

Supplemental Information for

Decoding Gene Responsiveness to Synthetic Chromatin Reader-Actuators with Multi-Modal Epigenomic Profiling

Seong Hu Kim¹, Isioma Enwerem-Lackland¹, Natecia L. Williams¹, Rachel Fisher², Christopher Plaisier^{2,3}, Karmella A. Haynes¹

1. Wallace H. Coulter Department of Biomedical Engineering, Emory University, Atlanta, GA
2. School of Biological and Health Systems Engineering, Ira A. Fulton School of Engineering, Arizona State University, Tempe, AZ
3. John Schufeldt School of Medicine and Medical Engineering, Arizona State University, Tempe, AZ

Supplemental Figures

1. Figure S1. ChIP-qPCR validation of SRA and PCD-RFP occupancy at the HOXB9 locus
2. Figure S2. ChIP-Atlas fold enrichment analysis of histone PTM at fusion protein ChIP-seq peaks
3. Figure S3. Comparison of histone PTM overlaps with PCD-RFP and SRA ChIP-seq peaks with ChromHMM regulatory states
4. Figure S4. XGBoost prediction of PCD-fusion genome-wide binding, extended data
5. Figure S5. XGBoost prediction of endogenous CBX8 genome-wide binding
6. Figure S6. Clustering and principal component analysis of RNA-seq samples across experimental conditions
7. Figure S7. Differential histone PTM overlap at downstream proximal promoter (DSPP) loci in MCF7 and SK-N-SH cells

Supplemental Tables

1. Table S1. Transcription factor target genes within SRA UpDEG classes
2. Table S2. RT-qPCR Primer Pairs
3. Table S3. ENCODE MCF7 Datasets
4. Table S4. ENCODE SK-N-SH Datasets

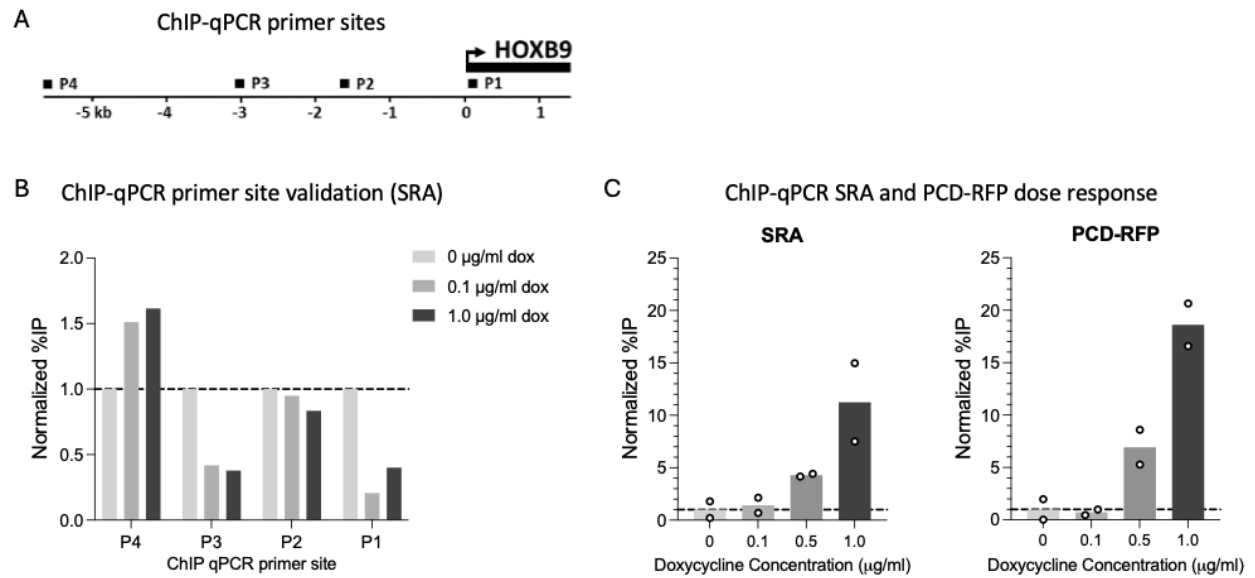


Figure S1. ChIP-qPCR validation of SRA and PCD-RFP occupancy at the HOXB9 locus. (A) Map of primer target sites for the qPCR assay. (B) The highest levels of enrichment for SRA were observed for site P4 in an optimization experiment using single replicates of ChIP DNA samples from SRA-expressing cells treated with 0, 0.1, or 1.0 µg/mL of inducer (doxycycline, dox). (C) New clonal isolates of SRA and PCD-RFP expressing cell lines were selected for strong activation of the fusion proteins, and dose-dependent ChIP enrichment was measured. At primer target site P4, enrichment increases with doxycycline concentration for both constructs, consistent with inducible binding observed in ChIP-seq. Bars represent the mean of two ChIP DNA replicates (dots).

