

Supplementary Methods

Conservation Analysis

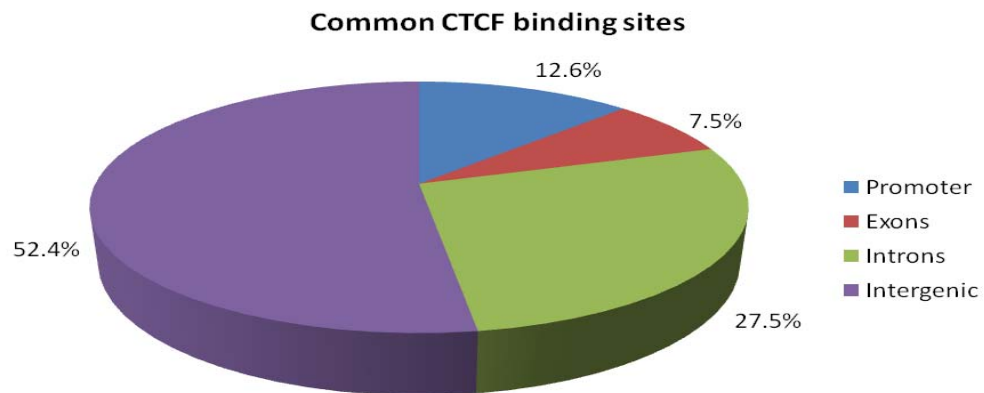
To make sure the CTCF binding sites that are common in all three cell lines (Jurkat, HeLa, and CD4) are evolutionarily conserved before we use them to demarcate the genome, we use *cis*-regulatory element annotation system (CEAS) to generate an average conservation plot (37). Each common CTCF binding sites is extended to 3kb and aligned at the center. An average score based on a phylogenetic Hidden Markov Model that estimates the probability that each nucleotide belong to a conserved region are then computed using PhastCons (38).

DNase I hypersensitive sites

A site detected by its hypersensitivity to DNase I is associated with a region of chromatin that are relaxed. Here, we set out to find whether the CTCF binding sites shared by three cell lines (i.e. Jurkat, CD4 and HeLa) overlap with these sites. This information can be important in understanding whether CTCF can demarcate genome into distinct region of gene activation. DNase I hypersensitive sites from HeLa-S3 cell were downloaded from UCSC genome browser. HotSpot algorithm with an FDR of 0.005 were used to identify these DNase I hypersensitive zones (39).

