Supplemental Figures

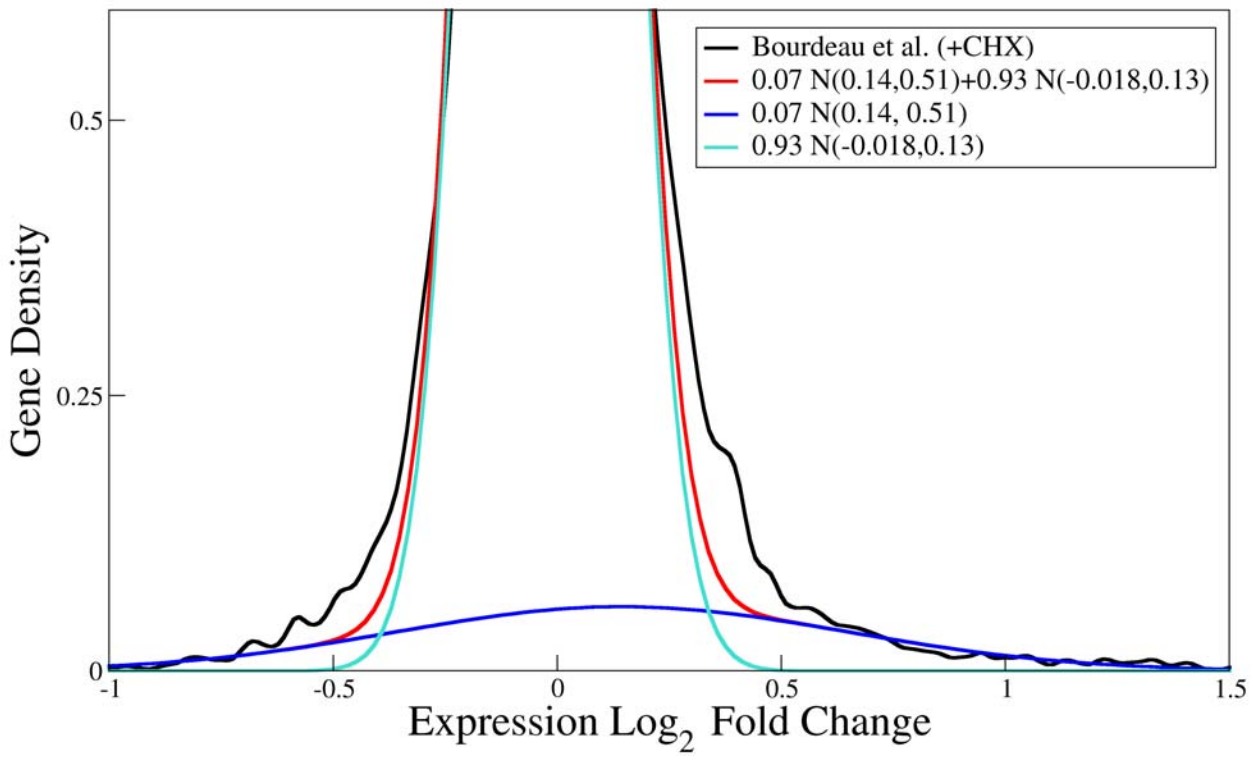


Figure S1: 2-component Gaussian mixture model of Bourdeau et al.'s fold-change distribution