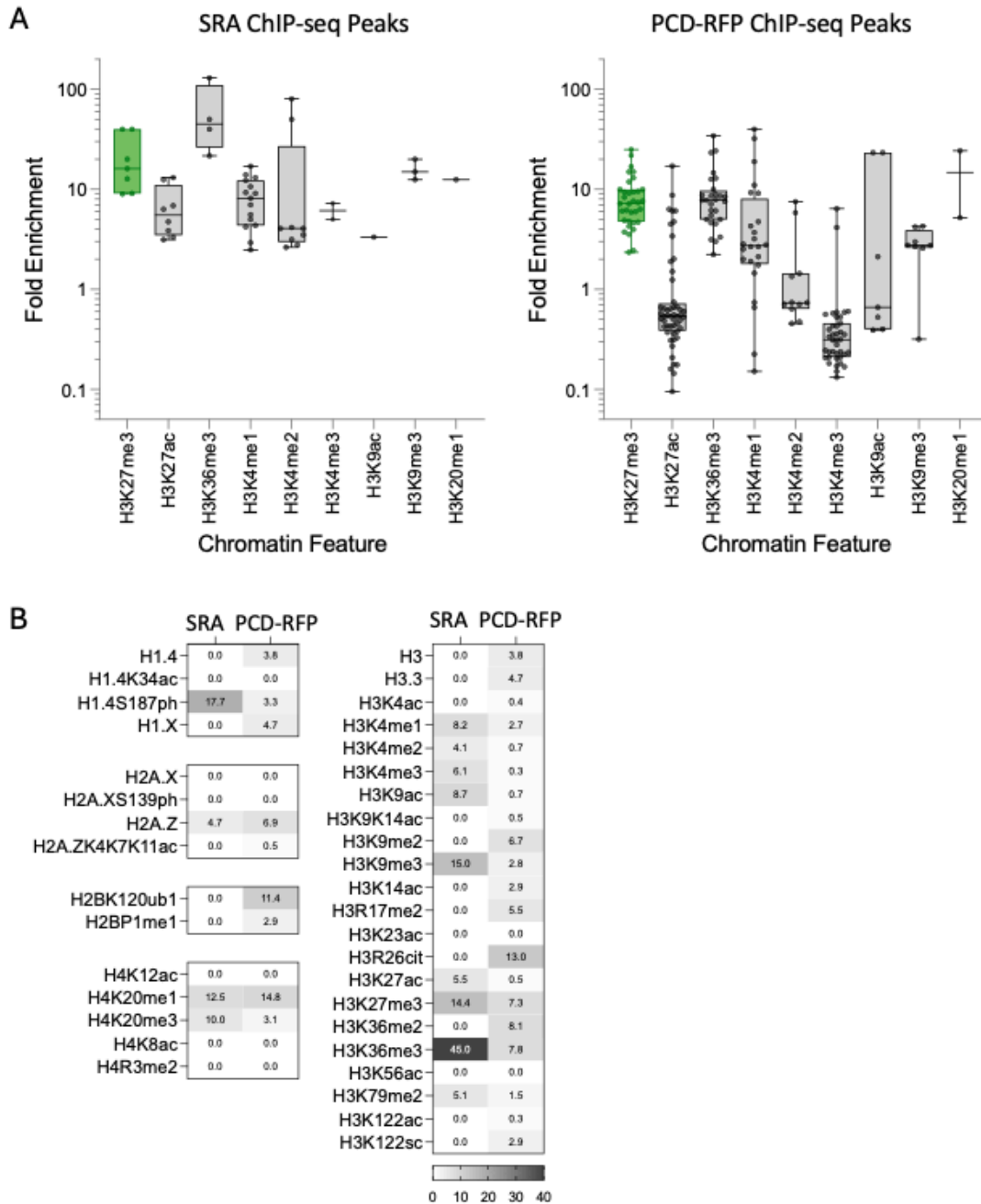


Figure S2. ChIP-Atlas fold enrichment analysis of histone PTM at fusion protein ChIP-seq peaks.

(A) The ChIP-Atlas Enrichment Analysis tool (https://chip-atlas.org/enrichment_analysis) was used to identify chromatin features enriched at SRA ChIP-seq peaks ($n = 295$) or PCD-RFP ChIP-seq peaks ($n = 13,477$) in untreated parental MCF7 cells. Settings were: genome, hg38; experiment type, "ChIP: Histone"; cell type class, "Breast"; significance threshold, 50; Dataset B, 10× random permutation of Dataset A. Enrichment p -values were calculated using Fisher's exact test. Features from drug-treated, hormone-treated, or genetically manipulated MCF7 cells, as well as infinite fold-enrichment values caused by zero Dataset B overlaps, were excluded from the analyses. Box plots include data for histone PTMs H3K27me3, H3K27ac, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me3, and H3K20me1. Only statistically significant values ($p < 0.001$) are included. (B) Tables show median enrichments for all histone modifications that showed at least one significant fold enrichment value.

