.



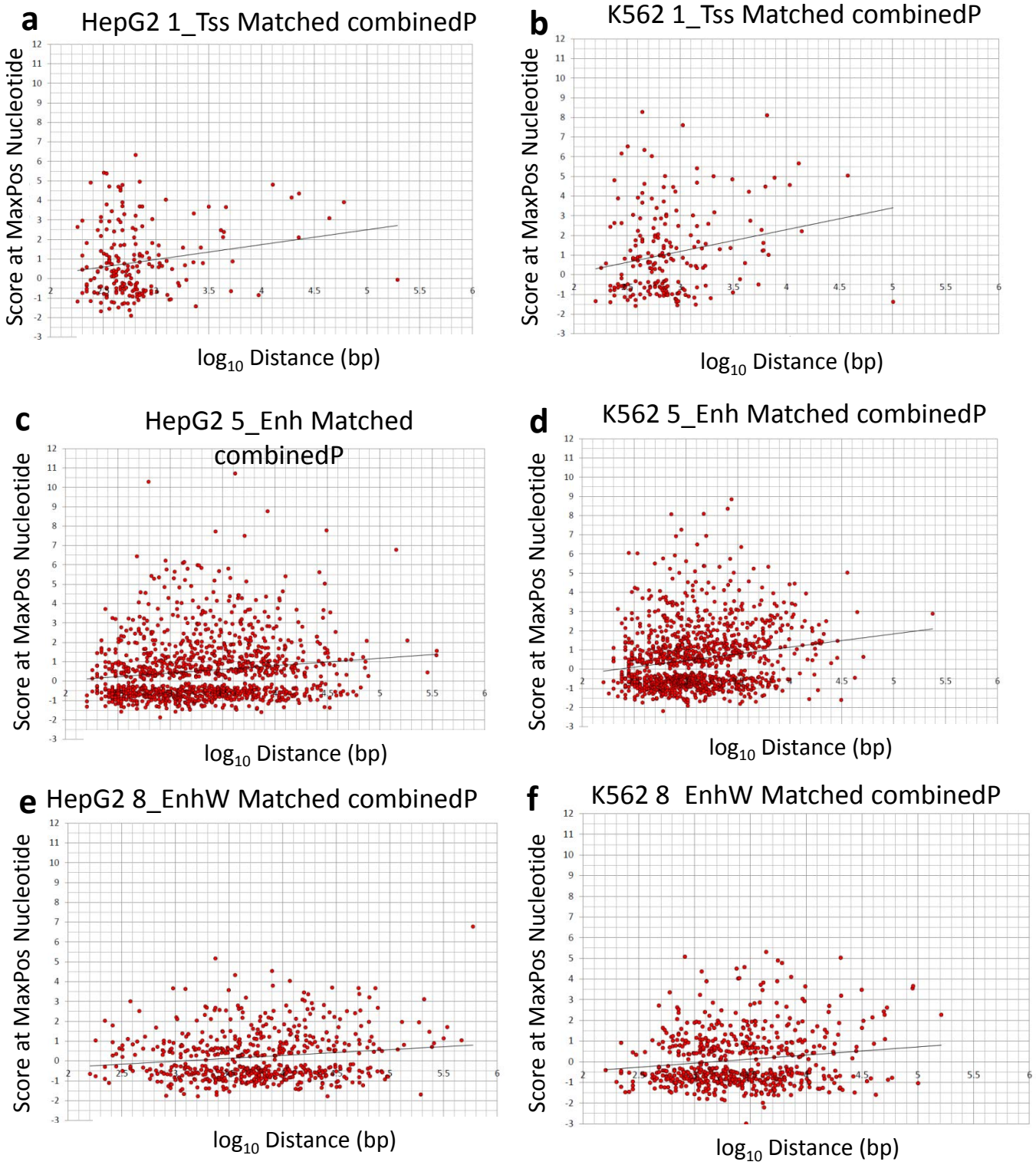
c. 50 pilot regions with high K562 dip scores tested in K562 vs. TF binding in K562.



