PNAS
www.pnas.org

Supplementary Information for

Positional specificity of different transcription factor classes within enhancers

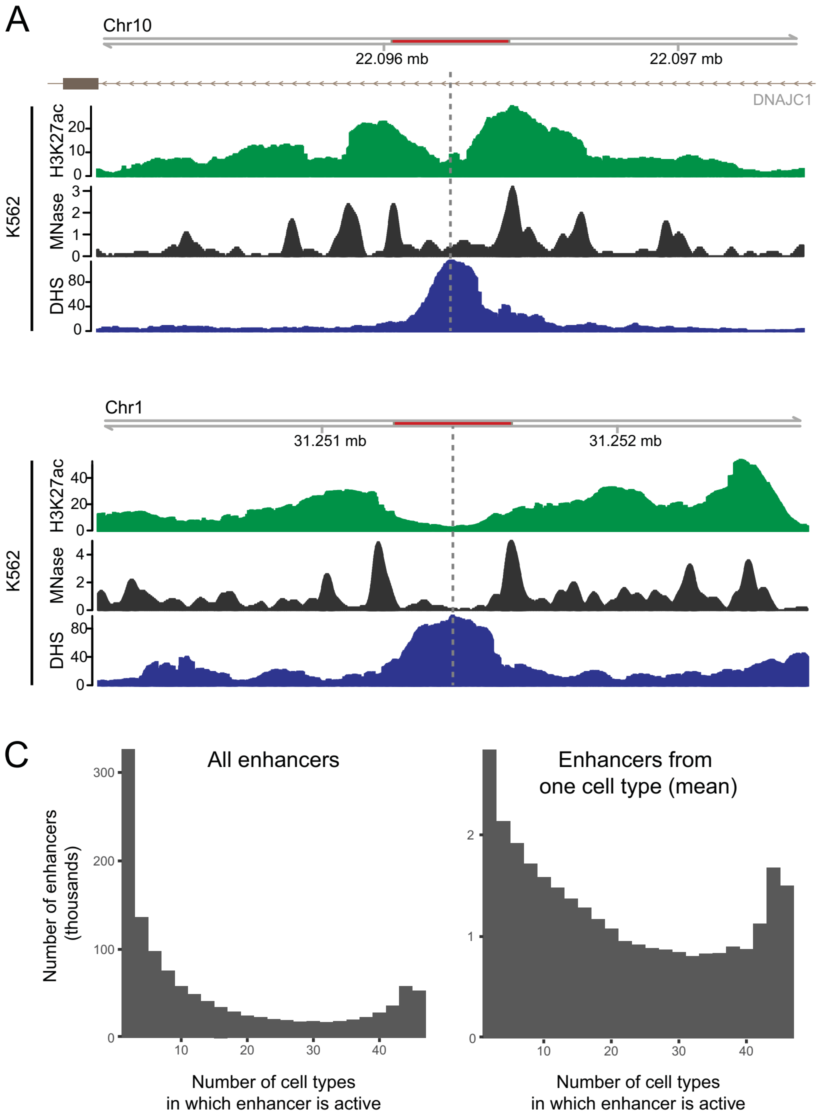
Sharon R. Grossman, Jesse Engreitz, John P. Ray, Tung H. Nguyen, Nir Hacohen, Eric S. Lander

Eric S. Lander
Email: lander@broadinstitute.org

This PDF file includes:

Figs. S1 to S13
Tables S1 to S3

www.pnas.org/cgi/doi/10.1073/pnas.1804663115



B

Cell Types

- IMR90 fetal lung fibroblasts Cell Line
- H9 Cells
- H1 Cells
- iPS DF 6.9 Cells
- iPS DF 19.11 Cells
- H1 Derived Neuronal Progenitor Cultured Cells
- H1 BMP4 Derived Mesendoderm Cultured Cells
- H1 BMP4 Derived Trophoblast Cultured Cells
- H1 Derived Mesenchymal Stem Cells
- Primary T cells from peripheral blood
- Primary monocytes from peripheral blood
- Primary hematopoietic stem cells G-CSF-mobilized
- Primary B cells from peripheral blood
- Primary Natural Killer cells from peripheral blood
- Foreskin Fibroblast Primary Cells skin 1
- Foreskin Fibroblast Primary Cells skin 2
- Foreskin Melanocyte Primary Cells skin 1
- Fetal Thymus
- Psoas Muscle
- Fetal Muscle Trunk
- Fetal Muscle Leg
- Fetal Stomach
- Fetal Intestine Small
- Fetal Intestine Large
- Small Intestine
- Gastric
- Ovary
- Fetal Adrenal Gland
- Placenta
- Pancreas
- A549 EtOH 0.02pct Lung Carcinoma Cell Line
- GM12878 Lymphoblastoid Cells
- HeLa-S3 Cervical Carcinoma Cell Line
- HepG2 Hepatocellular Carcinoma Cell Line
- HMEC Mammary Epithelial Primary Cells
- HSMM Skeletal Muscle Myoblasts Cells
- HSMM cell derived Skeletal Muscle Myotubes Cells
- HUVEC Umbilical Vein Endothelial Primary Cells
- K562 Leukemia Cells
- Monocytes-CD14+ RO01746 Primary Cells
- NH-A Astrocytes Primary Cells
- NHDF-Ad Adult Dermal Fibroblast Primary Cells
- NHEK-Epidermal Keratinocyte Primary Cells
- Jurkat T lymphocytes
- HCT 116 colon colorectal carcinoma
- U937 myeloid lymphocyte

Figure S1. Selection of nucleosome-depleted regions used in analysis. (A) Chromatin structure around two example putative regulatory regions from K562 cells. DNaseI hypersensitive regions (DHS) (blue track) surrounded by regions marked with H3K27ac (green track) centered around the peak of the signal (dotted line) were selected. Micrococcal nuclease accessibility data (MNase-seq, black track) shows these NDRs are flanked by well-positioned nucleosomes. (B) List of 47 cell types in which 577,669 putative regulatory NDRs were identified. (C) Distribution of number of cell types with activity (defined by significant DHS/ATAC-seq and H3K27ac signal) for all NDRs (left) and average distribution (right) for NDRs active in a cell type (averaged across all cell types).

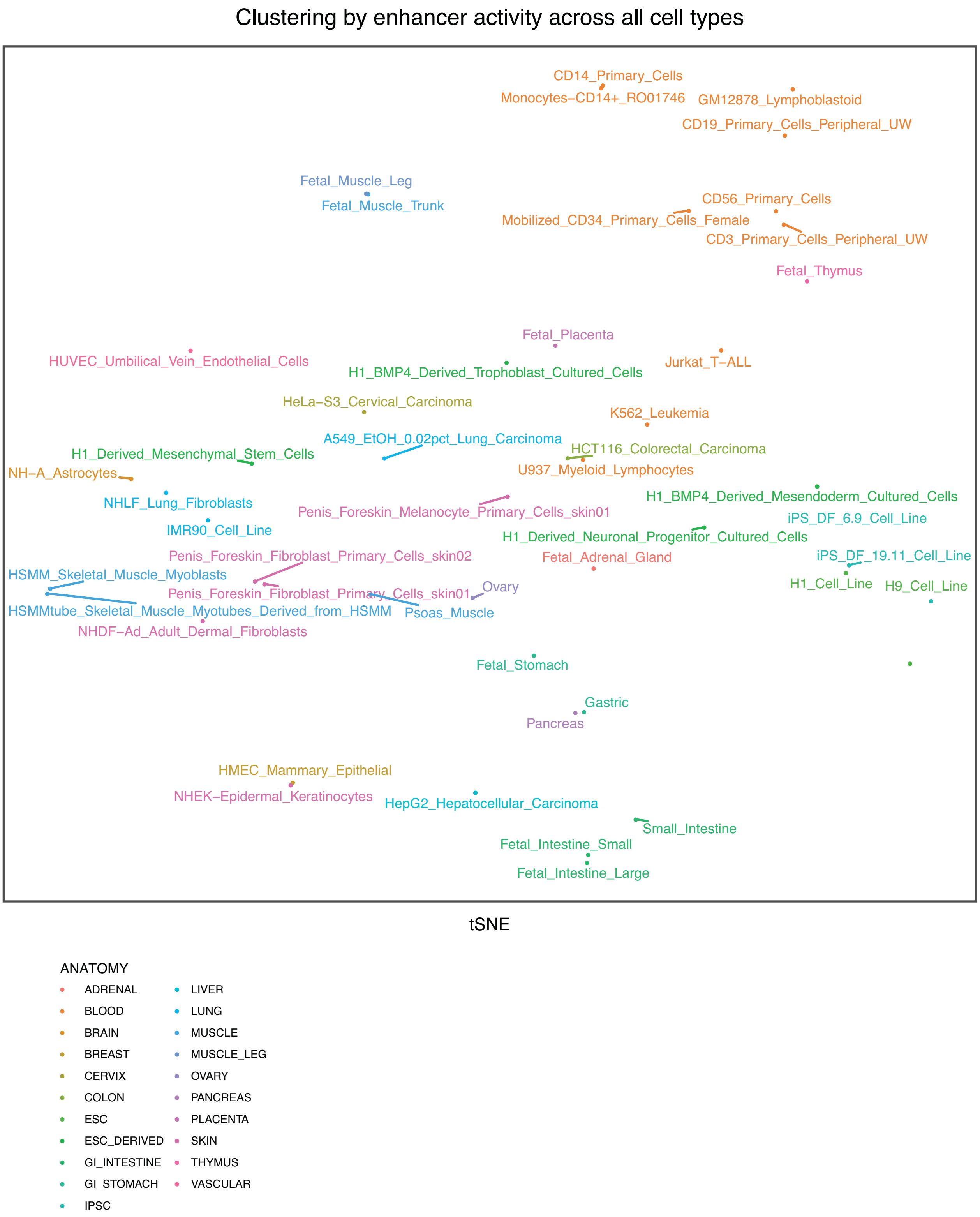


Fig. S2. Cell types clustered by activity of putative regulatory elements. t-SNE plot of 47 cell types clustered based on the activity of the 577,669 putative regulatory elements. Regions were defined as active in a given cell type if they overlapped significant DHS/ATAC-seq and H3K27ac ChIP-seq peaks. Each point represents one cell type. Cell type are colored based on their anatomical origin.