		H3K27me3	H3K27ac	H3K36me3	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H4K20me1	H3K9me3
PCD-RFP		15.86	0.22	45.91	10.90	2.88	-7.36	-6.55	13.08	4.84
SRA		7.46	0.97	22.49	3.09	1.31	-1.34	0.03	13.08	6.03

15-state model	1 - Active Promoter	2	96	2	17	93	99	98	6	
	2 - Weak Promoter	2	14	6	53	94	95	44	9	
	3 - Inactive/poised Promoter	72	1	0	48	78	49	10	9	
	4 - Strong enhancer	1	97	15	96	99	75	86	11	
	5 - Strong enhancer	0	84	10	88	57	5	25	3	
	6 - Weak/poised enhancer	1	6	1	58	75	8	5	3	
	7 - Weak/poised enhancer	1	6	2	56	3	0	2	1	
	8 - Insulator	2	0	1	6	3	0	1	3	
	9 - Transcriptional transition	0	9	43	37	11	2	4	43	
	10 - Transcriptional elongation	0	0	47	0	0	0	0	3	
	11 - Weak Transcribed	0	0	3	0	0	0	0	2	
	12 - Polycomb-Repressed	27	0	0	0	0	0	0	2	
	13 - Heterochromatin; low signal	0	0	0	0	0	0	0	0	
	14 - Repetitive/Copy Number Variation	28	5	19	6	5	26	13	41	
	15 - Repetitive/Copy Number Variation	85	73	91	76	77	91	85	88	

18-state model	1 - Active TSS	0.030	0.954	0.028	0.030		0.986			0.003
	2 - Flanking TSS	0.008	0.055	0.010	0.078		0.780			0.004
	3 - Flanking TSS Upstream	0.013	0.995	0.022	0.963		0.980			0.003
	4 - Flanking TSS Downstream	0.015	0.107	0.041	0.915		0.923			0.007
	5 - Strong transcription	0.001	0.007	0.665	0.002		0.001			0.008
	6 - Weak transcription		0.002	0.042	0.008					0.002
	7 - Genic enhancer1	0.003	0.186	0.694	0.447		0.022			0.008
	8 - Genic enhancer2	0.003	0.848	0.788	0.906		0.289			0.013
	9 - Active Enhancer 1	0.003	0.916	0.018	0.920		0.061			0.007
	10 - Active Enhancer 2	0.003	0.502	0.011	0.145		0.009			0.007
	11 - Weak Transcribed	0.004	0.041	0.010	0.659		0.005			0.004
	12 - ZNF genes & repeats	0.016	0.004	0.084	0.005		0.045			0.778
	13 - Heterochromatin	0.001					0.001			0.109
	14 - Bivalent/Poised TSS	0.848	0.135	0.020	0.583		0.964			0.051
	15 - Bivalent Enhancer	0.723	0.049	0.011	0.711		0.034			0.035
	16 - Repressed PolyComb	0.664	0.002	0.004	0.002		0.002			0.027
	17 - Weak Repressed PolyComb	0.064			0.003					0.005
	18 - Quiescent/Low	0.002			0.001					0.003

Figure S3. Comparison of histone PTM overlaps with PCD-RFP and SRA ChIP-seq peaks with ChromHMM regulatory states. Top table: Z-score values of histone PTM ChIP-seq peaks and PCD-RFP or SRA ChIP-seq peak overlaps. Middle table: 15-state ChromHMM model with chromatin mark observation frequencies from Figure 1 in J. Ernst et al.¹ Bottom table: 18-state Chrom HMM model from the Epigenome Roadmap (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html).