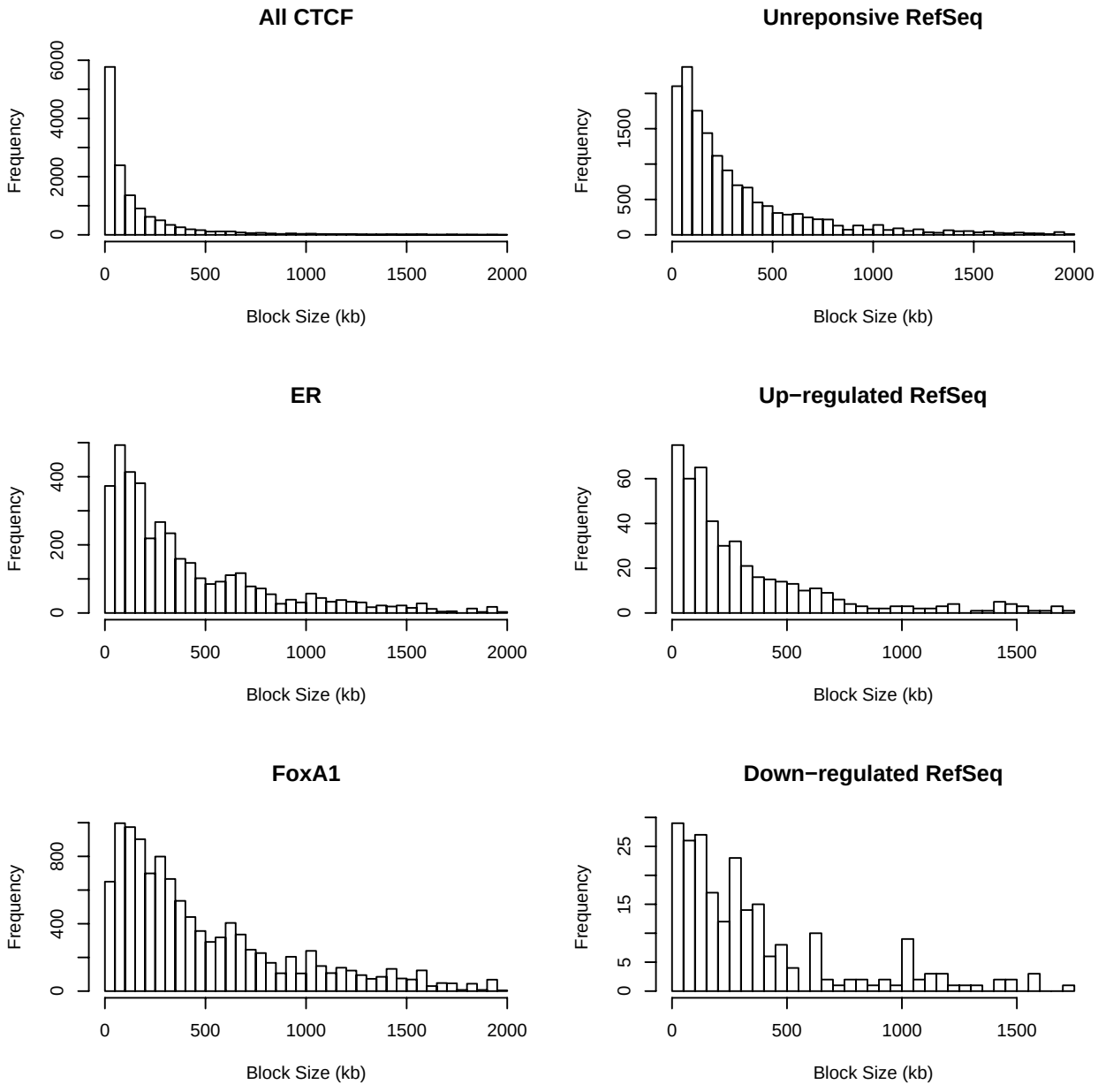
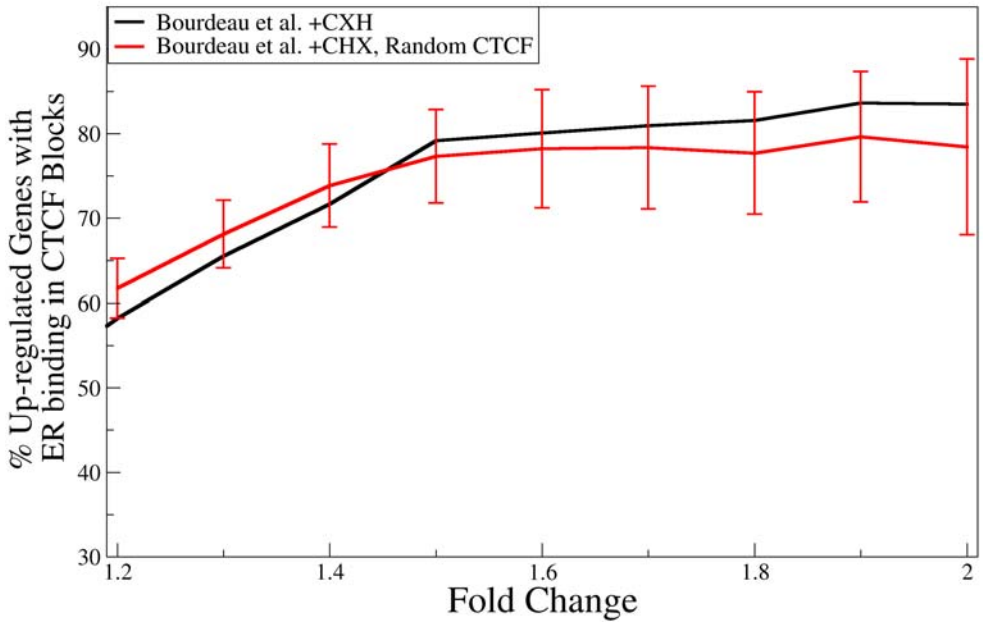