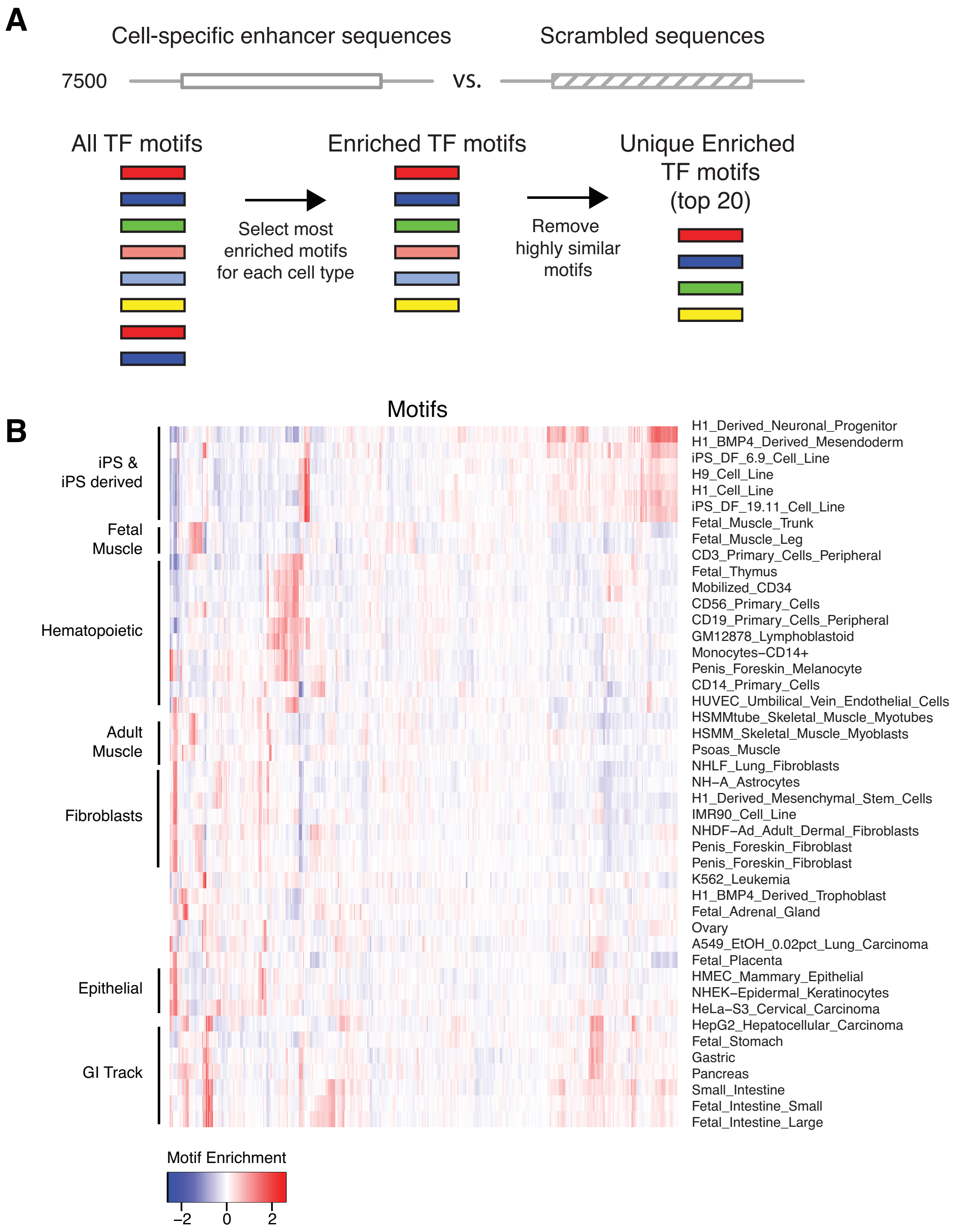


Fig. S3. Identification of enriched TF motifs in each cell type. (A) TF motifs from JASPAR, TRANSFAC, and CIS-BP enriched in genomic sequences of NDRs relative to shuffled sequences were identified for each cell type. For TFs with multiple motifs in the combined database, the most enriched motif corresponding to each TF in each cell type was selected, and motifs with high similarity to a more highly enriched motif were removed (see Methods). The 20 most enriched TF motifs from the filtered list were used for positional analysis. (B) Enrichment of TF motifs across cell types. Related cell types generally show correlated motif enrichment profiles.

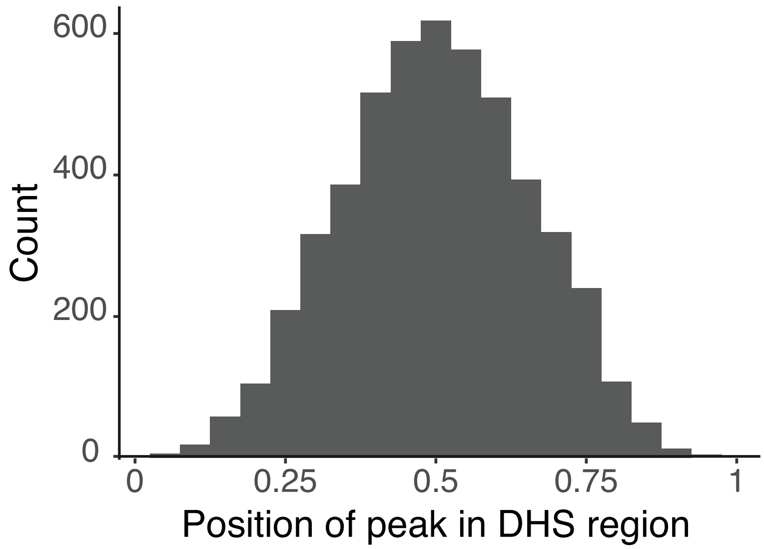
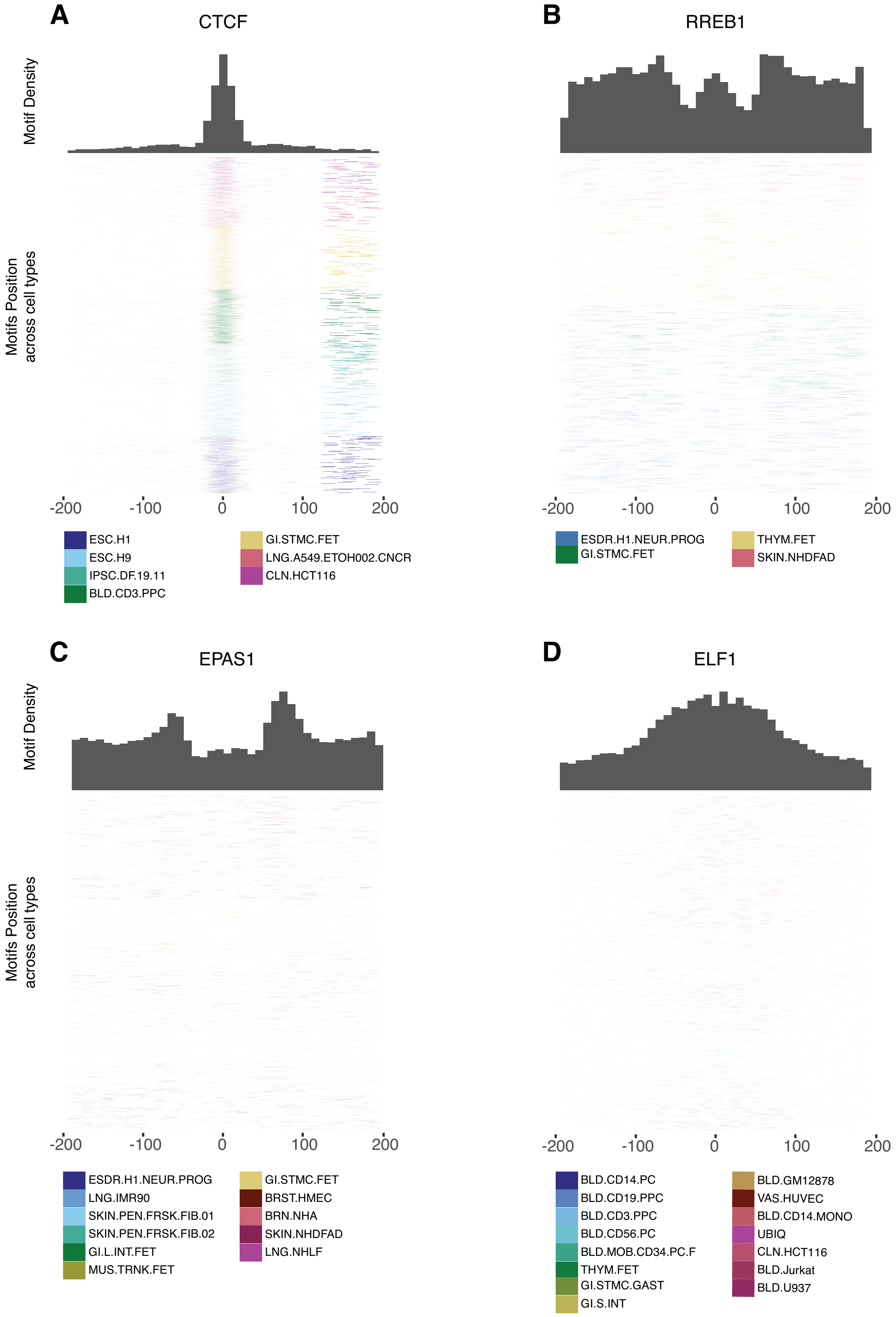


Fig. S4. Position of peak of DHS signal within full DHS region. Distribution of the relative position of the peak of the DHS signal within the overall DHS region (as defined by the MACS2 peakcaller). The peak of the signal falls within the central third of the region in 72% of NDRs.


A
CTCF
Motif Density
Motifs Position across cell types
-200
-100
0
100
200
ESC.H1
ESC.H9
IPSC.DF.19.11
BLD.CD3.PPC
GI.STMC.FET
LNG.A549.ETOH002.CNCR
CLN.HCT116
B
RREB1
-200
-100
0
100
200
ESDR.H1.NEUR.PROG
GI.STMC.FET
THYM.FET
SKIN.NHDFAD
C
EPAS1
Motif Density
Motifs Position across cell types
-200
-100
0
100
200
ESDR.H1.NEUR.PROG
LNG.IMR90
SKIN.PEN.FRSK.FIB.01
SKIN.PEN.FRSK.FIB.02
GI.L.INT.FET
MUS.TRNK.FET
GI.STMC.FET
BRST.HMEC
BRN.NHA
SKIN.NHDFAD
LNG.NHLF
D
ELF1
-200
-100
0
100
200
BLD.CD14.PC
BLD.CD19.PPC
BLD.CD3.PPC
BLD.CD56.PC
BLD.MOB.CD34.PC.F
THYM.FET
GI.STMC.GAST
GI.S.INT
BLD.GM12878
VAS.HUVEC
BLD.CD14.MONO
UBIQ
CLN.HCT116
BLD.Jurkat
BLD.U937

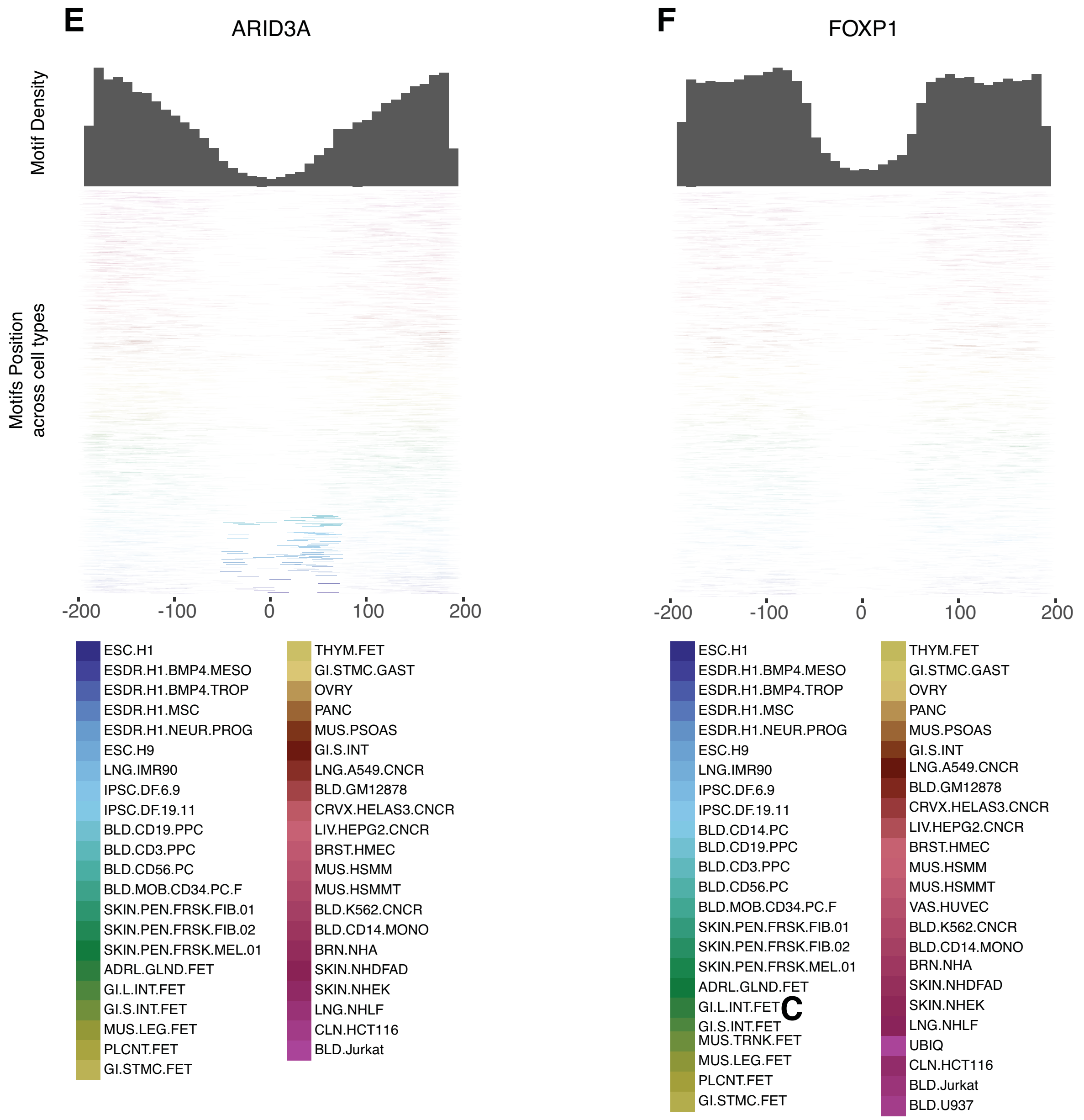


Fig. S5. Positional binding patterns of TF motifs show striking differences. Distribution of position of motif sites for CTCF (A), RREB1 (B), EPAS1 (C), ELF1 (D), ARID3A (E) and FOXP1 (F) across NDR regions, centered around the peak of the DHS signal. Histograms show density of motif sites in 10 bp bins tiled across the NDR (top). Heatmaps show the position of 10,000 motif sites in NDRs. Colors indicate motif sites in different cell types.