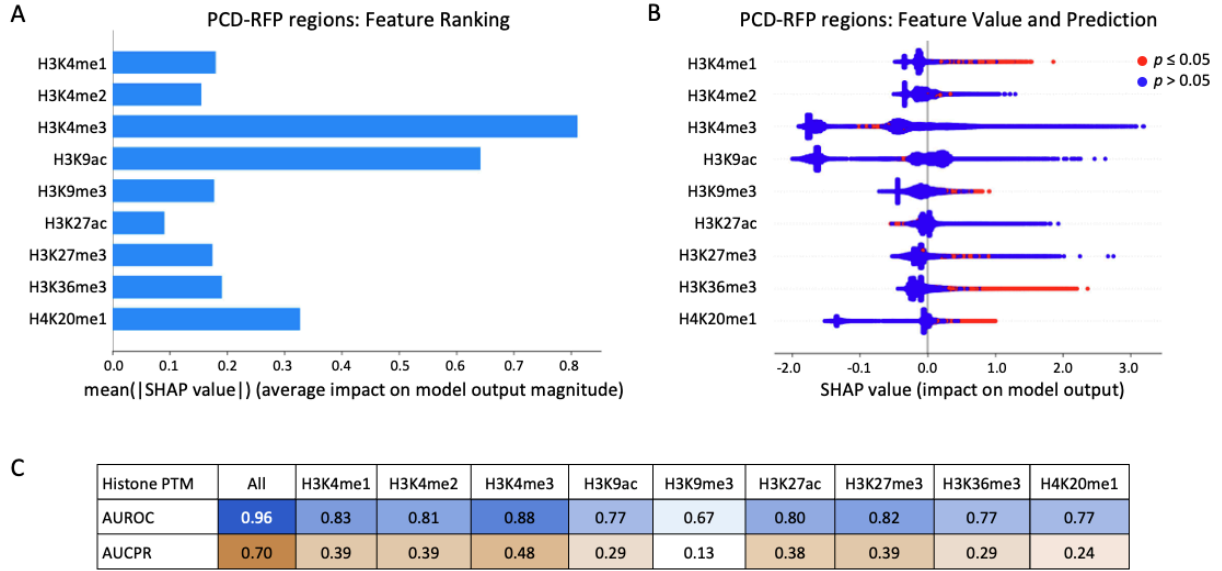


Figure S4. XGBoost prediction of PCD-fusion genome-wide binding, extended data. (A) SHAP bar chart ranking feature importance for PCD-RFP binding prediction. (B) SHAP bee swarm chart indicating distribution of each feature values respective to SHAP values. P-values indicate the significance of a histone PTM enrichment in a 200 bp bin. (C) AUROC and AUCPR values when predicting PCD-RFP binding using all 9 histone PTMs or only using a single indicated histone PTM.

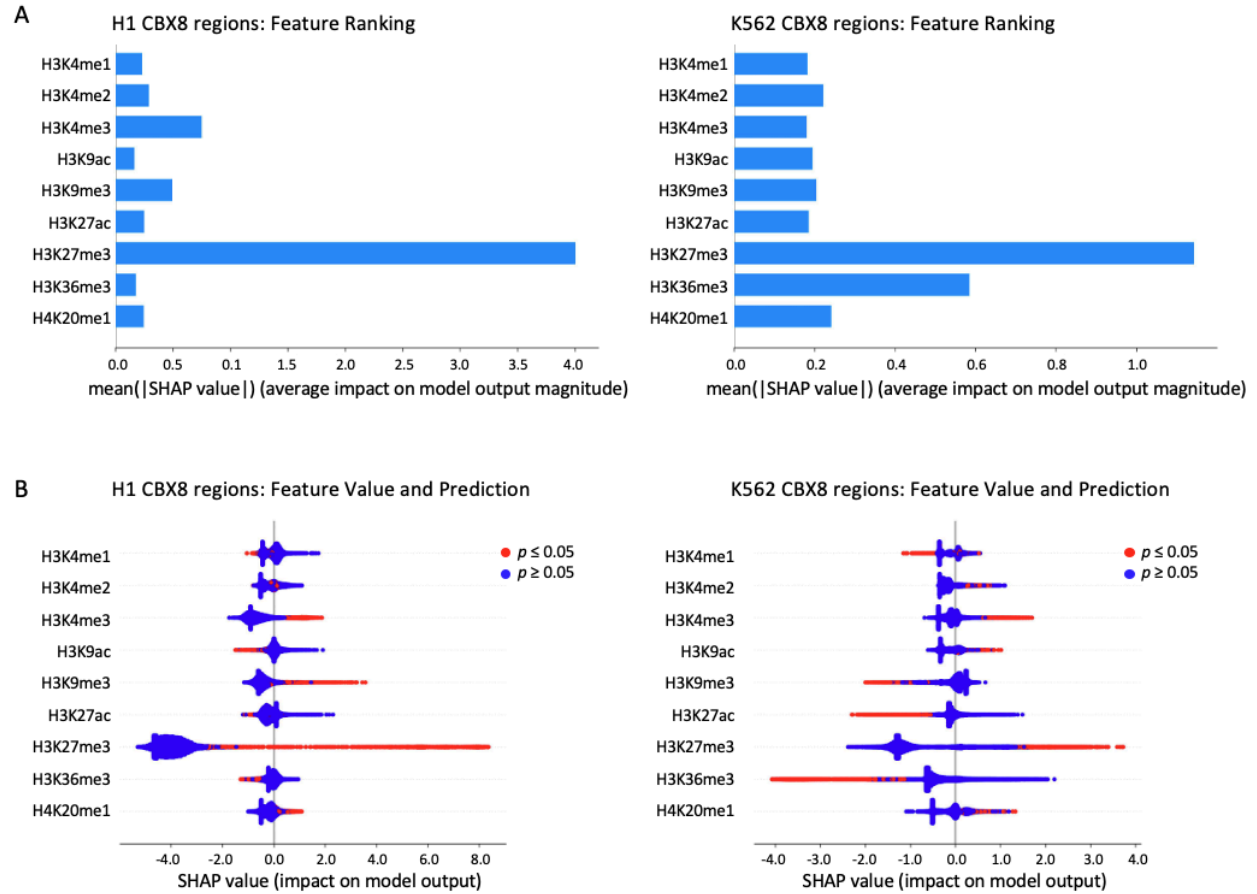


Figure S5. XGBoost prediction of endogenous CBX8 genome-wide binding. (A) SHAP bar chart ranking feature importance for CBX8 binding prediction. (B) SHAP bee swarm chart indicating distribution of each feature values respective to SHAP values. P-values indicate the significance of a histone PTM enrichment in a 200 bp bin.