Figure S2: Distributions of CTCF block sizes

(A)



(B)

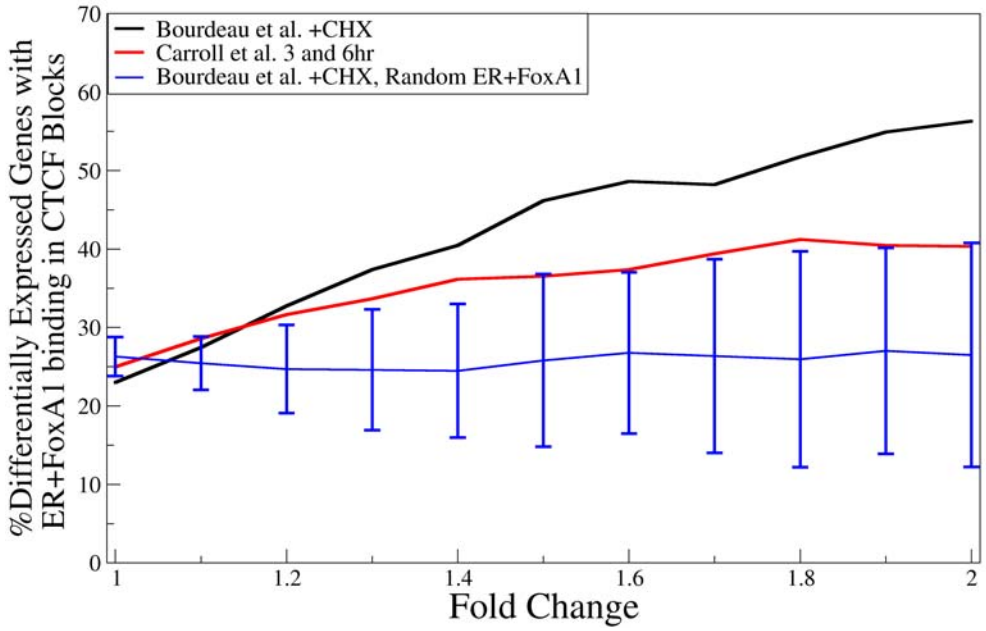
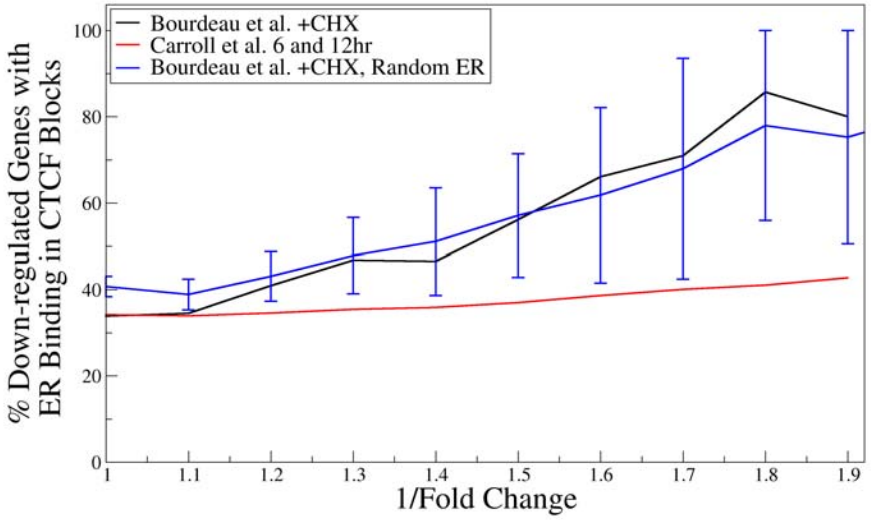