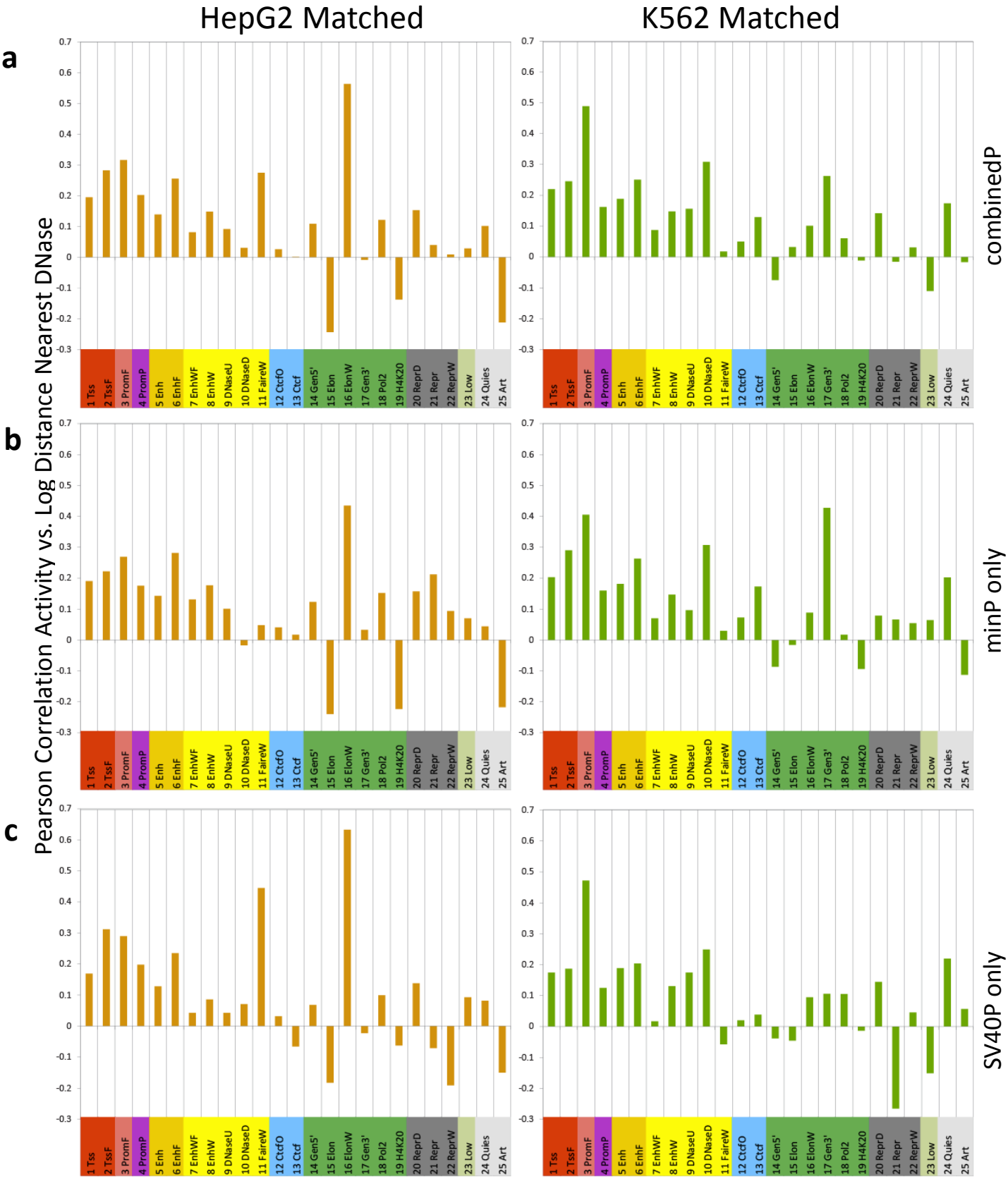
Supplementary Figure 35 – *In vivo* TF binding is higher and more centrally positioned for higher-scoring regions in pilot experiments. (a,b) Number of HepG2 ENCODE TF binding ChIP-Seq peak calls out of 77 sets (y-axis) overlapping regions with higher reporter scores (above threshold, red) and lower reporter scores (below-threshold, blue) for each nucleotide position relative to the selected regions center (x-axis) for: **(a)** the 100 regions tested in HepG2 and selected to have the highest *in vivo* dip scores in HepG2; **(b)** and the 100 regions selected to represent a range of dip scores in HepG2 cells. **(c)** The same plot as in (a,b) except for the 50 regions selected based on being in an active enhancer in K562 cells and for the 150 ENCODE peak call sets from K562 cells. Reporter activity threshold selected at the 95th percentile of outmost tile offsets. Vertical error bars indicate one standard error. These plots indicate that regions with higher reporter activity show higher TF binding within the tested region (higher red peak in the center), and more concentrated TF binding (lower red peaks in the surrounding) compared to lower reporter activity regions.