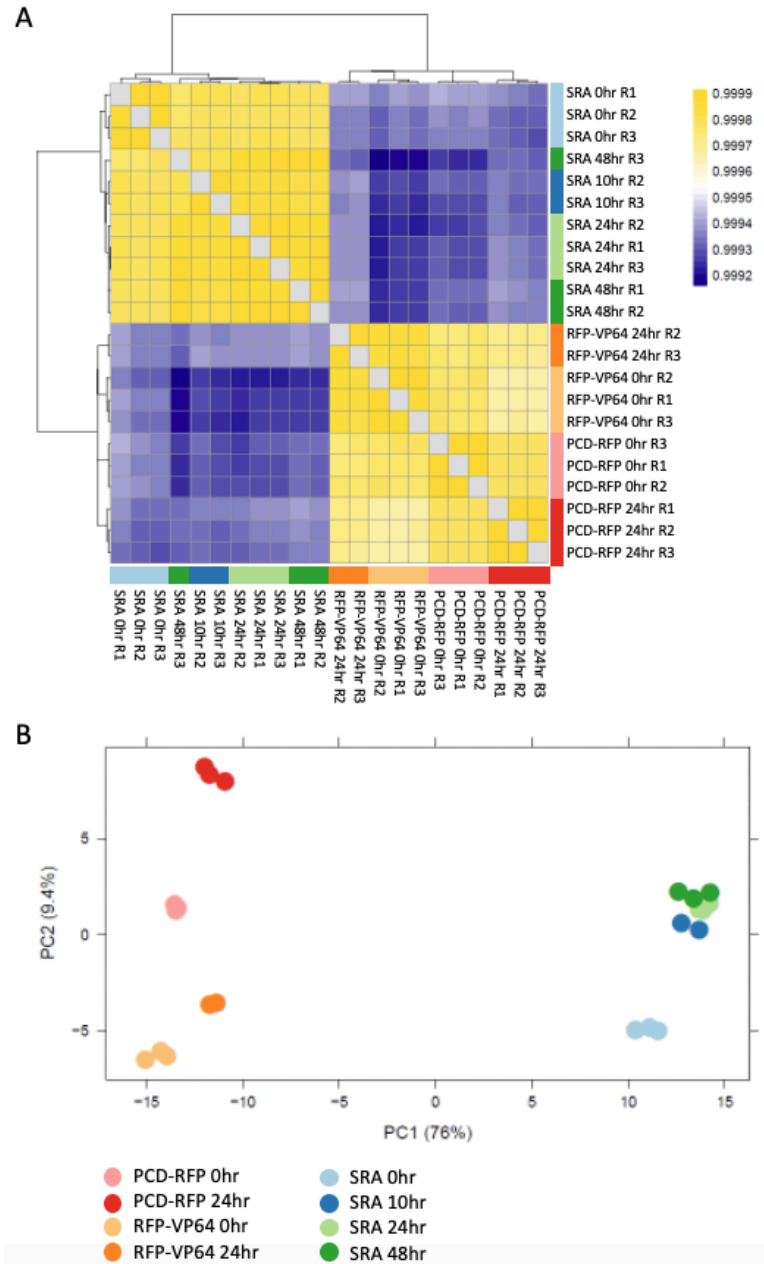


Figure S6. Clustering and principal component analysis of RNA-seq samples across experimental conditions. (A) Hierarchical clustering heatmap showing pairwise similarity of RNA-seq samples across conditions and time points, including SRA (PCD-RFP-VP64), PCD-RFP, and RFP-VP64 constructs at 0, 10, 24, and 48 hours. Samples are labeled by condition, time point, and biological replicate. Color scale represents similarity (e.g., correlation). (B) Principal component analysis (PCA) of RNA-seq samples. Each point represents a biological replicate, colored by construct and time point. The first two principal components (PC1 and PC2) are shown, with the percentage of variance indicated on each axis. Some conditions include two biological replicates, as described in the Methods.