Figure S3: Percentage of up-regulated genes with ER/FoxA1. (A) Percentage of up-regulated genes with ER within the same CTCF blocks as the genes. Red line shows the result from 10,000 simulations of random CTCF sites, with the error bars representing the maximum deviation from the mean. (B) Percentage of up-regulated genes with ER and FoxA1 overlapping sites within the same CTCF blocks as the genes. 2,656 random ER and FoxA1 overlapping binding sites were simulated 10,000 times. The error bars represent the maximum deviation from the mean among the 10,000 simulations. The fraction of the early up-genes identified by Carroll et al. having ER+FoxA1 binding within the same CTCF blocks as the genes is less significant than that found under cycloheximide treatment.

(A)



(B)

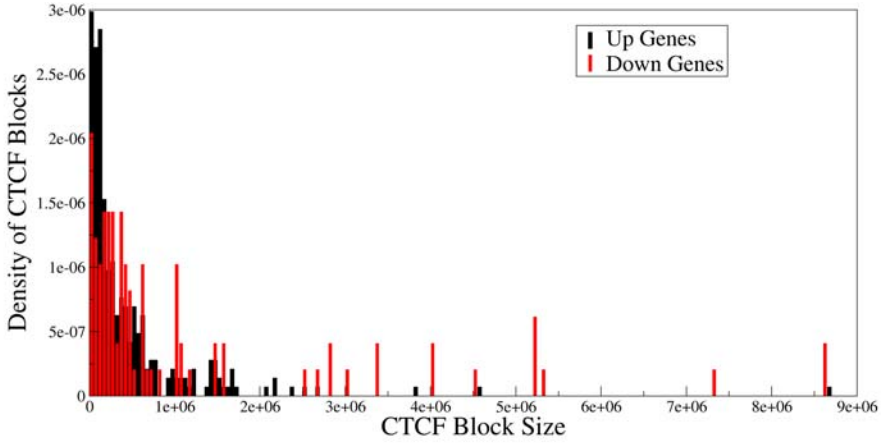


Figure S4: Percentage of down-regulated genes with ER and FoxA1 overlapping sites within the same CTCF blocks as the genes. (A) The distribution of ER was not biased towards down-regulated genes compared to random distributions. Because AP1 and NRIP, 2 essential components of ER-mediated down-regulation mechanism, were not translated after estrogen induction in the presence of CHX, estrogen-induced down-regulation was greatly reduced compared to the case in the absence of CHX. (B) Distribution of CTCF block sizes containing up-genes with fold-change greater than 1.5 and down-genes with fold-change less than 1.0/1.5. It can be seen that down-genes tend to have a greater number of large CTCF blocks than up-genes. The high percentage observed for randomly distributed ER binding sites in (A) can be explained by the fact that the number of down-genes with fold-change less than 1.0/1.5 was small (< 100) while the average CTCF block size was large (> 1.1 Mbp).