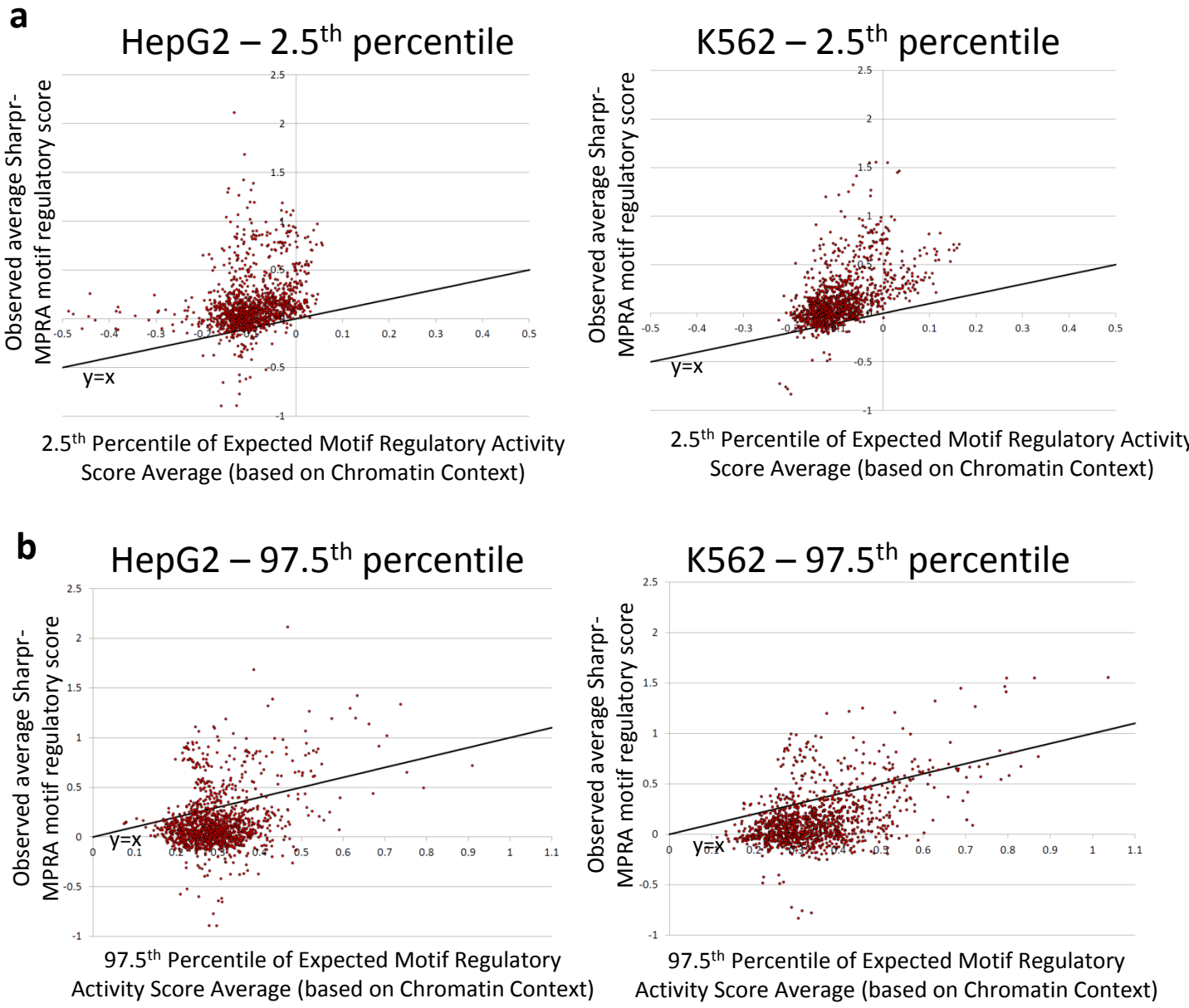
Supplementary Figure 36 – *In vivo* TF binding is higher and more centrally positioned for activating regions in active regulatory chromatin states for scale-up experiments. Number of HepG2 (left) and K562 (right) ENCODE TF binding ChIP-Seq peak calls (y-axis) overlapping regions with activating regulatory scores (red, combinedP MaxPos score≥1), repressive regulatory scores (blue, combinedP MaxPos score≤-1) or neither (green) for each position relative to the selected regions center (x-axis) for DNase regions selected in the matched cell type where they were tested (DNase matched) for: **(a)** 200 active promoter states (1_Tss) per cell type; **(b)** 1200 strong enhancer states (5_Enh); **(c)** 600 weak enhancer states (8_EnhW). Vertical error bars indicate one standard error. These plots indicate that regions with higher regulatory activity show higher TF binding within the tested region (higher red peak in the center), and more concentrated TF binding (lower red peaks in the surrounding) compared to lower regulatory activity regions.