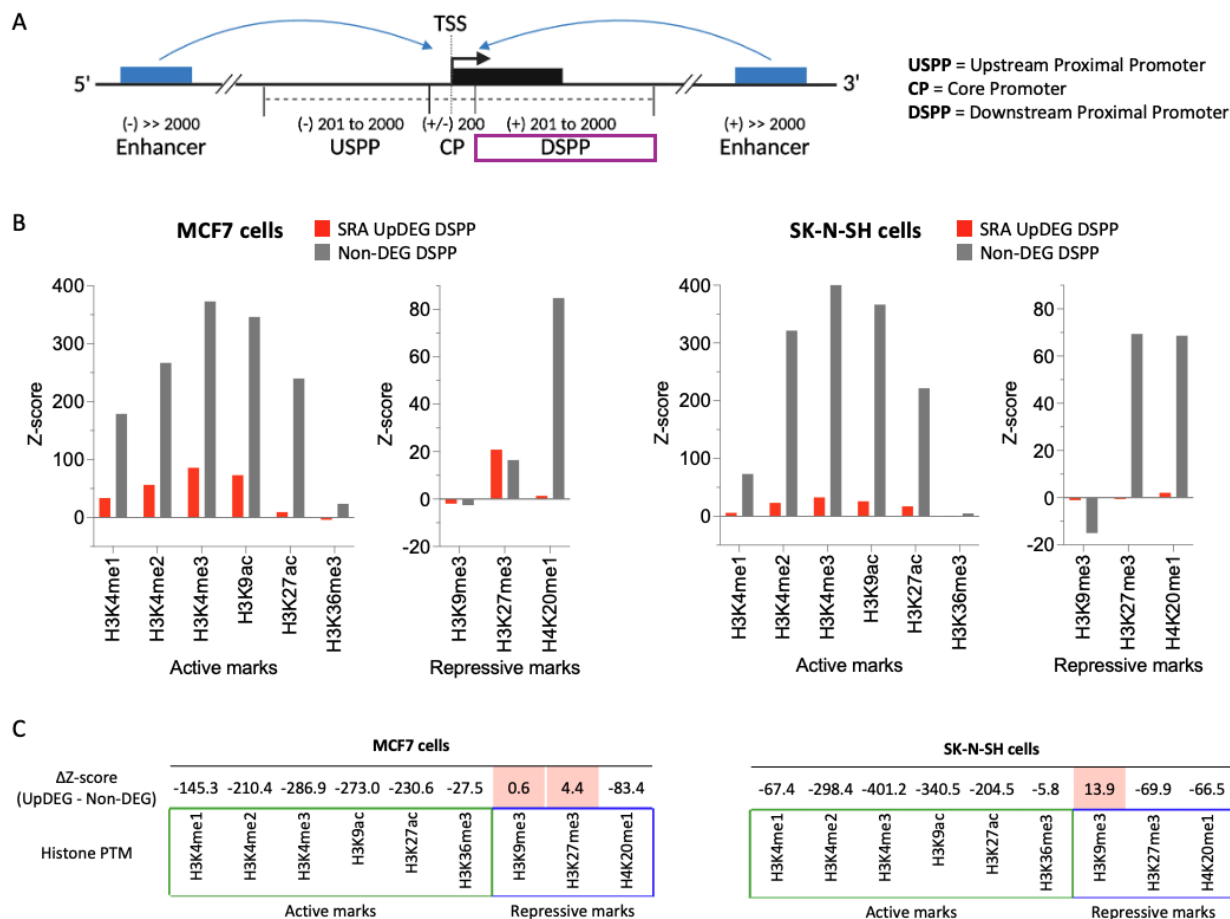


Figure S7. Conserved differential histone PTM signatures at downstream proximal promoter (DSPP) regions across two cell types. (A) Schematic of cis-regulatory element classification, highlighting the downstream proximal promoter (DSPP) region relative to the transcription start site (TSS), as defined in Fig. 7. (B) Overlap Z-scores for histone post-translational modifications (PTMs) at DSPP regions of SRA-upregulated genes (UpDEGs, red) and non-differentially expressed genes (Non-DEGs, gray) in MCF7 (left) and SK-N-SH (right) cells. Histone marks are grouped by association with active or repressive chromatin states. (C) Differential enrichment of histone PTMs, calculated as $\Delta Z = Z(\text{UpDEG}) - Z(\text{Non-DEG})$. Positive values are shaded red. ΔZ patterns are broadly consistent across cell types for active promoter-associated marks (e.g., H3K4me3, H3K4me2, H3K9ac), while repressive marks, including H3K27me3, show greater variability. The SK-N-SH dataset² is based on single-replicate RNA-seq and does not include SRA ChIP-seq; accordingly, it is used here for qualitative comparison of chromatin feature trends rather than direct equivalence to the MCF7 dataset.