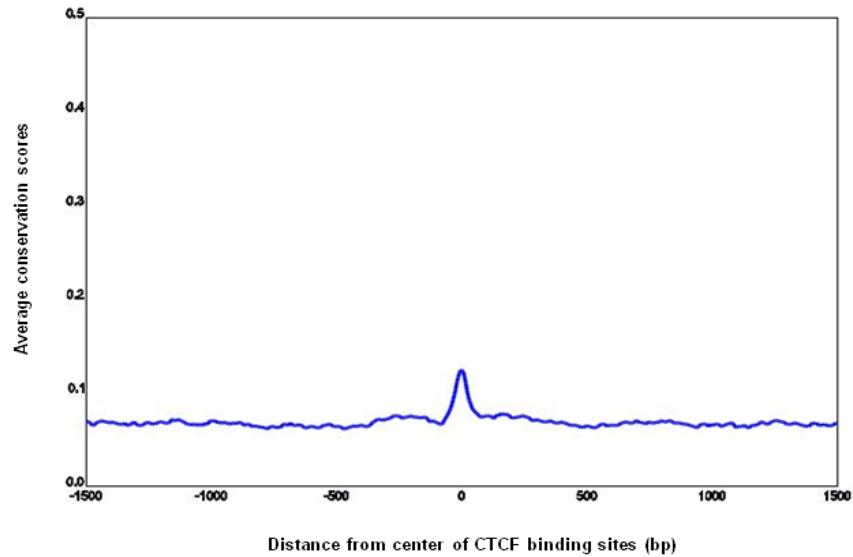
Supplementary Figures

Supplementary Figure 1



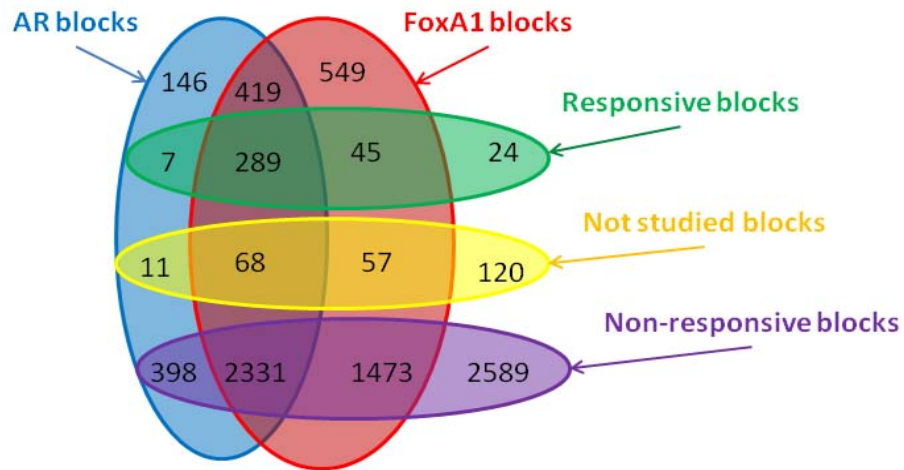
Supplementary Figure 1. Genome-wide distribution of CTCF binding sites shared by Jurkat, CD4 and HeLa cells. Majority of common CTCF binding sites are located in intergenic regions but over $\frac{1}{4}$ are in introns and many are occupying promoter and exons regions as well.

Supplementary Figure 2



Supplementary Figure 2. Conservation plot for all common CTCF binding sites. Y-axis indicating average conservation score of all common CTCF binding sites which are extended to 3kb, aligned and centered, calculated using PhastCons which are based on phylogenetic hidden markov model. The plot shows that area around the center of the CTCF binding sites are elevated more than two times the genomic background region which indicate that the binding sites are evolutionarily conserved.

Supplementary Figure 3



Supplementary Figure 3. Number of overlap between different block types. Number of overlap between blocks with AR sites (AR blocks), blocks with FoxA1 sites (FoxA1 blocks), blocks with at least one responsive genes (Responsive blocks), blocks with only responsive genes (Non-responsive blocks) and blocks which contain genes which are not being studied. Most of the responsive blocks (80%) have both AR and FoxA1 inside.

Supplementary Tables

Supplementary Table 1. Tested genes by TaqMan® Array.

Responsive genes in AR blocks
AMIGO2
ANKRD29
CLDN8
DUSP4
GRHL2
JAG1
KIAA0355
KLF5
LPAR3
MBOAT2
MTMR9
NFIL3
OPRK1
SLC16A6
SLC19A2
STK17B
TIPARP
ARL4C
C1orf116
C3orf25
FKBP5
FZD5
IGF1R
KLK4
LRIG1
MAGEC2
NKX3
NT5DC3
PMEPA1
PPFIBP2
PTGER4
SGK1
SNAI2
TMCC3
TMPRSS2
TNFAIP8

Responsive genes in no AR blocks
APBB2
ATXN2
CD2AP
CXXC5
CYBASC3
ELF3
GDF15
IGFBP3
KIAA1432
NET1
OSBPL8
OSR2
PFKFB2
PIK3R3
RCOR3
RFX3
RNF144A
SNRK
ST3GAL1
TNK2
VEGFA
XBP1
YOD1
ZBTB43

Supplementary Table 2. ChIP primers

Region 1+	AGGGGGCGCAAGAATCCCAC
Region 1-	GCAGCAGAGTGGCAAGAAGGC
Region2+	GGCCTCCTGCCATCTGGTG
Region2-	AGCCTGTGTGGGAAGCGTGA
Region3+	TGCATCAGACAGCTCGGCC
Region3-	GTTTCCCAGGCCTGCCACGA
Region 4+	AGGGCTGCGGACATTTGGGA
Region 4-	CACACTGCCCCCTGCCGTTA
Region 5+	TGTGGGAGGAAGGGGGAGGAC
Region 5-	CCTGCCCTGTGATGCCACCT
Region6+	AGCCACCATGCCAGCCTTT
Region 6-	GCGGGCAACCTCTTCAGTCAT

Supplementary Table 3. A summary of number of CTCF blocks with and without RefSeq genes^a.