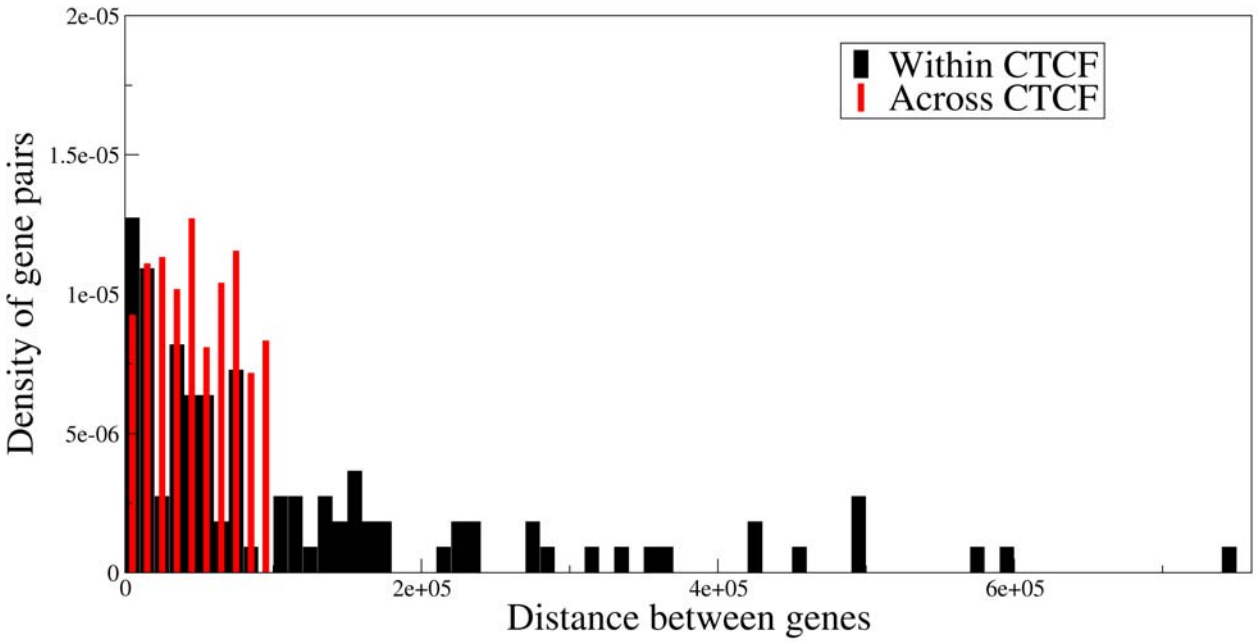
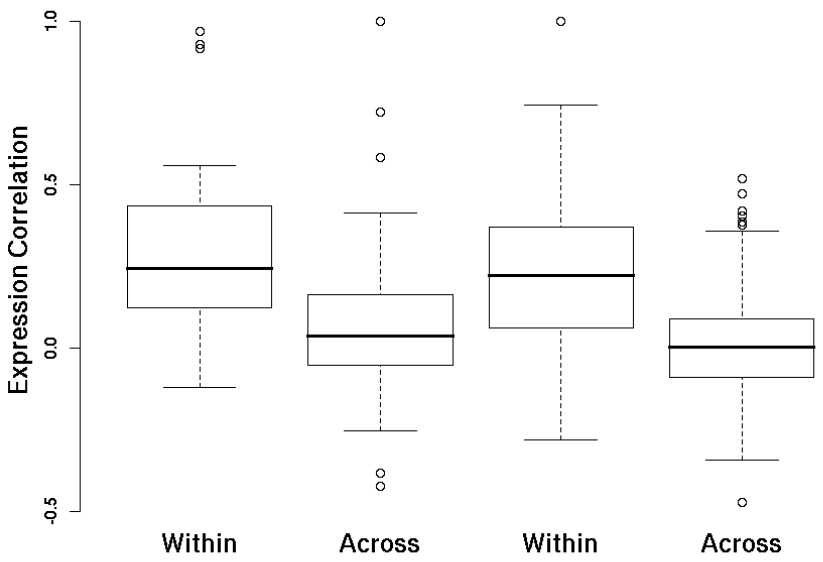


Figure S5: Distribution of pairwise distance between estrogen-reponsive genes within CTCF and across CTCF blocks. For the correlation analysis shown in Figure 3B, we considered only those gene pairs across CTCF blocks with a maximum separation distance of 100kb between the genes. Thus, the high correlation of gene expression within CTCF blocks was not biased by the distance between genes.