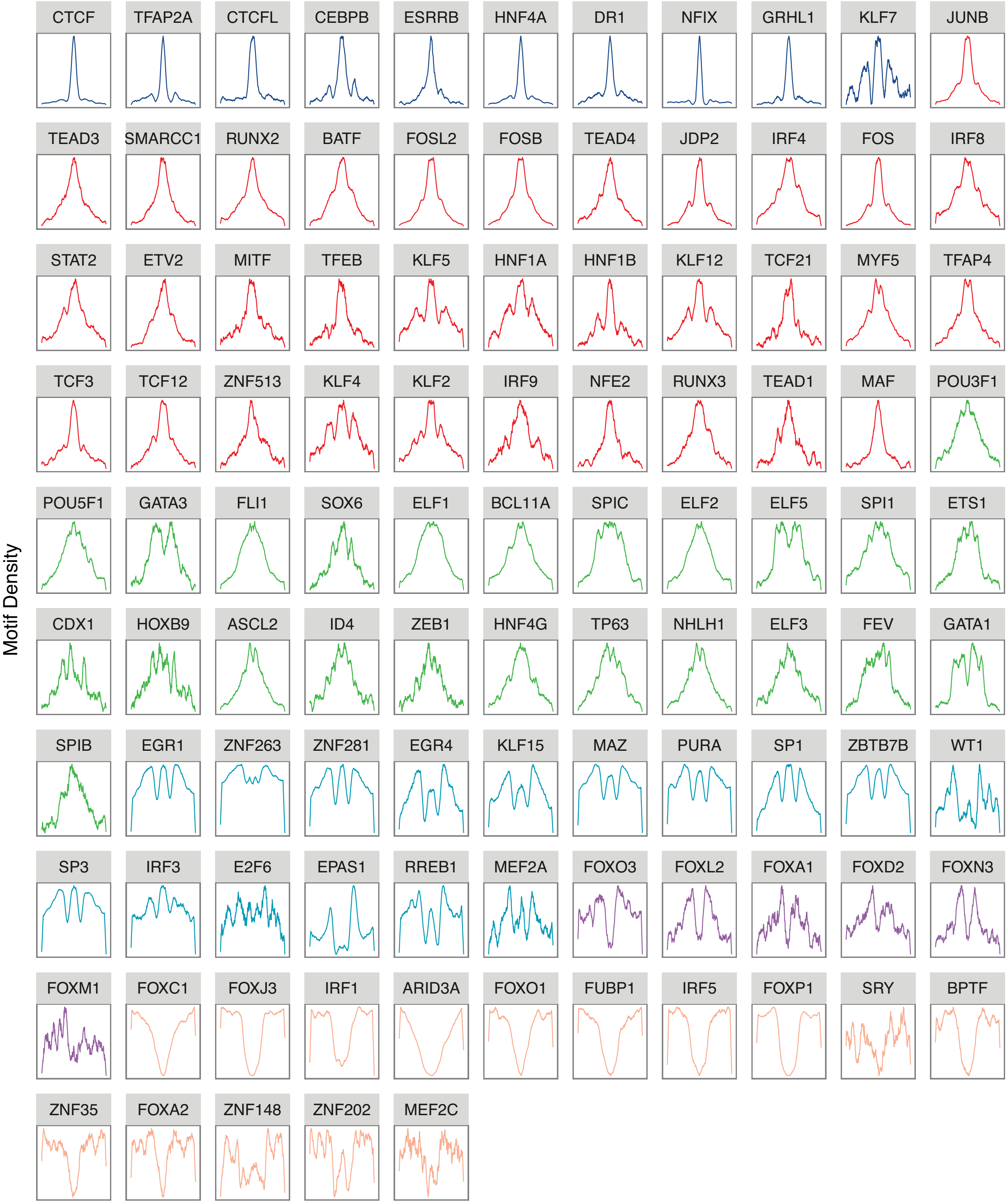


Fig. S6. Motif position profiles of 103 TF motifs. Motif position profiles for each TF were calculated using NDRs in all cell types that contained the motif among the top 20 most enriched motifs. Motif sites were collapsed to their central position, and the density of motif sites in 20-bp bins tiled every 1 bp across 400 bp centered around the peak of the DHS signal is shown. Motif position profiles are colored by their cluster (blue - cluster 1; red – cluster 2; green – cluster 3; cyan – cluster 4; purple – cluster 5; orange – cluster 6).

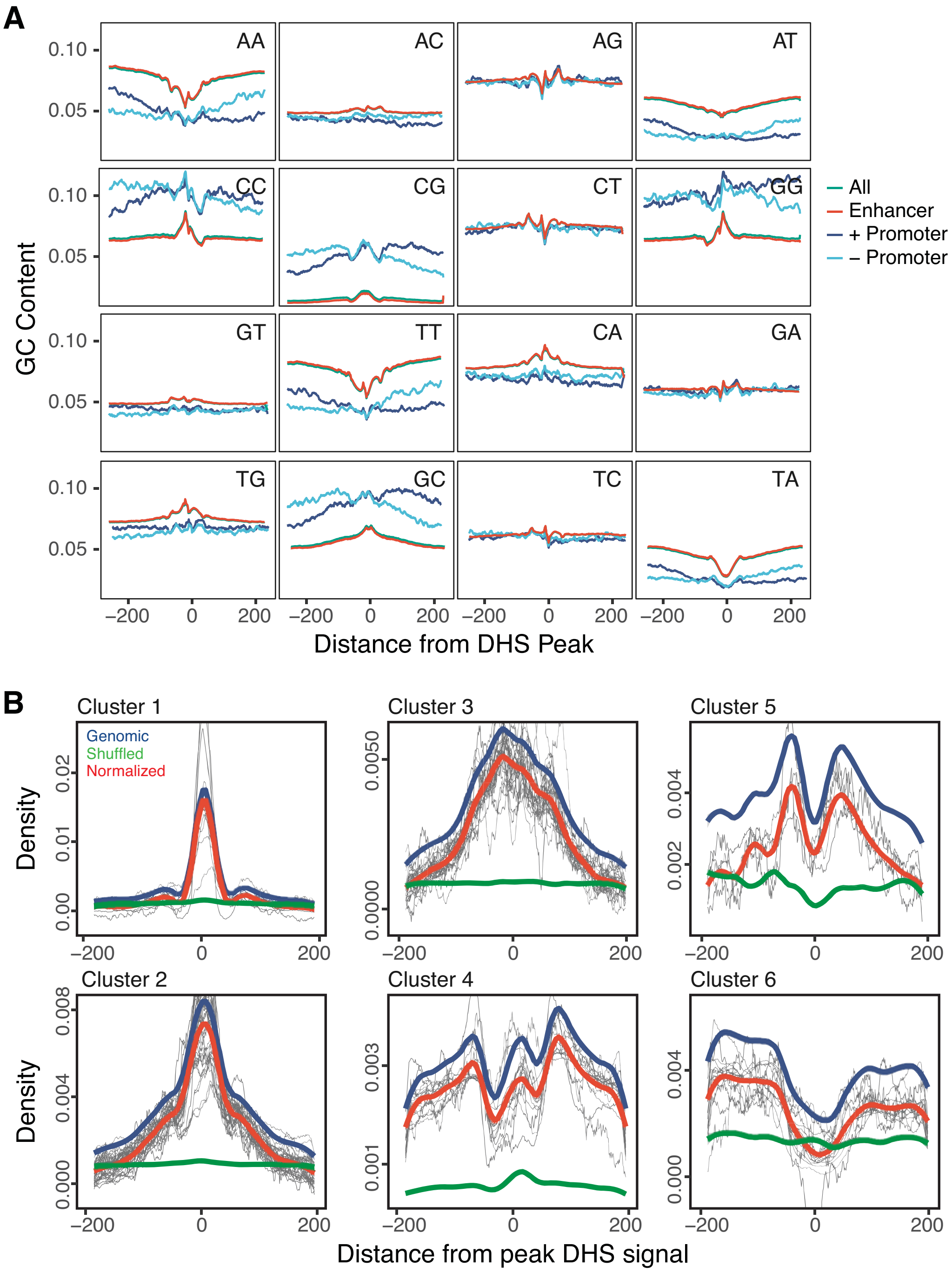


Fig. S7. Variation in dinucleotide content does not account for motif position patterns. (A) Average dinucleotide content as a function of distance from the peak of the DHS signal for all NDRs (green lines), distal enhancers (>1kb from annotated TSS; yellow lines), and promoters (<1kb from annotated TSS) on the plus strand (dark blue lines) and minus strand (light blue lines). (B) To control for variation in dinucleotide frequency across the regions, we calculated the background density profiles for each motif in shuffled sequences holding dinucleotide content at each position, and normalized the genomic density profiles by subtracting the background density profiles. Plots show the average cluster profiles for each motif in genomic sequences (blue lines, same as Fig. 5A), shuffled sequences (green lines), and the normalized density profiles (red lines).

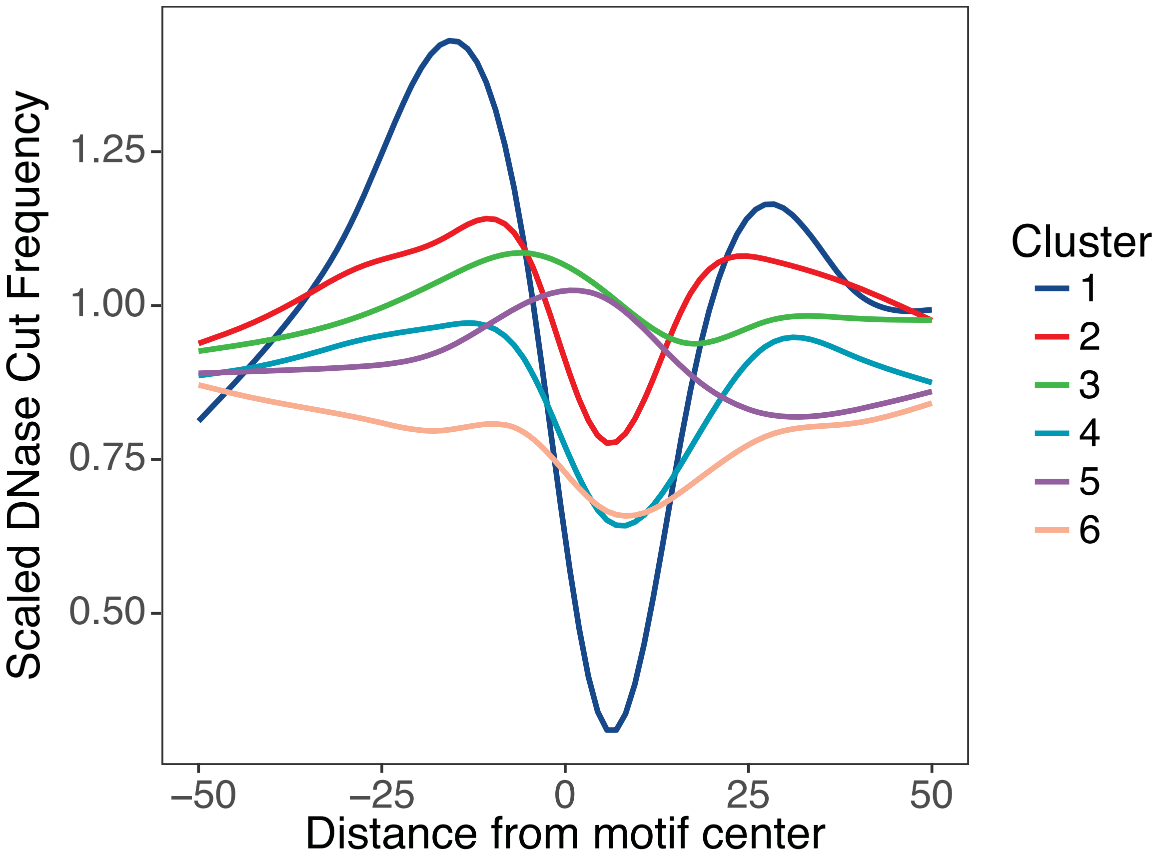


Fig. S8. Average DNase footprints of motifs in each TF cluster. Cut count profiles for each TF motif were averaged over all motif sites within NDRs. Cut count profiles for each TF were scaled to a mean of zero, and averaged over all TFs in each cluster.