CTCF Blocks		Number of blocks
Blocks w/o RefSeq gene		4,164
Blocks w/ RefSeq gene	Not on microarray	256
	Responsive gene	365
	Only non-responsive gene	6,791

^aBlocks with RefSeq gene not on microarray are blocks which contains RefSeq gene which are not being studied (i.e. their gene expression are not measured). Blocks with responsive genes are those that contain at least 1 responsive gene. Blocks with non-responsive gene are those that include only genes which are not responsive to androgen induction.

Supplementary Table 4. A summary of number of AR blocks^b.

AR Blocks	Number of blocks
At least one responsive gene	296
Only non-responsive genes	2,729
No genes studied	644

^bAR Blocks with non-responsive genes are those that contain only non-responsive genes. AR Blocks with at least one responsive gene are those that contain at least one gene that are responsive to androgen induction. AR blocks with no genes studied are those with RefSeq genes but their expressions are not measured on our microarray.

Supplementary Table 5. A summary of FoxA1 blocks^c.

FoxA1 Blocks	Number of blocks
At least one responsive gene	334
Only non-responsive genes	3,804
No genes studied	1,093

^cFoxA1 Blocks with non-responsive genes are those that contain only non-responsive genes. FoxA1 Blocks with at least one responsive gene are those that contain at least one gene that are responsive to androgen induction. FoxA1 blocks with no genes studied are those with RefSeq genes but their expressions are not measured on our microarray.

Supplementary Table 6. A summary of blocks containing AR and/or FoxA1 and the genes that are being studied (on microarray)^d.

AR-FoxA1 Blocks			
AR	FoxA1	Genes	Number of blocks
Yes	Yes	Responsive	289
		Non-responsive	2,331
Yes	No	Responsive	7
		Non-responsive	398
No	Yes	Responsive	45
		None responsive	1,473
No	No	Responsive	24
		None responsive	2,589

^dAR: yes, FoxA1: yes, Genes: responsive refers to blocks containing AR and FoxA1 sites and have at least one responsive gene. AR: no, FoxA1: no, Genes: non-responsive refers to blocks that do not have AR nor FoxA1 and none of the gene it contained is responsive to androgen induction. None of the blocks is overlapping. 2,620 blocks have AR, FoxA1 and genes that are being studied. Of these, 289 blocks (11%) contain at least 1 responsive gene. The remaining 2331 blocks (89%) have genes that are non-responsive to androgen treatment. There are 405 blocks that contain AR, no FoxA1 and at least one RefSeq genes which are being studied. Of these, only 7 blocks (2%) consist of at least one responsive gene. The remaining 398 blocks (98%) have genes that are all non-responsive. There are 1,518 blocks with FoxA1, no AR and at least one RefSeq genes which are being studied. Of these, 45 blocks (3%) consist of at least one responsive gene. The remaining 1,473 blocks (97%) have genes that are all non-responsive. There are 2,613 blocks with no FoxA1, no AR and at least one RefSeq genes that are being studied. Of these, 24 blocks (1%) consist of at least one responsive gene. The remaining 2,589 blocks (99%) have genes that are all non-responsive.

References

37. Ji, X., Li, W., Song, J., Wei, L. and Liu, X.S. (2006) CEAS: cis-regulatory element annotation system. *Nucleic Acids Res*, **34**, W551-554.
38. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, **15**, 1034-1050.
39. Sabo, P.J., Hawrylycz, M., Wallace, J.C., Humbert, R., Yu, M., Shafer, A., Kawamoto, J., Hall, R., Mack, J., Dorschner, M.O. *et al.* (2004) Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc Natl Acad Sci U S A*, **101**, 16837-16842.