Supplementary Figure 37 – Scatter Plot Relationship between Region Activity and Distance to Nearest DNase Site for Selected Chromatin States. (a,b) (a) Scatter plot showing for each tiled region selected based on being in a DNase site in promoter state 1_Tss in HepG2 cells on the x-axis the $\log_{10}$ distance in bp to the nearest DNase site in HepG2 (excluding itself) and on the y-axis the Sharpr-MPRA regulatory activity score at the MaxPos nucleotide for the region based on the HepG2 combinedP data. In black is shown a line of best fit for the scatter plots. **(b)** The same as (a) except for K562 cells. **(c,d)** The same as (a,b) except for enhancer state 5_Enh. **(e,f)** The same as (a,b) except for weak enhancer state 8_EnhW. The plots show for these states that more isolated DNase sites tend to have greater positive activity. The correlation for all states are shown in **Supplementary Fig. 38**.



Supplementary Figure 38 – Correlation of Each Chromatin State for Region Activity and Distance to Nearest DNase Site. Bar graph showing for the regions selected based on each of the 25 chromatin states the correlation between the log of the distance to the nearest DNase site in that cell type (excluding itself) and the regulatory activity score at the MaxPos nucleotide for HepG2 chromatin states (left) and K562 chromatin states (right), using: **(a)** the combined minP and SV40P score; **(b)** the minP data only; **(c)** the SV40P data only. Full scatter plot for selected states are found in **Supplementary Fig. 37**.



Supplementary Figure 39 – 95% Confidence Interval Bounds for Expected Motif Regulatory Activity Score Averages based on Chromatin Context. (a) Scatter plots analogous to **Fig. 6c** showing the observed average regulatory activity (y-axis) compared with the *lower bound* of a 95% confidence interval (2.5th percentile) on the motif regulatory score expected based on the chromatin context (x-axis) for HepG2 (left) and K562 (right). **(b)** Same as (a) except the x-axis represents the *upper bound* on a 95% confidence interval (97.5th percentile).