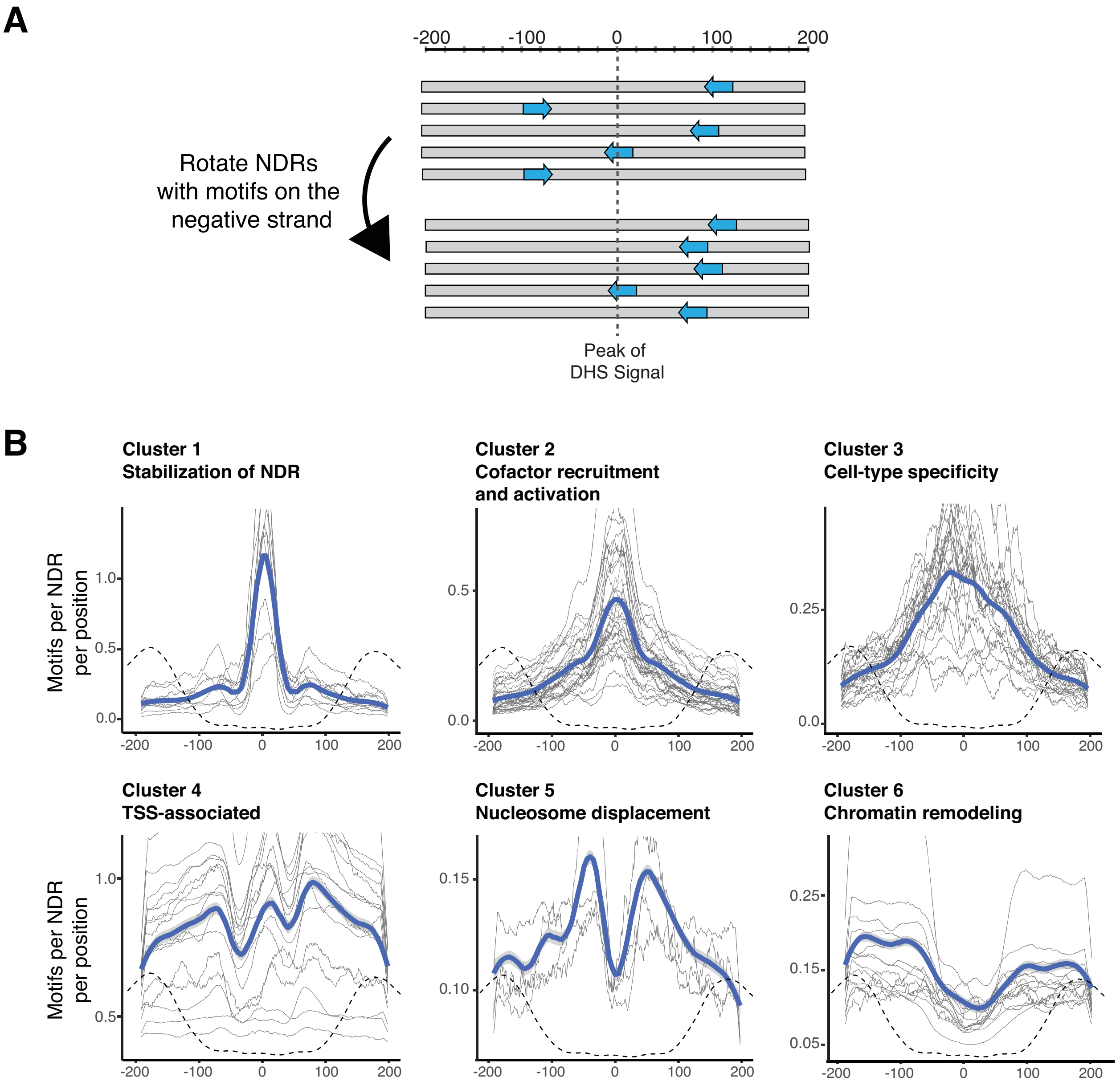
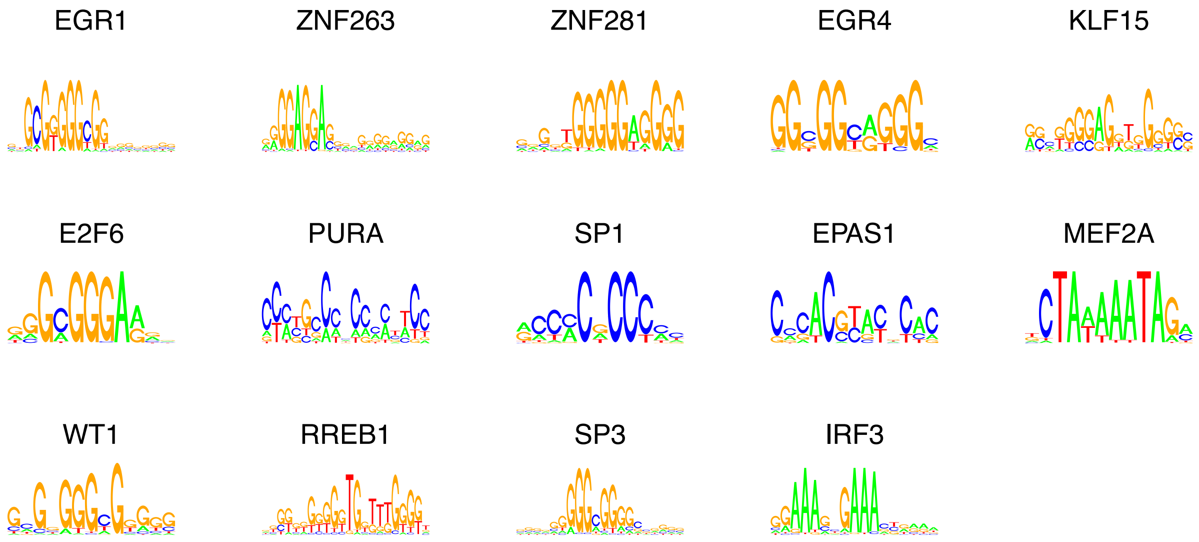


Fig. S9. Directional TF motif density for the 6 clusters. (A) Schematic of asymmetric motif analysis. For each asymmetric TF motif, we considered both the motif sequence and its reverse-complementary sequence. When we oriented the NDRs so that the sequence occurred on the ‘positive’ strand, we found that the sequences showed a clear preferred orientation relative to the center of the NDRs—that is, either the motif sequence or its reverse-complementary sequences preferentially points inward. In aggregating across the motifs in Clusters 4 and 6, we chose the motif sequence that pointed inward. Density profiles were then calculated using the aligned NDRs. Asymmetric occupancy profiles indicate that the TFs binds with a preferred orientation relative to the center of the NDR. (B) Motifs were aligned on the same strand, and motif density profiles in 400 bp regions centered around the peak of the DHS signal (gray lines) were calculated for each TF in each of the 6 clusters. Blue line depicts the overall density profile of the cluster, calculated by fitting a loess model with span=0.2. MNase-seq read density (indicating the position of the flanking nucleosomes) is shown in dotted lines for context.

A Cluster 4 motifs



B Cluster 6 motifs

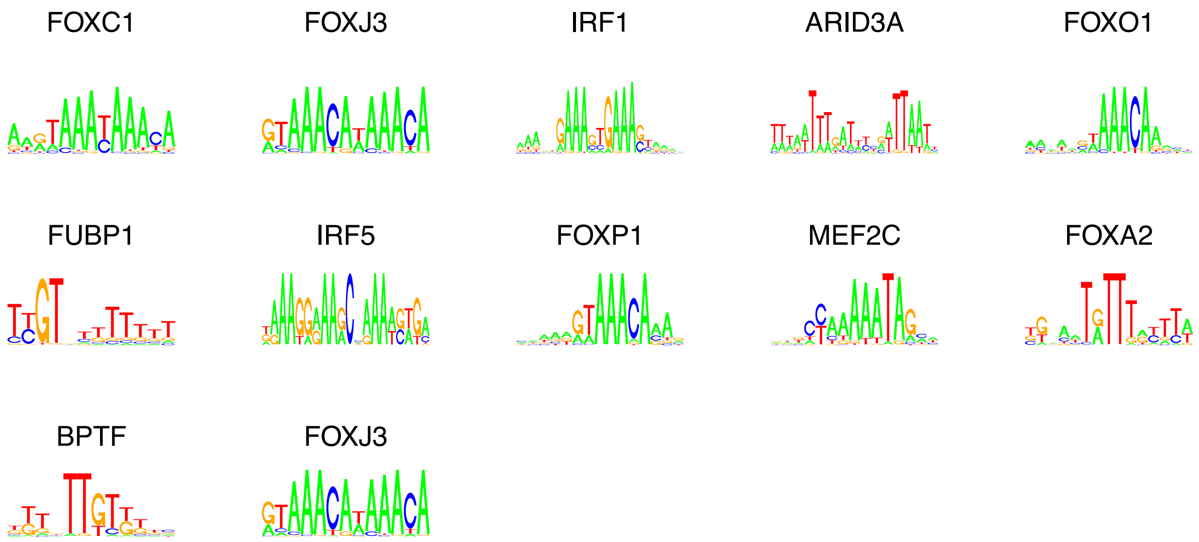


Fig. S10. Asymmetric motifs from clusters 4 and 6 that show a preferred orientation relative to the center of the NDR. Sequences of TF motifs in cluster 4 (A) and cluster 6 (B) that have asymnmetic occupancy profiles around the NDR center.

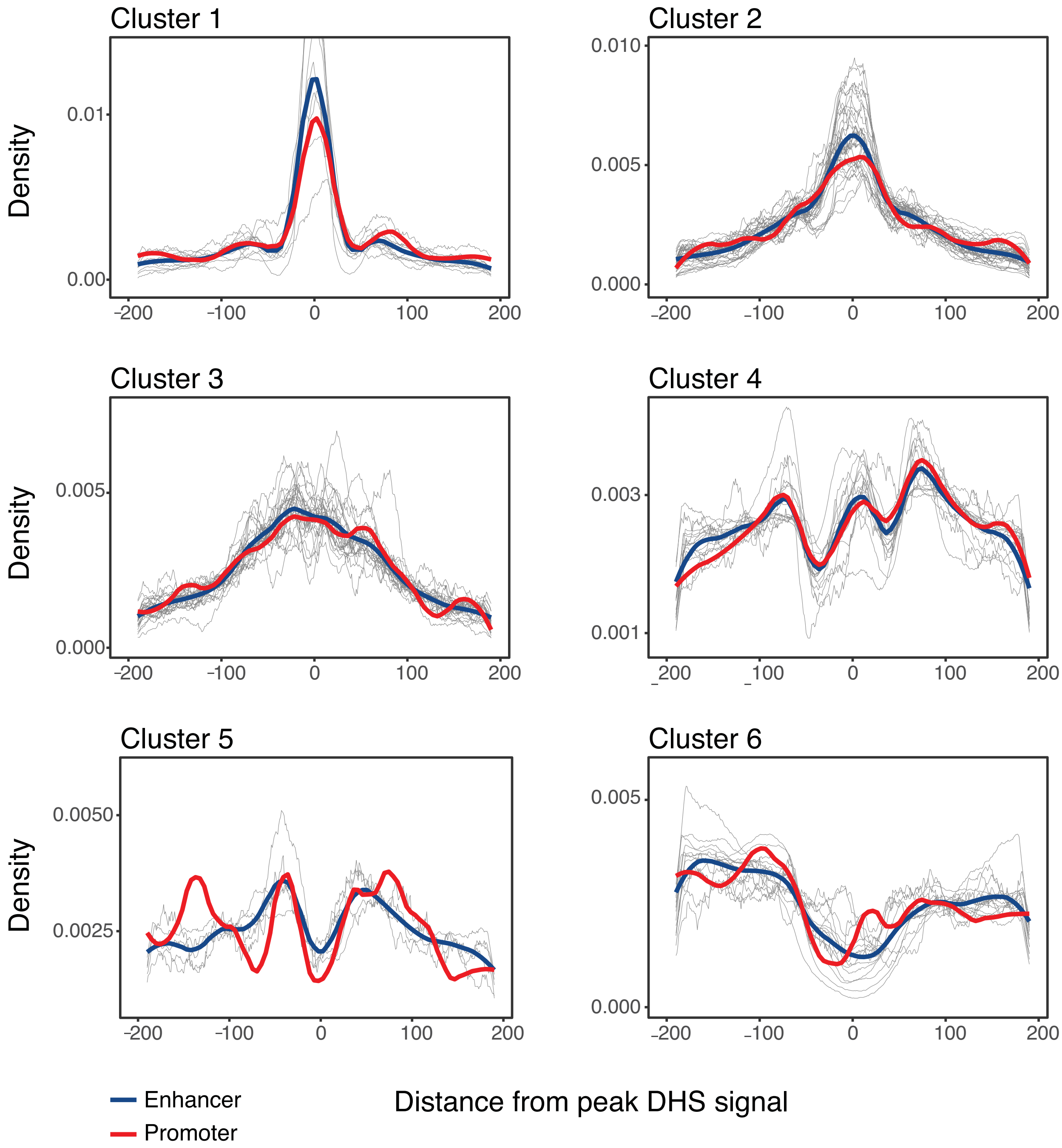


Fig. S11. Motif density profiles at promoters versus enhancers. Motif sites for each NDRs were split into promoters (<1 kb from an annotated TSS; red lines) and enhancers (>1 kb from an annotated TSS; blue lines). TF were aligned on the same strand, and the density of motif sites in 20-bp bins tiled every 1 bp in 400 bp regions centered around the peak of the DHS signal was calculated.