SUPPLEMENTAL TABLES

Table S1. Transcription factor target genes within SRA UpDEG classes

TF		Target gene UpDEGs					
Symbol	FC	Early transient	Early sustained	Late transient	Late sustained	Delayed	Stochastic
HOXB6 (UpDEG)	0.60	(11) CSRP2, EYA4, HOXC9, ISM2, LRRC8C, MAFA, MFAP3L, NPL, PPP4R4, SRPX, TFPI2	(5) BOC, EPHB2, FAM171A1, FLT1, SEMA6B	(5) ARHGDIG, BACH2, ETV4, PANX2, ZMYND15	(3) COLEC12, MAPRE2, PDIA2	(6) ABCA10, ADD3, DMD, FAT4, TECTA, WIPF1	none
HOXB9 (UpDEG)	0.83	(16) CSRP2, DDR2, DLX3, FGD1, GPR156, IGF2BP1, ISM1, KBTBD11, KHDRBS3, MFAP3L, NPL, OLFML2A, PARVB, PLSCR4, PRR5L, SMAD9	(6) EPHB2, ESRRG, FLT1, LIMD2, LIPG, SEMA6B	(5) ETV4, ETV5, F2RL1, PNPLA3, TMEM169	(4) MAPRE2, PNCK, SPRY4, TIMP2	(10) DISC1, IGF2, LTBP3, MMP24, NR2F1, RBMS2, RPRM, SEMA6A, TMPRSS11E, TNIK	(1) CHSY3
HOXB13 (UpDEG)	1.02	(22) CLIP2, ERG, HOXC9, HTR6, IGF2BP1, ISM1, KBTBD11, KCNG1, KHDRBS3, KIF21B, MFAP3L, PLCL2, PLSCR4, PRR5L, RGMA, RGS9, RIMS3, STEAP2, SYNE3, TNFAIP8L3, TRNP1, TSHZ3	(4) EPHB2, GJB2, HTR1D, SEMA6B	(6) ARHGDIG, ANTXR2, F2RL1, GXYLT2, TMEM169, WFDC2	(2) PDIA2, SPRY4	(8) RBMS2, RPRM, SEMA6A, SLC25A27, STON1, TMPRSS11E, TNIK, WIPF1	(1) LRP12
HOXC8 (UpDEG)	1.08	(12) BCL2L10, CSRP2, DDR2, FAM78A, HOXC9, ISM2, MAFA, NPL, NRTN, PPP4R4, STEAP2, TRNP1	(5) ESRRG, FAM171A1, GMPR, RASGRP2, SEMA6B	(5) BACH2, ETV4, PDLIM3, PIK3AP1, ZMYND15	(2) DPP4, MAPRE2	(4) ADD3, SEMA6A, SMPD1, WIPF1	(1) JAZF1
HOXC9 (UpDEG)	0.64	(10) CSRP2, DPYSL5, FAM78A, HOXC9, KBTBD11, KHDRBS3, KIAA1549L, NRTN, PRR5L, RCAN2	(2) ESRRG, FLT1	(3) BACH2, ETV4, THRA	(2) BEND7, MAPRE2	(5) DISC1, FAT4, MMP24, SEMA6A, SEMA6D	(1) JAZF1
HOXC10 (UpDEG)	0.64	(7) DPYSL5, FAM78A, HOXC8, KHDRBS3, KIF21B, RYR1, TUBB2B	(4) ESRRG, FLT1, KIF1A, SERPINE2	(3) GXYLT2, PNPLA3, SLC29A4	(2) BEND7, FBLN7	(3) DISC1, DMD, STON1	(1) JAZF1
YY1	n/a	(14) BCL2L10, DDR2, DPYSL5, GLIPR2, HAP1, HOXC10, KCNG1, MFAP3L, PRIMA1, PRR5L, RASSF5, STEAP2, TNFAIP8L3, TUBB2B	(8) CYBRD1, FXYP6, GJB2, GMPR, HTR1D, LARP6, LIPG, RASL10B	(4) ETV4, GXYLT2, PDLIM3, TMEM86A	(5) ATP1B2, BEAN1, COL23A1, COLEC12, DBN1	(4) ADD3, IGF2, STON1, WIPF1	(1) LGI2

The TF Database³ was used to retrieve all target promoters and genes for each transcription factor (TF) shown in the first column. Target gene UpDEGs were identified by cross-checking this output with the entire list of 397 UpDEGs. For TFs that are also UpDEGs, the maximum log2 fold change (FC) across the three time points (10, 24, 48 hours) is shown in the second column.

Table S2. RT-qPCR Primer Pairs

Target	Forward 5' to 3'	Reverse 5' to 3'
SRA	ATCGCCTGGAGCCAATTCC	AACCTCCCACATCTCCCCC
PCD-RFP	ATCGCCTGGAGCCAATTCC	AACCTCCCACATCTCCCCC
GAPDH	TCTCCTCTGACTTCAACAGCGAC	CCCTGTTGCTGTAGCCAAATTC

The same primer pair was used for the fusion protein transcripts, SRA and PCD-RFP, which binds to the mCherry sequence that is present in both mRNA sequences.

Table S3. ENCODE MCF7 Datasets

Type	Feature	ENCODE Accession
Histone ChIP-seq	H3K4me1	ENCFF763NCP
	H3K4me2	ENCFF442RRY
	H3K4me3	ENCFF163MXP
	H3K9ac	ENCFF327XJC
	H3K9me3	ENCFF481DZL
	H3K27ac	ENCFF138YNG
	H3K27me3	ENCFF163QKN
	H3K36me3	ENCFF910BRP
	H4K20me1	ENCFF366GLZ
Transcription Factor ChIP-seq	ARID3A	ENCFF907DWT
	ATF7	ENCFF703ZRP
	BMI1	ENCFF977TJK
	CEBPB	ENCFF464JOC
	CHD1	ENCFF923MDX
	CLOCK	ENCFF045RXD
	COPS2	ENCFF102ABZ
	CREB1	ENCFF954WHX
	CSDE1	ENCFF722GUE
	CTBP1	ENCFF922FKY
	CTCF	ENCFF507CRU
	CUX1	ENCFF663YXL
	DDX20	ENCFF895IXP
	DPF2	ENCFF142KHL
	E2F8	ENCFF919YCR
	E4F1	ENCFF555HAI
	EGR1	ENCFF487DKA
	ELF1	ENCFF751TAA
	ELK1	ENCFF045XJM
	EP300	ENCFF708NMR
	ESRRA	ENCFF059ZSC
	FOS	ENCFF327LTA
	FOSL2	ENCFF704HAX
	FOXA1	ENCFF512UGW
	FO XK2	ENCFF477VZD