(A)



(B)

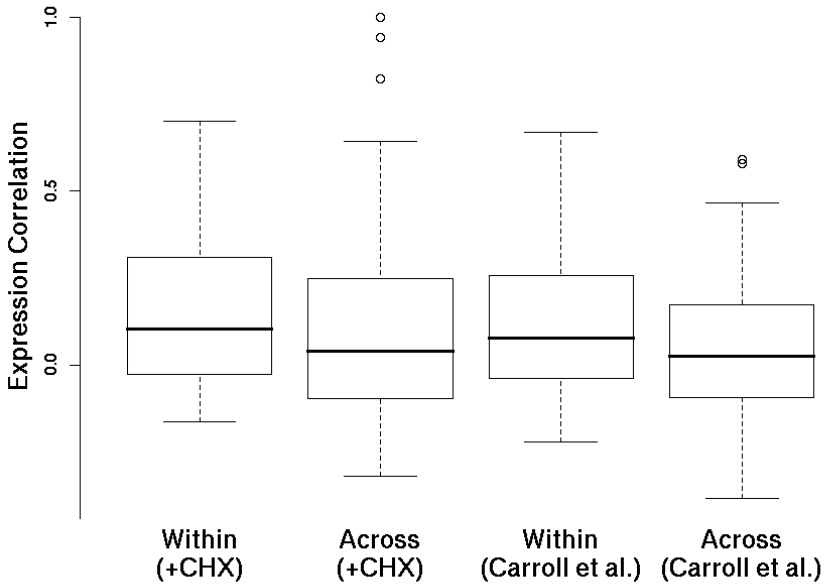
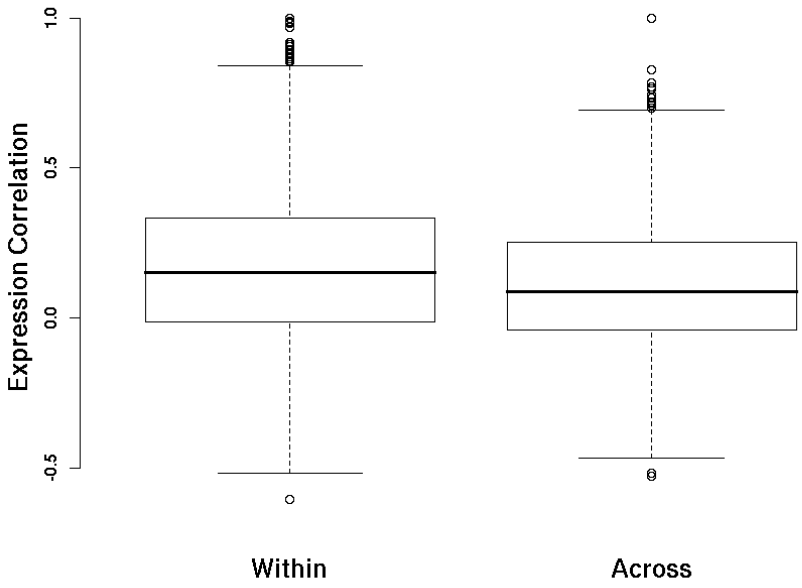


Figure S6: Correlation of expression of estrogen-responsive genes within and across CTCF blocks. (A) Correlation across 286 primary breast cancer samples (Wang et al. 2005). p-value for the difference $= 1.07 \times 10^{-7}$ (Bourdeau et al. +CHX), p-value $= 6.636 \times 10^{-11}$ (Carroll et al.) (B) Correlation across NCI60 cell lines with randomized CTCF blocks, where the mid-points of original blocks were taken as new boundaries, unless a block contained more than 1 differentially expressed gene, in which case the mid-point between 2 randomly chosen differentially expressed gene TSSs were chosen. p-value $= 0.3$ (Bourdeau et al. +CHX), p-value $= 0.2$ (Carroll et al.) Because CTCF blocks can be large, some gene pairs were assigned to same blocks even after randomization, leaving some residual correlation. In this analysis, we imposed the minimum distance between genes within CTCF blocks to be 50kb.