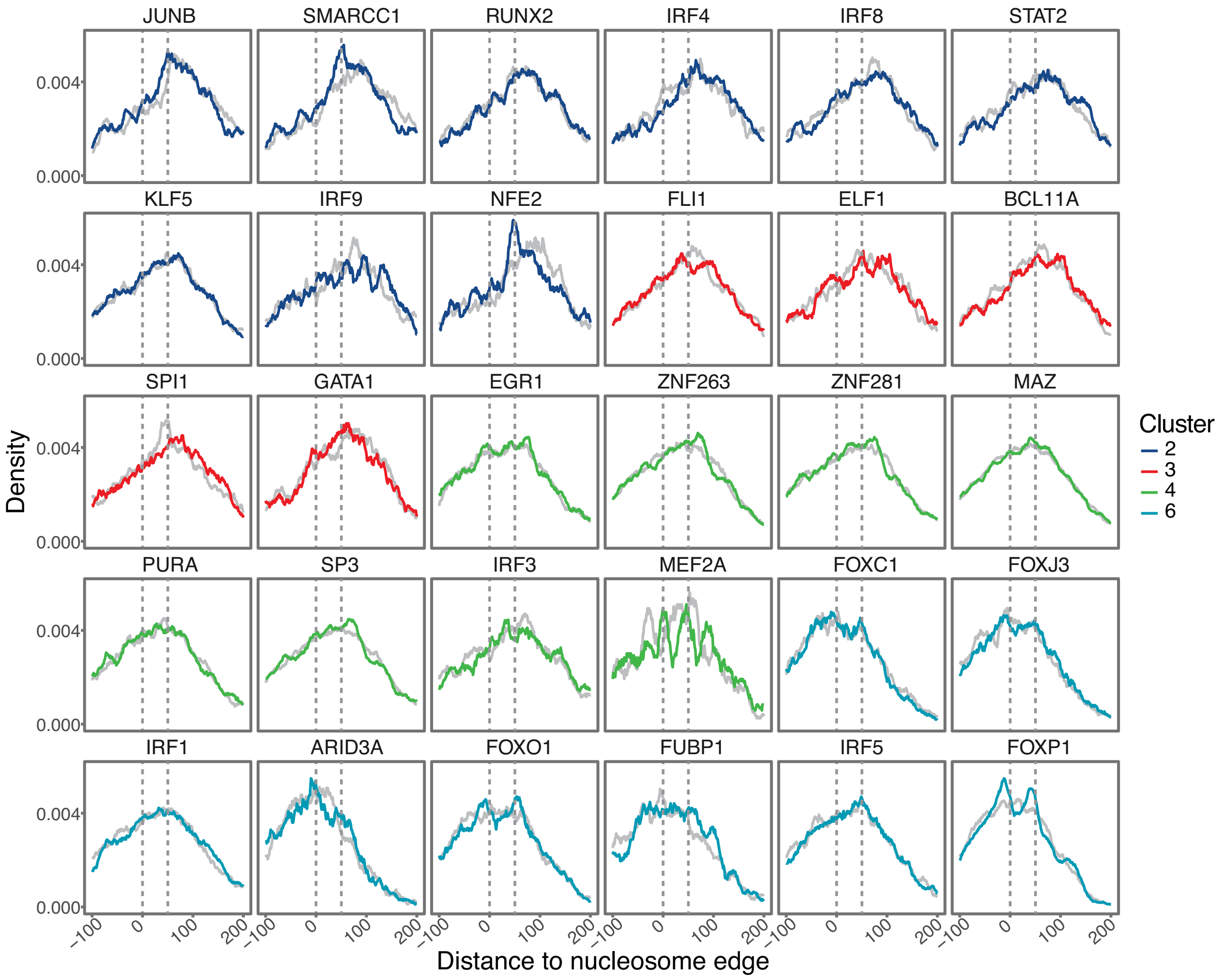


Fig. S12. Motif density relative to flanking nucleosome edge. The edges of the nucleosomes flanking each NDR were estimated from MNase-seq data (see Methods), and TF motifs were aligned relative to the nearest flanking nucleosome. Density plots represent the frequency of motif sites in 20-bp bins tiled every 1 bp in genomic NDRs (colored lines) and permuted data (gray line; permuting nucleosome position across motif sites).

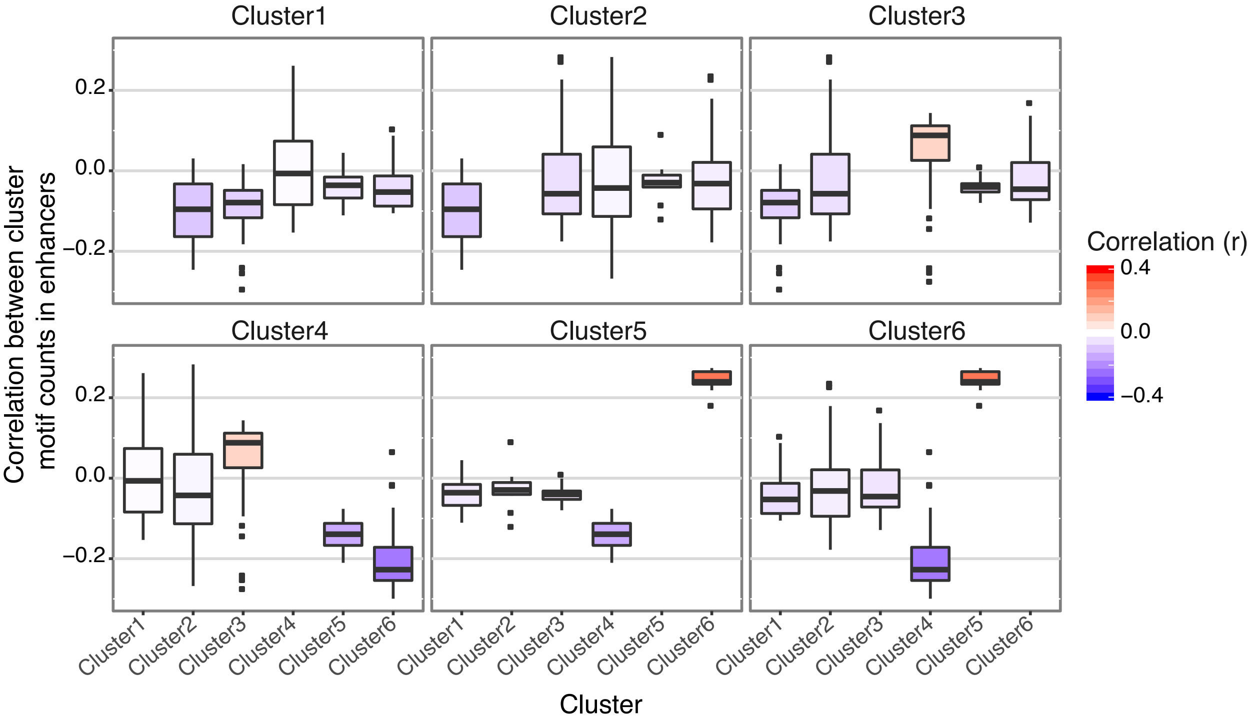


Fig. S13. Co-enrichment and -depletion of TF motif clusters is constant across cell types. For each pair of clusters, the correlation (Spearman ρ) between the number of motif sites in NDRs from cluster A and cluster B was calculated in each cell type. Positively correlated cluster pairs tend to occur together in NDRs, whereas negatively cluster pairs are co-depleted. Boxplots depict the range of ρ values for each cluster pair across the 47 cell types. The correlations of motif counts from each pair of clusters is relatively constant across cell types, indicating the clusters may represent general functional classes of TFs that interact similarly in different cellular contexts.

Table S1. TF motifs clusters identified based on the positional distribution of motif sites within NDRs

	Family	TF	Motif symmetry	Cofactor binding	AA Bias	Median genomic enrichment	Number of cell types with enrichment	Median expression [range]	Cells
Cluster 1	AP2	TFAP2A	S			2.20	16	84.8	ESDR.H1.BMP4.TROP, PLCNT.FET
	CEBP	CEBPB				3.39	50	n.d.	BLD.CD14.PC, BLD.CD14.MONO
	CTCF	CTCF				2.43	57	35.2 [30.7-39.8]	ESC.H1, ESC.H9, IPSC.DF.19.11, BLD.CD3.PPC, GI.STMC.FET, LNG.A549.ETOH002.CNCR, CLN.HCT116
		CTCFL		TBP		1.57	1	n.d.	ESC.H9
	GRHL1	GRHL1	S			4.39	2	n.d.	PLCNT.FET
	KLF	KLF7				2.66	40	n.d.	PLCNT.FET
	NFI	NFIX	S			2.30	36	54.5	MUS.TRNK.FET, MUS.LEG.FET, OVRY
	NR	DR1				2.23	57	21.3	GI.L.INT.FET
		ESRRB				3.30	1	n.d.	ADRL.GLND.FET
		HNF4A		TBP		3.84	13	132.9 [97.6-170.9]	GI.L.INT.FET, GI.S.INT.FET, GI.S.INT
Cluster 2	AP-1	FOS	S	p300, CBP, SWI/SNF, RB1		4.02	51	17.2 [15.8-214.2]	BLD.CD14.PC, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, MUS.PSOAS, GI.S.INT, MUS.HSMMT, BLD.CD14.MONO, SKIN.NHDFAD
		FOSB				4.34	29	2.4 [0.1-90.5]	ESDR.H1.MSC, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, LNG.A549.ETOH002.CNCR, CRVX.HELAS3.CNCR, BRST.HMEC, VAS.HUVEC, SKIN.NHEK, LNG.NHLF
		FOSL2	S			5.79	44	25.8 [12.7-45.2]	ESDR.H1.MSC, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, LNG.A549.ETOH002.CNCR, CRVX.HELAS3.CNCR, BRST.HMEC, MUS.HSMM, VAS.HUVEC, SKIN.NHEK, LNG.NHLF
		JDP2	S			6.15	26	n.d.	LNG.IMR90, BRN.NHA
		JUNB	S	p300, SWI/SNF		3.07	54	64.9 [2.2-329.2]	ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, BLD.CD14.PC, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.STMC.GAST, PANC, MUS.PSOAS, GI.S.INT, BLD.K562.CNCR, BLD.CD14.MONO
		BATF		KDM1A		2.58	11	0.0 [0.0-0.3]	ESDR.H1.MSC, LNG.IMR90, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, CRVX.HELAS3.CNCR, BRST.HMEC, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116
		MAF		CBP, HDAC		2.86	29	n.d.	CLN.HCT116
		NFE2		p300, CBP	Poly-pro	2.52	5	143.2	BLD.K562.CNCR
	bHLH	TFAP4	S	SWI/SNF, HDAC	Poly-pro	2.42	42	2.5	MUS.TRNK.FET, MUS.LEG.FET, MUS.HSMM, MUS.HSMMT