a

HepG2 PI-conserved elements (AUC 1%/expected)
Active Ranking

	0.1	0.5	1	5	10	50	100
0.1	2.41	2.42	2.43	2.49	2.51	2.60	2.65
0.5		2.42	2.44	2.49	2.52	2.60	2.66
1			2.44	2.49	2.52	2.61	2.67
5				2.52	2.55	2.67	2.71
10					2.56	2.70	2.73
50						2.73	2.71
100							2.69

K562 PI-conserved elements (AUC 1%/expected)
Active Ranking

	0.1	0.5	1	5	10	50	100
0.1	2.41	2.42	2.42	2.45	2.47	2.53	2.54
0.5		2.42	2.42	2.45	2.48	2.54	2.54
1			2.43	2.45	2.48	2.55	2.55
5				2.47	2.50	2.57	2.57
10					2.51	2.59	2.58
50						2.62	2.60
100							2.59

b

Repressive Ranking

	0.1	0.5	1	5	10	50	100
0.1	1.69	1.69	1.68	1.66	1.63	1.54	1.48
0.5		1.69	1.68	1.66	1.62	1.54	1.48
1			1.68	1.65	1.62	1.53	1.48
5				1.63	1.60	1.52	1.46
10					1.59	1.48	1.40
50						1.31	1.24
100							1.21

Repressive Ranking

	0.1	0.5	1	5	10	50	100
0.1	1.43	1.43	1.42	1.41	1.38	1.24	1.17
0.5		1.42	1.42	1.41	1.38	1.24	1.16
1			1.42	1.40	1.37	1.23	1.16
5				1.37	1.33	1.21	1.13
10					1.31	1.17	1.10
50						1.10	1.04
100							0.99

c

Absolute Ranking

	0.1	0.5	1	5	10	50	100
0.1	2.46	2.47	2.48	2.55	2.59	2.72	2.77
0.5		2.47	2.48	2.56	2.60	2.72	2.78
1			2.49	2.56	2.60	2.73	2.79
5				2.58	2.63	2.78	2.82
10					2.64	2.81	2.81
50						2.78	2.67
100							2.61

Absolute Ranking

	0.1	0.5	1	5	10	50	100
0.1	2.42	2.43	2.43	2.46	2.48	2.54	2.55
0.5		2.43	2.43	2.46	2.48	2.55	2.56
1			2.43	2.46	2.49	2.55	2.56
5				2.48	2.50	2.58	2.57
10					2.52	2.59	2.59
50						2.60	2.53
100							2.49

d

HepG2 PI-conserved elements (AUC 5%/expected)
Active Ranking

	0.1	0.5	1	5	10	50	100
0.1	1.85	1.86	1.86	1.88	1.90	1.94	1.93
0.5		1.86	1.86	1.89	1.90	1.94	1.93
1			1.87	1.89	1.90	1.94	1.93
5				1.91	1.92	1.95	1.94
10					1.93	1.95	1.93
50						1.93	1.88
100							1.85

K562 PI-conserved elements (AUC 5%/expected)
Active Ranking

	0.1	0.5	1	5	10	50	100
0.1	1.82	1.82	1.83	1.84	1.86	1.88	1.88
0.5		1.83	1.83	1.85	1.86	1.88	1.88
1			1.83	1.85	1.86	1.89	1.88
5				1.86	1.87	1.89	1.88
10					1.88	1.89	1.87
50						1.86	1.81
100							1.77

e

Repressive Ranking

	0.1	0.5	1	5	10	50	100
0.1	1.28	1.28	1.28	1.27	1.26	1.21	1.18
0.5		1.28	1.28	1.27	1.26	1.21	1.18
1			1.28	1.27	1.26	1.20	1.18
5				1.26	1.25	1.18	1.15
10					1.23	1.15	1.13
50						1.08	1.05
100							1.03