	FOXM1	ENCFF405RBL
	GABPA	ENCFF494GAR
	GATA3	ENCFF971AZB
	GATAD2B	ENCFF560WGB
	GTF2F1	ENCFF736KUY
	HCFC1	ENCFF103ABY
	HDAC2	ENCFF178WAQ
	HDGF	ENCFF178LPH
	HES1	ENCFF509HTJ
	HSF1	ENCFF505ZBM
	JUN	ENCFF730TVS
	JUND	ENCFF990FGN
	LARP7	ENCFF148SFF
	MAFK	ENCFF174HAP
	MAX	ENCFF315FIO
	MAZ	ENCFF297WGG
	MBD2	ENCFF517UQD
	MLLT1	ENCFF654ISN
	MNT	ENCFF623IRB
	MTA1	ENCFF883HBF
	MTA2	ENCFF557XGU
	MTA3	ENCFF486PVE
	NBN	ENCFF242QKO
	NCOA3	ENCFF770MLI
	NEUROD1	ENCFF501HJF
	NFIB	ENCFF631HNZ
	NFRKB	ENCFF411WFM
	NFXL1	ENCFF537TBX
	NONO	ENCFF875CIS
	NR2F2	ENCFF678MPN
	NRF1	ENCFF113MFU
	PAX8	ENCFF775POG
	PKNOX1	ENCFF737NUU
	PML	ENCFF911FVQ
	POLR2A	ENCFF827YIP
	PPP1R10	ENCFF167JMR

	RAD21	ENCFF775EKJ
	RAD51	ENCFF522WFI
	RCOR1	ENCFF536KZN
	REST	ENCFF068TEY
	RFX1	ENCFF310ZPD
	RFX5	ENCFF085XIV
	SIN3A	ENCFF807QKK
	SIX4	ENCFF032MSQ
	SMARCA5	ENCFF297JNO
	SMARCE1	ENCFF375LRP
	SNIP1	ENCFF231LYV
	SP1	ENCFF367HVV
	SREBF1	ENCFF791DFW
	SRF	ENCFF534AJF
	SUZ12	ENCFF598JRC
	TAF1	ENCFF418WWV
	TARDBP	ENCFF065RQM
	TCF7L2	ENCFF711CKB
	TCF12	ENCFF031XVC
	TEAD4	ENCFF977ZGZ
	TOE1	ENCFF318YIK
	TRIM22	ENCFF100OCC
	YBX1	ENCFF033VON
	ZBTB1	ENCFF062ZHZ
	ZBTB7B	ENCFF073COQ
	ZBTB11	ENCFF510SQD
	ZBTB33	ENCFF506CTQ
	ZBTB40	ENCFF539ZHK
	ZFX	ENCFF482RYD
	ZHX2	ENCFF724GPW
	ZKSCAN1	ENCFF239BWH
	ZNF8	ENCFF550FAH
	ZNF24	ENCFF505ORH
	ZNF207	ENCFF668FJE
	ZNF217	ENCFF491CXP
	ZNF217	ENCFF343WRL

	ZNF444	ENCFF101ION
	ZNF507	ENCFF659LTF
	ZNF512B	ENCFF413BLF
	ZNF574	ENCFF914VHT
	ZNF579	ENCFF913XOW
	ZNF592	ENCFF950SDY
	ZNF687	ENCFF910XUA
Chromatin accessibility	DNase-seq	ENCFF137FTO
	ATAC-seq	ENCFF976UNK

Table S4. ENCODE SK-N-SH Datasets

Type	Feature	ENCODE Accession
Histone ChIP-seq	H3K4me1	ENCFF864MVD
	H3K4me2	ENCFF348FTU
	H3K4me3	ENCFF948FXU
	H3K9ac	ENCFF671RKZ
	H3K9me3	ENCFF218PAX
	H3K27ac	ENCFF790UBM
	H3K27me3	ENCFF493XGA
	H3K36me3	ENCFF140RVM
	H4K20me1	ENCFF626ZQQ

SUPPLEMENTAL REFERENCES

1. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
2. Nyer, D. B., Daer, R. M., Vargas, D., Hom, C. & Haynes, K. A. Regulation of cancer epigenomes with a histone-binding synthetic transcription factor. *NPJ Genom Med* **2**, (2017).
3. Plaisier, C. L. *et al.* Causal Mechanistic Regulatory Network for Glioblastoma Deciphered Using Systems Genetics Network Analysis. *Cell Syst* **3**, 172–186 (2016).