		TCF3		p300, CBP, HDAC, KAT2B, KDM1A, RB1		2.14	56	n.d.	MUS.TRNK.FET, MUS.LEG.FET, THYM.FET
		TCF12	S	p300, CBP		2.32	57	5.9	MUS.LEG.FET, THYM.FET, PANC
		TCF21	S			3.26	13	n.d.	MUS.TRNK.FET
		MITF				2.41	14	63.9	SKIN.PEN.FRSK.MEL.01
		MYF5	S			2.63	1	2.6 [0.0-558.1]	MUS.TRNK.FET, MUS.LEG.FET, PANC, MUS.PSOAS, MUS.HSMM, MUS.HSMMT
		TFEB	S		Poly-pro	3.75	19	4.4	SKIN.PEN.FRSK.MEL.01
	ETS	ETV2				3.76	0	0.5 [0.4-0.6]	BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, VAS.HUVEC, BLD.Jurkat
	IRF	IRF4				2.07	11	0.0 [0.0-229.4]	LNG.IMR90, SKIN.PEN.FRSK.FIB.02, BLD.GM12878, BRST.HMEC, VAS.HUVEC, BRN.NHA, SKIN.NHEK, LNG.NHLF, CLN.HCT116
		IRF8				2.74	14	104.6 [4.4-204.8]	BLD.CD14.PC, BLD.CD19.PPC, BLD.MOB.CD34.PC.F, BLD.GM12878, BLD.CD14.MONO, BLD.U937
		IRF9				3.50	46	43.7	BLD.GM12878
	Zinc finger, C2H2	KLF12				3.05	20	2.4 [1.1-6.7]	GI.S.INT.FET, GI.S.INT, CRVX.HELAS3.CNCR, SKIN.NHEK, CLN.HCT116
		KLF2		KAT2B	Poly-ala, Poly-pro	2.66	34	15.5 [0.1-31.0]	GI.S.INT, CRVX.HELAS3.CNCR
		KLF4		p300, CBP, HDAC, SWI/SNF	Poly-pro	2.20	33	18.8 [17.9-19.6]	GI.STMC.GAST, SKIN.NHEK
		KLF5		p300, CBP, YAP1, HDAC		2.23	29	65.3 [0.5-101.7]	GI.L.INT.FET, BRST.HMEC, MUS.HSMMT, BLD.K562.CNCR, SKIN.NHDFAD
		ZNF513				2.50	53	n.d.	MUS.LEG.FET, SKIN.NHDFAD
	RUNX	RUNX2		p300, RB1	Poly-ala	2.36	8	3.3 [0.0-7.8]	ESDR.H1.MSC, BLD.CD3.PPC, BLD.CD56.PC, BLD.GM12878, MUS.HSMM, MUS.HSMMT, BRN.NHA, LNG.NHLF, BLD.Jurkat
		RUNX3		p300		3.33	12	0.8	BRN.NHA, LNG.NHLF, BLD.Jurkat
	SMARCC	SMARCC1	S	SWI/SNF, KDM1A	Poly-pro, Poly-ala	3.12	57	42.1 [41.5-42.6]	ESDR.H1.BMP4.TROP, BLD.K562.CNCR
	STAT	STAT2		p300, CBP, SWI/SNF		3.46	56	35.4	BLD.CD19.PPC, BLD.CD56.PC, BLD.GM12878
	TCF	HNF1B	S			4.50	9	7.9 [6.1-9.6]	GI.S.INT.FET, GI.S.INT

		HNF1A	S	p300, CBP, KAT2B		4.13	7	8.4 [4.5-19.1]	GI.L.INT.FET, GI.STMC.GAST, PANC, LIV.HEPG2.CNCR
	TEAD	TEAD1			Poly-pro	2.34	48	45.7	LNG.NHLF
		TEAD3		YAP1		3.20	45	97.5 [34.8-160.2]	ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, IPSC.DF.6.9, BRN.NHA
		TEAD4		YAP1		2.95	42	48.6	ESC.H9, PLCNT.FET, MUS.HSMM, MUS.HSMMT
Cluster 3	bHLH	ASCL2				3.23	4	0.0 [0.0-0.0]	MUS.TRNK.FET, MUS.PSOAS, MUS.HSMM, MUS.HSMMT
		ID4				2.26	31	24.8	PANC
		NHLH1				3.08	2	0.3	MUS.HSMM, MUS.HSMMT
	BCL11A	BCL11A				2.59	10	0.0 [0.0-12.9]	BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, BLD.GM12878, VAS.HUVEC, BLD.CD14.MONO, BLD.U937
	ETS	ETS1		p300, MED23		3.41	49	31.3	BLD.MOB.CD34.PC.F
		ELF1				3.01	53	28.9 [16.2-144.6]	BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, THYM.FET, GI.STMC.GAST, GI.S.INT, BLD.GM12878, VAS.HUVEC, BLD.CD14.MONO, CLN.HCT116, BLD.Jurkat, BLD.U937, UBIQ
		ELF2				2.47	41	9.3 [4.6-14.0]	BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, THYM.FET, VAS.HUVEC, BLD.CD14.MONO, BLD.Jurkat, BLD.U937
		ELF3		p300, MED23		2.67	18	0.0	VAS.HUVEC
		ELF5				2.86	0	0.0 [0.0-0.1]	BLD.CD14.PC, BLD.MOB.CD34.PC.F, GI.STMC.GAST, VAS.HUVEC, BLD.U937
		FEV				3.63	3	0.0	VAS.HUVEC
		FLI1				2.95	15	1.9 [0.0-52.4]	ESDR.H1.BMP4.MESO, ESDR.H1.MSC, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, THYM.FET, MUS.PSOAS, BLD.GM12878, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, CLN.HCT116
		SPI1				2.40	18	23.7 [4.0-43.5]	BLD.CD56.PC, BLD.MOB.CD34.PC.F, BLD.GM12878, BLD.Jurkat
		SPIC				3.84	1	n.d.	BLD.CD14.PC, BLD.CD19.PPC, BLD.CD14.MONO
	GATA	GATA1		p300, MED17		4.54	3	125.2	BLD.K562.CNCR
		GATA3				3.22	19	3.0 [0.1-5.8]	ESDR.H1.BMP4.MESO, GI.STMC.GAST
	NR	HNF4G				2.59	7	8.4	LIV.HEPG2.CNCR
	P63	TP63	S	p300		2.98	8	68.3 [45.8-90.8]	BRST.HMEC, SKIN.NHEK
	Homeobox	POU3F1				4.69	10	15.7	ESC.H1, ESC.H9, IPSC.DF.6.9, IPSC.DF.19.11

		POU5F1				4.19	7	204.7 [183.0-226.4]	ESC.H1, ESDR.H1.BMP4.MESO, ESC.H9, IPSC.DF.6.9, IPSC.DF.19.11
		POU5F1				4.19	7	204.7 [183.0-226.4]	ESC.H1, ESDR.H1.BMP4.MESO, ESC.H9, IPSC.DF.6.9, IPSC.DF.19.11
		CDX1				2.70	5	84.4 [65.7-103.0]	GI.L.INT.FET, GI.S.INT.FET
		HOXB9				2.48	10	10.8	GI.S.INT.FET
		ZEB1		p300		2.26	41	6.5	PANC
	SOX	SOX6				2.38	5	3.6	ESDR.H1.NEUR.PROG
		SPIB				3.10	13	n.d.	BLD.U937
Cluster 4	E2F	E2F6		HDAC		1.98	27	7.3	ESDR.H1.NEUR.PROG
	EGR	EGR1		p300		2.46	50	11.2 [4.8-130.3]	ESC.H1, ESDR.H1.NEUR.PROG, ESC.H9, IPSC.DF.6.9, IPSC.DF.19.11, MUS.TRNK.FET, GI.STMC.FET, THYM.FET, OVRY, LIV.HEPG2.CNCR, BLD.K562.CNCR, UBIQ
		EGR4				1.87	2	0.0	ESC.H1, IPSC.DF.19.11, UBIQ
	EPAS2	EPAS1		p300, CBP		5.26	43	28.5 [0.1-163.0]	ESDR.H1.NEUR.PROG, LNG.IMR90, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, GI.L.INT.FET, MUS.TRNK.FET, GI.STMC.FET, BRST.HMEC, BRN.NHA, SKIN.NHDFAD, LNG.NHLF
	IRF	IRF3		p300, CBP	Poly-pro	2.22	56	32.6 [10.9-44.3]	ESDR.H1.MSC, ESDR.H1.NEUR.PROG, LNG.IMR90, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, GI.S.INT.FET, MUS.TRNK.FET, OVRY, MUS.PSOAS, BLD.GM12878, BRST.HMEC, VAS.HUVEC, BLD.CD14.MONO, BLD.U937
	KLF	KLF15		p300		2.01	25	5.5 [5.2-5.9]	ESC.H1, ESDR.H1.NEUR.PROG, ESC.H9, IPSC.DF.6.9, IPSC.DF.19.11
	MEF	MEF2A		p300, HDAC		3.77	57	28.3 [27.7-28.9]	ADRL.GLND.FET, MUS.PSOAS, BLD.GM12878, MUS.HSMMT
	PUR	PURA				2.31	46	7.1 [2.1-13.1]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, ADRL.GLND.FET, GI.S.INT.FET, MUS.LEG.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, LIV.HEPG2.CNCR, BLD.K562.CNCR, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, BLD.Jurkat, BLD.U937, UBIQ
	SP1	SP1		p300, DMT1, HDAC, CBP		2.32	57	25.9 [20.0-38.7]	ESC.H1, ESDR.H1.BMP4.MESO, IPSC.DF.6.9, IPSC.DF.19.11, GI.S.INT.FET, GI.STMC.FET, BLD.Jurkat, UBIQ

		SP3		p300, DMT1, HDAC		2.44	53	12.7 [3.6-22.0]	ESDR.H1.BMP4.TROP, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, BLD.K562.CNCR, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, BLD.U937
	Zinc finger, C2H2	WT1		DMT1	Poly-pro	2.41	2	0.5	ESDR.H1.BMP4.MESO
		MAZ			Poly-pro	2.53	57	88.5 [24.2-436.0]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, BLD.GM12878, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937
		RREB1		HDAC	Poly-pro	3.03	29	2.4	ESDR.H1.NEUR.PROG, GI.STMC.FET, THYM.FET, SKIN.NHDFAD
		ZNF263				2.48	53	10.4 [3.5-20.1]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937, UBIQ
	ZNF	ZBTB7B		p300, HDAC	Poly-pro	2.41	51	15.0 [4.3-25.0]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.NEUR.PROG, ESC.H9, IPSC.DF.6.9, IPSC.DF.19.11, GI.S.INT.FET, MUS.LEG.FET, GI.STMC.FET, OVRY, LIV.HEPG2.CNCR, SKIN.NHDFAD, UBIQ

		ZNF281			Poly-pro	2.25	43	11.0 [2.7-45.1]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, LNG.A549.ETOH002.CNCR, LIV.HEPG2.CNCR, MUS.HSMM, MUS.HSMMT, BLD.K562.CNCR, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, BLD.Jurkat, BLD.U937, UBIQ
Cluster 5	FOX	FOXA1				2.32	12	14.3	GI.L.INT.FET
		FOXD2	S			3.67	0	0.1 [0.0-0.2]	GI.STMC.GAST, PANC, LNG.A549.ETOH002.CNCR, LIV.HEPG2.CNCR
		FOXL2				2.39	1	2.1 [0.0-71.0]	ESDR.H1.MSC, LNG.IMR90, SKIN.PEN.FRSK.FIB.02, PLCNT.FET, GI.STMC.FET, OVRY, CRVX.HELAS3.CNCR
		FOXM1				1.91	1	9.1	LIV.HEPG2.CNCR
		FOXN3				1.77	50	19.6 [3.0-36.3]	LNG.A549.ETOH002.CNCR, LIV.HEPG2.CNCR
		FOXO3				2.05	54	14.0 [4.9-42.0]	ESDR.H1.BMP4.TROP, ESDR.H1.MSC, LNG.IMR90, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, GI.STMC.GAST, OVRY, PANC, LNG.A549.ETOH002.CNCR, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR
Cluster 6	ARID	ARID3A			Poly-ala	2.62	27	7.6 [1.4-284.7]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, BLD.GM12878, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat
	BPTF	BPTF				1.85	57	14.7 [6.7-24.7]	ESDR.H1.BMP4.TROP, ESDR.H1.MSC, BLD.CD19.PPC, BLD.CD3.PPC, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, PLCNT.FET, OVRY, LNG.A549.ETOH002.CNCR, CRVX.HELAS3.CNCR, BRST.HMEC, CLN.HCT116
	FOX	FOXA2		SATB2		2.07	12	0.0 [0.0-0.0]	ESDR.H1.MSC, LNG.IMR90, SKIN.PEN.FRSK.FIB.01, CRVX.HELAS3.CNCR