Repressive Ranking

	0.1	0.5	1	5	10	50	100
0.1	1.18	1.18	1.18	1.16	1.15	1.09	1.06
0.5		1.18	1.18	1.16	1.15	1.09	1.06
1			1.18	1.16	1.15	1.08	1.06
5				1.14	1.12	1.06	1.03
10					1.11	1.04	1.02
50						0.99	0.97
100							0.96

f

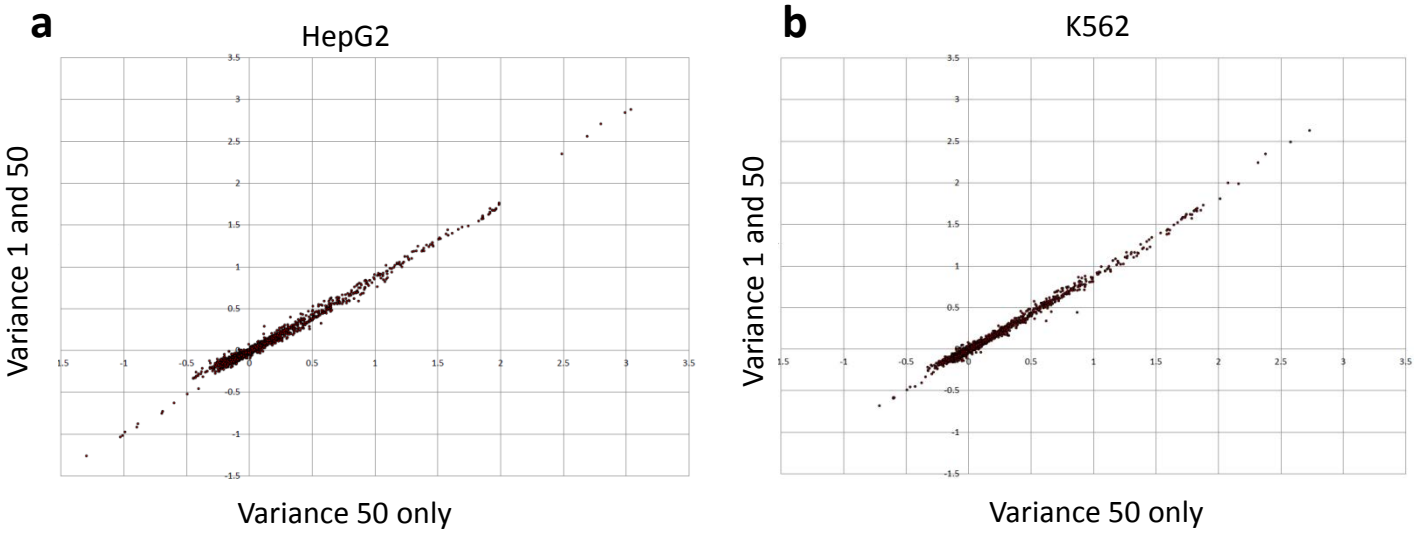
Absolute Ranking

	0.1	0.5	1	5	10	50	100
0.1	1.88	1.89	1.89	1.91	1.92	1.95	1.93
0.5		1.89	1.89	1.91	1.92	1.95	1.93
1			1.90	1.92	1.93	1.95	1.93
5				1.93	1.95	1.96	1.92
10					1.95	1.95	1.90
50						1.88	1.78
100							1.75

Absolute Ranking

	0.1	0.5	1	5	10	50	100
0.1	1.82	1.82	1.83	1.84	1.85	1.86	1.83
0.5		1.83	1.83	1.85	1.85	1.86	1.83
1			1.83	1.85	1.86	1.86	1.83
5				1.86	1.86	1.86	1.82
10					1.87	1.86	1.81
50						1.80	1.72
100							1.67

Supplementary Figure 40 – Conserved Element Recovery as a Function of Variance Prior Parameters (a) The table shows for each indicated combination of the variance prior parameters for the regulatory activity variables (see Methods), when ranking the resulting nucleotides by how activating their regulatory activity score is, the ratio of the Area under the ROC curve (AUC) up to a 1% false positive rate for recovering SiPhy-PI conserved elements[42,65] to what would be expected by random guessing (0.00005) for HepG2 (left) and K562 (right). **(b)** The same as (a) except based on ranking nucleotides by how repressive their regulatory score is. **(c)** The same as (a) and (b) except ranking nucleotides based on the absolute value of their regulatory score. **(d-f)** The same as (a-c) except based on an AUC to a 5% false positive rate, where the expected AUC to a 5% false positive rate by random guessing is 0.00125. The color shading scale is specific to each heatmap.