		FOXC1		SATB2	Poly-ala	3.00	14	1.3 [0.0-31.1]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, BLD.GM12878, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937
		FOXJ3				2.98	57	16.6 [9.1-27.0]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, BLD.GM12878, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937
		FOXO1		SATB2	Poly-ala	2.17	42	7.2 [0.8-61.8]	ESC.H1, ESDR.H1.BMP4.TROP, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, ADRL.GLND.FET, GI.L.INT.FET, MUS.TRNK.FET, PLCNT.FET, THYM.FET, OVRY, MUS.PSOAS, GI.S.INT, BLD.GM12878, BRST.HMEC, MUS.HSMM, MUS.HSMMT, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937

		FOXP1		SATB2		6.01	35	6.7 [0.0-15.7]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, BLD.GM12878, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937, UBIQ
	FUBP	FUBP1				2.11	57	33.8 [8.9-100.6]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, BLD.GM12878, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937
	IRF	IRF1				2.60	22	3.3 [0.3-27.4]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, ESC.H9, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, PANC, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, BLD.GM12878, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, MUS.HSMMT, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937

		IRF5				2.08	10	0.4 [0.0-26.7]	ESC.H1, ESDR.H1.BMP4.MESO, ESDR.H1.BMP4.TROP, ESDR.H1.MSC, ESDR.H1.NEUR.PROG, LNG.IMR90, IPSC.DF.6.9, IPSC.DF.19.11, BLD.CD14.PC, BLD.CD19.PPC, BLD.CD3.PPC, BLD.CD56.PC, BLD.MOB.CD34.PC.F, SKIN.PEN.FRSK.FIB.01, SKIN.PEN.FRSK.FIB.02, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, GI.L.INT.FET, GI.S.INT.FET, MUS.TRNK.FET, MUS.LEG.FET, PLCNT.FET, GI.STMC.FET, THYM.FET, GI.STMC.GAST, OVRY, MUS.PSOAS, GI.S.INT, LNG.A549.ETOH002.CNCR, BLD.GM12878, CRVX.HELAS3.CNCR, LIV.HEPG2.CNCR, BRST.HMEC, MUS.HSMM, VAS.HUVEC, BLD.K562.CNCR, BLD.CD14.MONO, BRN.NHA, SKIN.NHDFAD, SKIN.NHEK, LNG.NHLF, CLN.HCT116, BLD.Jurkat, BLD.U937
	MEF	MEF2C				2.01	21	61.1	MUS.HSMM
	SRY	SRY				2.21	1	0.5 [0.0-1.1]	ESDR.H1.BMP4.MESO, ESC.H9, IPSC.DF.6.9, SKIN.PEN.FRSK.MEL.01
	ZNF	ZNF148				2.43	43	2.3	SKIN.PEN.FRSK.MEL.01
		ZNF202				0.44	33	4.0 [3.7-12.9]	PLCNT.FET, GI.STMC.GAST, PANC, GI.S.INT, CLN.HCT116, UBIQ
		ZNF35				1.64	25	4.8 [1.4-8.7]	ESDR.H1.MSC, SKIN.PEN.FRSK.MEL.01, ADRL.GLND.FET, PLCNT.FET, LNG.A549.ETOH002.CNCR, CRVX.HELAS3.CNCR

Table S2. Enrichment of functional properties in TF motif clusters

	Term	1	2	3	4	5	6
BIOCARTA	h_tertPathway:Overview of telomerase protein component gene hTert Transcriptional Regulation	0.0	0.0	2.4	0.0	0.0	0.0
BIOGRID_INTERACTION	2033:EP300~E1A binding protein p300	0.0	3.1	5.6	0.8	7.2	0.0
	1387:CREBBP~CREB binding protein	0.0	0.0	0.1	0.0	7.1	0.0
	2353:FOS~Fos proto-oncogene, AP-1 transcription factor subunit	0.0	0.0	0.0	0.0	4.8	0.0
	3239:HOXD13~homeobox D13	0.0	3.3	0.0	0.0	0.0	0.0
	23314:SATB2~SATB homeobox 2	0.0	2.9	0.0	0.0	0.0	0.0
	6667:SP1~Sp1 transcription factor	1.8	0.0	2.8	0.0	0.0	0.0
	468:ATF4~activating transcription factor 4	0.0	0.0	0.0	0.0	1.8	0.0
	1386:ATF2~activating transcription factor 2	0.0	0.0	0.0	0.0	1.7	0.0
	6929:TCF3~transcription factor 3	0.0	0.0	0.0	0.0	1.7	0.0
	3725:JUN~Jun proto-oncogene, AP-1 transcription factor subunit	0.0	0.0	1.3	0.0	1.7	0.0
	57154:SMURF1~SMAD specific E3 ubiquitin protein ligase 1	0.0	0.0	0.0	0.0	1.6	0.0
	4088:SMAD3~SMAD family member 3	0.0	0.0	0.0	0.0	1.6	0.2
	10413:YAP1~Yes associated protein 1	0.0	0.0	0.0	0.0	1.5	0.0
	10014:HDAC5~histone deacetylase 5	0.0	0.0	0.0	0.0	1.5	0.0
	1786:DNMT1~DNA methyltransferase 1	0.0	0.0	1.5	0.0	0.0	0.0
	7392:USF2~upstream transcription factor 2, c-fos interacting	0.0	0.0	0.0	0.0	1.5	0.0
	11016:ATF7~activating transcription factor 7	0.0	0.0	0.0	0.0	1.5	0.0
	6605:SMARCE1~SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1	0.0	0.0	0.0	0.0	1.5	0.0
	26145:IRF2BP1~interferon regulatory factor 2 binding protein 1	0.0	0.0	0.0	0.0	1.4	0.0
	10538:BATF~basic leucine zipper ATF-like transcription factor	0.0	0.0	0.0	0.0	1.4	0.0
	25937:WWTR1~WW domain containing transcription regulator 1	0.0	0.0	0.0	0.0	1.4	0.0
	6304:SATB1~SATB homeobox 1	0.0	1.3	0.0	0.0	0.0	0.0
COG	Transcription / Cell division and chromosome partitioning	0.0	0.0	3.2	0.0	0.0	0.0
	Transcription	0.0	2.5	0.0	0.0	0.0	0.0
GOTERM_BP	GO:0006357~regulation of transcription from RNA polymerase II promoter	0.1	1.8	0.0	5.5	19.7	0.1
	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	5.1	4.1	2.4	14.9	16.4	1.4
	GO:0006366~transcription from RNA polymerase II promoter	3.8	3.6	4.7	10.7	15.6	1.6

	GO:0045893~positive regulation of transcription, DNA-templated	2.5	3.8	2.6	5.1	9.8	1.9
	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	1.0	2.0	2.9	4.1	6.8	1.7
	GO:0030154~cell differentiation	0.0	0.0	0.0	4.2	0.0	0.1
	GO:0006355~regulation of transcription, DNA-templated	0.0	0.3	4.0	0.0	0.4	0.0
	GO:0006351~transcription, DNA-templated	1.3	3.5	1.7	2.7	3.0	0.1
	GO:0045892~negative regulation of transcription, DNA-templated	0.6	0.6	1.7	3.0	0.1	0.8
	GO:0035019~somatic stem cell population maintenance	0.0	0.0	0.0	2.6	0.0	0.0
	GO:0001824~blastocyst development	0.0	0.0	0.0	2.2	0.4	0.0
	GO:0009887~organ morphogenesis	0.0	0.0	0.0	2.2	0.0	0.0
	GO:1901653~cellular response to peptide	0.0	0.0	0.0	0.0	2.1	0.0
	GO:0060337~type I interferon signaling pathway	0.0	0.2	0.3	0.0	2.1	0.0
	GO:0042832~defense response to protozoan	0.0	0.0	0.0	0.0	1.7	0.0
	GO:0030218~erythrocyte differentiation	0.0	0.0	0.0	1.5	0.0	0.0
	GO:0035329~hippo signaling	0.0	0.0	0.0	0.0	1.5	0.0
	GO:0045595~regulation of cell differentiation	0.0	0.0	0.0	0.0	1.4	0.0
GOTERM_CC	GO:0005634~nucleus	3.6	4.8	6.6	8.2	6.1	1.7
	GO:0005654~nucleoplasm	1.4	5.3	4.0	2.7	6.7	0.0
	GO:0005667~transcription factor complex	0.0	0.0	1.1	6.4	3.9	0.0
	GO:0000790~nuclear chromatin	0.0	0.0	0.4	2.3	5.1	0.0
	GO:0090575~RNA polymerase II transcription factor complex	0.0	0.0	0.0	0.0	2.1	0.0
	GO:0016607~nuclear speck	0.0	0.3	1.3	0.0	0.0	0.0
GOTERM_MF	GO:0003700~transcription factor activity, sequence-specific DNA binding	4.3	7.9	9.0	11.1	23.8	5.2
	GO:0000978~RNA polymerase II core promoter proximal region sequence-specific DNA binding	3.7	2.8	2.7	4.4	17.2	0.0
	GO:0001077~transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	4.3	0.4	4.6	5.3	16.5	0.6
	GO:0043565~sequence-specific DNA binding	5.6	7.4	4.5	14.5	7.9	6.3
	GO:0000981~RNA polymerase II transcription factor activity, sequence-specific DNA binding	0.0	7.4	1.5	9.5	1.9	4.2
	GO:0000982~transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding	0.0	0.0	0.0	0.0	8.1	0.0
	GO:0003677~DNA binding	2.3	2.8	2.7	2.5	5.4	1.4
	GO:0070888~E-box binding	0.0	0.0	0.0	2.4	5.4	0.0
	GO:0044212~transcription regulatory region DNA binding	4.4	1.3	2.3	2.1	4.8	0.0
	GO:0035035~histone acetyltransferase binding	0.0	0.0	4.6	0.0	0.0	0.0
	GO:0046872~metal ion binding	0.0	0.3	4.5	0.0	0.1	0.0
	GO:0043425~bHLH transcription factor binding	0.0	0.0	0.0	0.0	4.2	0.0
	GO:0008134~transcription factor binding	1.4	2.0	1.1	1.8	4.2	0.0

	GO:0001078~transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding	0.8	0.6	1.8	4.0	1.3	0.0
	GO:0001085~RNA polymerase II transcription factor binding	0.0	0.0	0.0	3.7	0.7	0.0
	GO:0005515~protein binding	0.3	0.5	1.7	0.4	3.6	0.0
	GO:0046982~protein heterodimerization activity	0.4	0.0	0.2	0.0	3.1	0.0
	GO:0000975~regulatory region DNA binding	0.0	1.3	0.0	0.0	3.0	0.0
	GO:0001228~transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	0.8	0.6	0.0	2.8	0.5	0.0
	GO:0001158~enhancer sequence-specific DNA binding	0.0	0.0	0.0	2.8	0.0	0.0
	GO:0000979~RNA polymerase II core promoter sequence-specific DNA binding	2.6	0.0	2.2	2.1	1.8	0.0
	GO:0003705~transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding	2.5	2.0	0.8	0.0	1.7	0.0
	GO:0046983~protein dimerization activity	0.0	0.0	0.0	2.4	2.0	0.0
	GO:0003713~transcription coactivator activity	1.5	0.0	0.0	1.0	2.4	0.0
	GO:0000977~RNA polymerase II regulatory region sequence-specific DNA binding	0.0	1.3	2.3	2.0	0.9	0.0
	GO:0003676~nucleic acid binding	0.2	0.1	2.2	0.0	0.1	0.0
	GO:0000976~transcription regulatory region sequence-specific DNA binding	0.0	0.0	0.0	2.1	0.0	0.0
	GO:0000980~RNA polymerase II distal enhancer sequence-specific DNA binding	0.0	0.0	0.0	2.0	0.0	0.0
	GO:0044729~hemi-methylated DNA-binding	0.0	0.0	1.9	0.0	0.0	0.0
	GO:0010385~double-stranded methylated DNA binding	0.0	0.0	1.8	0.0	0.0	0.0
	GO:0003682~chromatin binding	0.4	1.7	0.3	1.5	1.8	0.0
	GO:0042826~histone deacetylase binding	0.0	0.0	0.7	0.0	1.4	0.0
INTACT	11016:activating transcription factor 7(ATF7)	0.0	0.0	0.0	0.0	4.0	0.0
	3726:JunB proto-oncogene, AP-1 transcription factor subunit(JUNB)	0.0	0.0	0.0	0.0	3.9	0.0
	1649:DNA damage inducible transcript 3(DDIT3)	0.0	0.0	0.0	0.0	3.8	0.0
	468:activating transcription factor 4(ATF4)	0.0	0.0	0.0	0.0	3.4	0.0
	23314:SATB homeobox 2(SATB2)	0.0	3.3	0.0	0.0	0.0	0.0
	3727:JunD proto-oncogene, AP-1 transcription factor subunit(JUND)	0.0	0.0	0.3	0.0	3.0	0.0
	9935:MAF bZIP transcription factor B(MAFB)	0.0	0.0	0.0	0.0	1.8	0.0
	3229:homeobox C13(HOXC13)	0.0	1.7	0.0	0.0	0.0	0.0
	4779:nuclear factor, erythroid 2 like 1(NFE2L1)	0.0	0.0	0.0	0.0	1.6	0.0
	6304:SATB homeobox 1(SATB1)	0.0	1.6	0.0	0.0	0.0	0.0
	6886:TAL bHLH transcription factor 1, erythroid differentiation factor(TAL1)	0.0	0.0	0.0	1.6	0.2	0.0
	1386:activating transcription factor 2(ATF2)	0.0	0.0	0.0	0.0	1.5	0.0

INTERPRO	IPR000418:Ets domain	0.0	0.0	0.0	16.2	0.0	0.0
	IPR001766:Transcription factor, fork head	0.0	6.1	0.0	0.0	0.0	12.3
	IPR018122:Transcription factor, fork head, conserved site	0.0	2.5	0.0	0.0	0.0	9.9
	IPR004827:Basic-leucine zipper domain	0.0	0.0	0.0	0.0	9.8	0.0
	IPR011991:Winged helix-turn-helix DNA-binding domain	0.0	6.4	0.1	8.7	2.4	9.3
	IPR015880:Zinc finger, C2H2-like	0.7	0.3	9.2	0.1	0.9	0.0
	IPR013087:Zinc finger C2H2-type/integrase DNA-binding domain	0.7	0.3	9.2	0.1	1.0	0.0
	IPR007087:Zinc finger, C2H2	0.8	0.3	9.2	0.1	0.9	0.0
	IPR000837:Fos transforming protein	0.0	0.0	0.0	0.0	7.6	0.0
	IPR011598:Myc-type, basic helix-loop-helix (bHLH) domain	0.0	0.0	0.0	1.5	6.2	0.0
	IPR003118:Pointed domain	0.0	0.0	0.0	5.6	0.0	0.0
	IPR013761:Sterile alpha motif/pointed domain	0.0	0.0	0.0	4.1	0.0	0.0
	IPR016361:Transcriptional enhancer factor	0.0	0.0	0.0	0.0	3.6	0.0
	IPR000818:TEA/ATTS	0.0	0.0	0.0	0.0	3.6	0.0
	IPR019471:Interferon regulatory factor-3	0.0	0.0	0.0	0.0	3.2	0.0
	IPR001346:Interferon regulatory factor DNA-binding domain	0.0	1.3	0.0	0.0	3.0	0.0
	IPR019817:Interferon regulatory factor, conserved site	0.0	1.3	0.0	0.0	3.0	0.0
	IPR001356:Homeodomain	0.0	0.0	0.0	2.9	0.0	0.0
	IPR008917:Eukaryotic transcription factor, Skn-1-like, DNA-binding	0.0	0.0	0.0	0.0	2.6	0.0
	IPR017855:SMAD domain-like	0.0	0.0	0.0	0.0	2.6	0.0
	IPR009057:Homeodomain-like	0.0	0.0	0.0	2.5	0.1	0.0
	IPR017970:Homeobox, conserved site	0.0	0.0	0.0	2.2	0.0	0.0
	IPR013088:Zinc finger, NHR/GATA-type	0.8	0.0	0.0	2.0	0.0	0.0
	IPR008967:p53-like transcription factor, DNA-binding	0.0	0.0	0.0	0.0	1.8	0.0
	IPR022084:Transcription factor Elf, N-terminal	0.0	0.0	0.0	1.8	0.0	0.0
	IPR008984:SMAD/FHA domain	0.0	0.0	0.0	0.0	1.7	0.0
	IPR013711:Runx, C-terminal domain	0.0	0.0	0.0	0.0	1.6	0.0
	IPR013524:Runt domain	0.0	0.0	0.0	0.0	1.6	0.0
	IPR000040:Acute myeloid leukemia 1 protein (AML1)/Runt	0.0	0.0	0.0	0.0	1.6	0.0
	IPR027384:Runx, central domain	0.0	0.0	0.0	0.0	1.6	0.0
	IPR021802:Basic helix-loop-helix leucine zipper transcrition factor MiT/TFE	0.0	0.0	0.0	0.0	1.5	0.0
	IPR012346:p53/RUNT-type transcription factor, DNA-binding domain	0.0	0.0	0.0	0.0	1.4	0.0

KEGG_PATHWAY	hsa04380:Osteoclast differentiation	0.0	0.0	0.0	0.0	4.6	0.0
PFAM	PF00178:Ets-domain	0.0	0.0	0.0	16.3	0.0	0.0
	PF00250:Fork head domain	0.0	6.1	0.0	0.0	0.0	12.4
	PF00170:bZIP transcription factor	0.0	0.0	0.0	0.0	7.6	0.0
	PF00010:Helix-loop-helix DNA-binding domain	0.0	0.0	0.0	1.6	6.4	0.0
	PF02198:Sterile alpha motif (SAM)/Pointed domain	0.0	0.0	0.0	5.7	0.0	0.0
	PF00096:Zinc finger, C2H2 type	0.0	0.0	4.6	0.0	0.8	0.0
	PF01285:TEA/ATTS domain family	0.0	0.0	0.0	0.0	3.8	0.0
	PF10401:Interferon-regulatory factor 3	0.0	0.0	0.0	0.0	3.4	0.0
	PF00605:Interferon regulatory factor transcription factor	0.0	1.2	0.0	0.0	3.2	0.0
	PF00046:Homeobox domain	0.0	0.0	0.0	3.1	0.0	0.0
	PF12310:Transcription factor protein N terminal	0.0	0.0	0.0	1.8	0.0	0.0
	PF00853:Runt domain	0.0	0.0	0.0	0.0	1.6	0.0
	PF08504:Runx inhibition domain	0.0	0.0	0.0	0.0	1.6	0.0
	PF11851:Domain of unknown function (DUF3371)	0.0	0.0	0.0	0.0	1.6	0.0
	PF00320:GATA zinc finger	0.0	0.0	0.0	1.4	0.0	0.0
	PF00157:Pou domain - N-terminal to homeobox domain	0.0	0.0	0.0	1.3	0.0	0.0
PIR_SUPERFAMILY	PIRSF002603:transcriptional enhancer factor	0.0	0.0	0.0	0.0	4.3	0.0
SMART	SM00413:ETS	0.0	0.0	0.0	14.7	0.0	0.0
	SM00339:FH	0.0	5.2	0.0	0.0	0.0	11.7
	SM00338:BRLZ	0.0	0.0	0.0	0.0	9.0	0.0
	SM00355:ZnF_C2H2	0.9	0.1	6.8	0.0	0.5	0.0
	SM00353:HLH	0.0	0.0	0.0	1.3	5.4	0.0
	SM00251:SAM_PNT	0.0	0.0	0.0	5.2	0.0	0.0
	SM00426:TEA	0.0	0.0	0.0	0.0	3.7	0.0
	SM01243:SM01243	0.0	0.0	0.0	0.0	3.3	0.0
	SM00348:IRF	0.0	1.1	0.0	0.0	3.1	0.0
	SM00389:HOX	0.0	0.0	0.0	2.3	0.0	0.0
	SM00401:ZnF_GATA	0.0	0.0	0.0	1.3	0.0	0.0
UP_KEYWORDS	DNA-binding	7.5	10.8	13.4	15.7	27.0	3.6
	Transcription regulation	7.3	11.7	12.9	16.1	21.7	3.6
	Transcription	7.3	11.8	12.9	16.1	21.5	3.7
	Nucleus	4.5	7.3	7.9	11.5	17.3	2.2

	Activator	4.5	2.4	7.3	8.6	16.3	1.3
	Zinc-finger	1.5	1.0	6.5	0.6	0.2	0.0
	Ubl conjugation	1.0	3.4	3.5	1.7	6.4	0.7
	Zinc	1.2	0.7	5.4	0.7	0.1	0.0
	Repressor	1.0	2.4	4.9	1.9	0.7	0.4
	Isopeptide bond	1.3	2.5	3.4	2.4	4.6	0.3
	Metal-binding	0.8	0.4	3.8	0.3	0.0	0.0
	Developmental protein	0.0	0.2	0.5	3.7	0.1	0.0
	Phosphoprotein	0.6	2.5	1.0	0.4	3.1	1.4
	Homeobox	0.0	0.0	0.0	3.1	0.0	0.0
	Differentiation	0.0	0.7	1.4	1.8	0.6	0.0
	Coiled coil	0.0	0.0	0.0	0.0	1.7	0.0
UP_SEQ_FEATURE	DNA-binding region:Basic motif	0.0	0.0	0.0	0.2	19.5	0.0
	DNA-binding region:ETS	0.0	0.0	0.0	16.3	0.0	0.0
	domain:Leucine-zipper	0.0	0.0	0.0	0.0	15.6	0.0
	DNA-binding region:Fork-head	0.0	5.5	0.0	0.0	0.0	11.9
	zinc finger region:C2H2-type 1	0.7	0.2	10.3	0.1	1.1	0.0
	zinc finger region:C2H2-type 3	0.8	0.3	10.1	0.0	1.0	0.0
	zinc finger region:C2H2-type 2	0.7	0.2	10.1	0.0	1.0	0.0
	domain:Helix-loop-helix motif	0.0	0.0	0.0	1.2	6.3	0.0
	domain:PNT	0.0	0.0	0.0	5.3	0.0	0.0
	cross-link:Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in SUMO)	1.6	0.3	1.3	4.1	0.2	0.0
	DNA-binding region:TEA	0.0	0.0	0.0	0.0	3.6	0.0
	DNA-binding region:Tryptophan pentad repeat	0.0	0.8	0.0	0.0	2.9	0.0
	DNA-binding region:Homeobox	0.0	0.0	0.0	1.8	0.0	0.0

Table S3. Co-enrichment and -depletion of motifs from different clusters

Cluster Pair	Odds Ratio (Genomic)	Odds Ratio (Shuffled)	p-value (Genomic)	p-value (Shuffled)
1_2	0.36 [0.36-0.37]	0.54 [0.53-0.55]	0.0E+00	0.0E+00
1_3	0.76 [0.75-0.77]	0.77 [0.75-0.78]	0.0E+00	9.2E-193
1_4	1.29 [1.26-1.32]	1.74 [1.70-1.77]	4.9E-98	0.0E+00
1_5	1.51 [1.49-1.54]	1.43 [1.40-1.46]	0.0E+00	1.0E-217
1_6	0.81 [0.79-0.83]	0.69 [0.68-0.71]	1.0E-61	0.0E+00
2_3	1.04 [1.03-1.06]	1.08 [1.06-1.09]	8.9E-08	5.3E-26
2_4	0.92 [0.90-0.94]	1.06 [1.05-1.08]	4.5E-12	9.0E-16
2_5	0.56 [0.55-0.57]	0.71 [0.70-0.72]	0.0E+00	2.9E-261
2_6	1.33 [1.30-1.36]	1.08 [1.06-1.10]	2.6E-114	3.0E-26
3_4	1.28 [1.26-1.31]	1.16 [1.14-1.18]	1.9E-112	7.3E-75
3_5	0.64 [0.64-0.65]	0.83 [0.81-0.84]	0.0E+00	8.8E-73
3_6	1.02 [1.00-1.05]	1.12 [1.10-1.13]	4.3E-02	2.0E-45
4_5	0.93 [0.91-0.95]	0.81 [0.79-0.83]	2.4E-09	1.2E-86
4_6	0.44 [0.41-0.46]	0.47 [0.46-0.48]	1.9E-275	0.0E+00
5_6	1.84 [1.79-1.89]	1.72 [1.68-1.75]	0.0E+00	0.0E+00