Supplementary Figure 41 – Effect on motif average Sharpr-MPRA scores when using just the higher-variance parameter. Scatter plots for **(a)** HepG2 and **(b)** K562 cells comparing the average motif regulatory score using (x-axis) just the larger variance parameter (50) and (y-axis) using the combined results from two variance parameters (1,50) as reported in **Fig. 4a**. Each point corresponds to one motif. Only motifs with >10 instances are shown. In both plots the correlation of the values are >0.99.

Supplementary Note

Here we derive the equations for the terms of $\Sigma_{M_{r,t},M_{r,t}}$ and $\Sigma_{A_{r,t},M_{r,t}}$ presented in the Methods.

For $\Sigma_{M_{r,t},M_{r,t}}$ we have

$$\Sigma_{M_{r,t,u},M_{r,t,u}} = \sigma_{m_r}^2 + \sigma_a^2/N$$

$$\Sigma_{M_{r,t,u},M_{r,t,v}} = \frac{\sigma_a^2(N - |u - v|)}{N^2} \text{ if } (0 < |u - v| < N)$$

$$\Sigma_{M_{r,t,u},M_{r,t,v}} = 0 \text{ if } (|u - v| \geq N)$$

The expression for $\Sigma_{M_{r,t,u},M_{r,t,u}}$ above follows from observing

$$var(M_{r,t,u}) = var\left(\sigma_{m_r} Z_{r,t,u} + \left(\frac{1}{N}\sum_{l=0}^{N-1} A_{r,t,u+l}\right)\right) = \sigma_{m_r}^2 + \frac{1}{N^2}\sum_{l=0}^{N-1} var(A_{r,t,u+l}) = \sigma_{m_r}^2 + \sigma_a^2/N$$

where $Z_{r,t,u}$ represents a standard normal random variable.

The expression for $\Sigma_{M_{r,t,u},M_{r,t,v}}$ follow since

$$cov(M_{r,t,u}, M_{r,t,v}) = cov\left(\sigma_{m_r} Z_{r,t,u} + \frac{1}{N}\sum_{l=0}^{N-1} A_{r,t,u+l}\,, \sigma_{m_r} Z_{r,t,v} + \frac{1}{N}\sum_{l=0}^{N-1} A_{r,t,v+l}\right)$$

$$= cov\left(\frac{1}{N}\sum_{l=0}^{N-1} A_{r,t,u+l}\,, \frac{1}{N}\sum_{l=0}^{N-1} A_{r,t,v+l}\right)$$

where $Z_{r,t,u}$ and $Z_{r,t,v}$ represent independent standard normal random variables. The expression is 0 if $(|u - v| \geq N)$ and otherwise is

$$\frac{1}{N^2}\sum_{l=|u-v|}^{N-1} var(A_{r,t,(\min(u,v)+l)}) = \frac{\sigma_a^2(N - |u - v|)}{N^2}$$

For $\Sigma_{A_{r,t},M_{r,t}}$ we have:

$$\Sigma_{A_{r,t,k},M_{r,t,u}} = \frac{\sigma_a^2}{N} \text{ if } u \leq k < u + N$$

$$\Sigma_{A_{r,t,k},M_{r,t,u}} = 0 \text{ otherwise}$$

This follows from observing

$$cov(A_{r,t,k}, M_{r,t,u}) = cov\left(A_{r,t,k}, \frac{1}{N}\sum_{l=0}^{N-1} A_{r,t,u+l}\right)$$

$$= \frac{1}{N}\sum_{l=0}^{N-1} cov(A_{r,t,k}, A_{r,t,u+l})$$

All terms are 0 except if $u \leq k < u + N$ in which case the expression reduces to

$$= \frac{1}{N} var(A_{r,t,k})$$