(A)



(B)

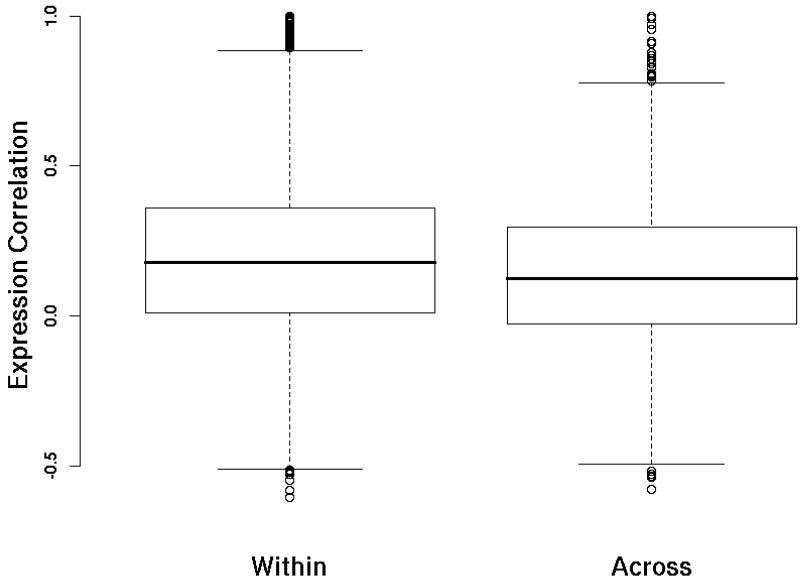


Figure S7: Correlation of gene expression within and across CTCF blocks. (A) Correlation of RefSeq genes in ER blocks across NCI60 cell lines is higher within CTCF blocks than across CTCF blocks (p-value $< 1.0 \times 10^{-300}$). (B) Correlation of all RefSeq genes across NCI60 cell lines is higher within CTCF blocks than across CTCF blocks (p-value $< 1.0 \times 10^{-300}$).

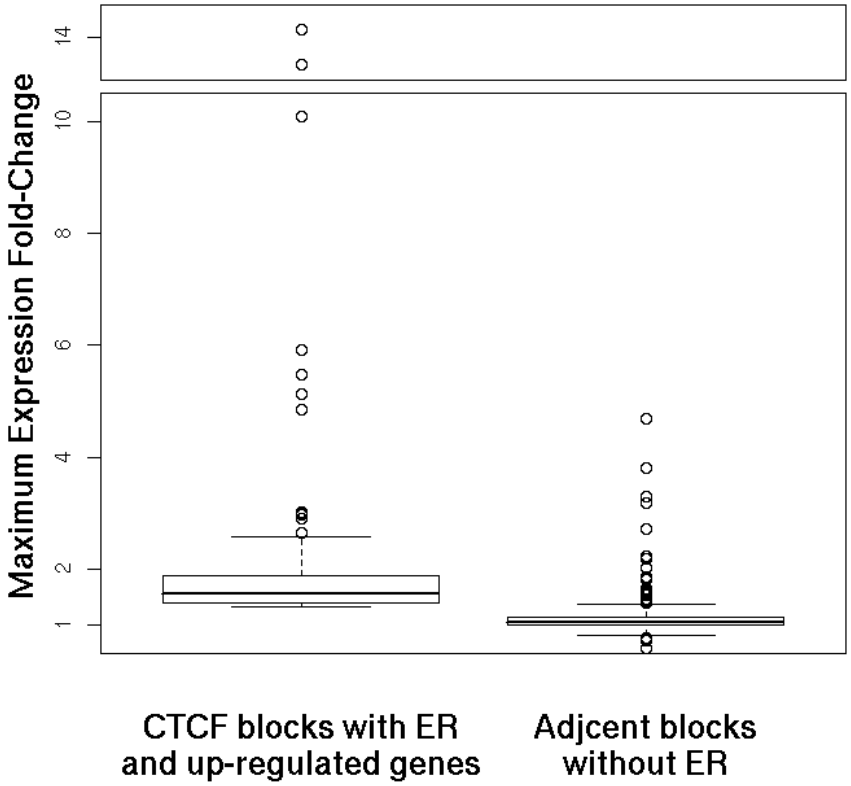


Figure S8: Maximum expression fold-changes in randomized CTCF blocks with ER and immediately adjacent blocks without ER. In randomized CTCF blocks without any ER, there were several highly estrogen-responsive genes, suggesting that ER activation "leaks" across these randomly chosen barriers.

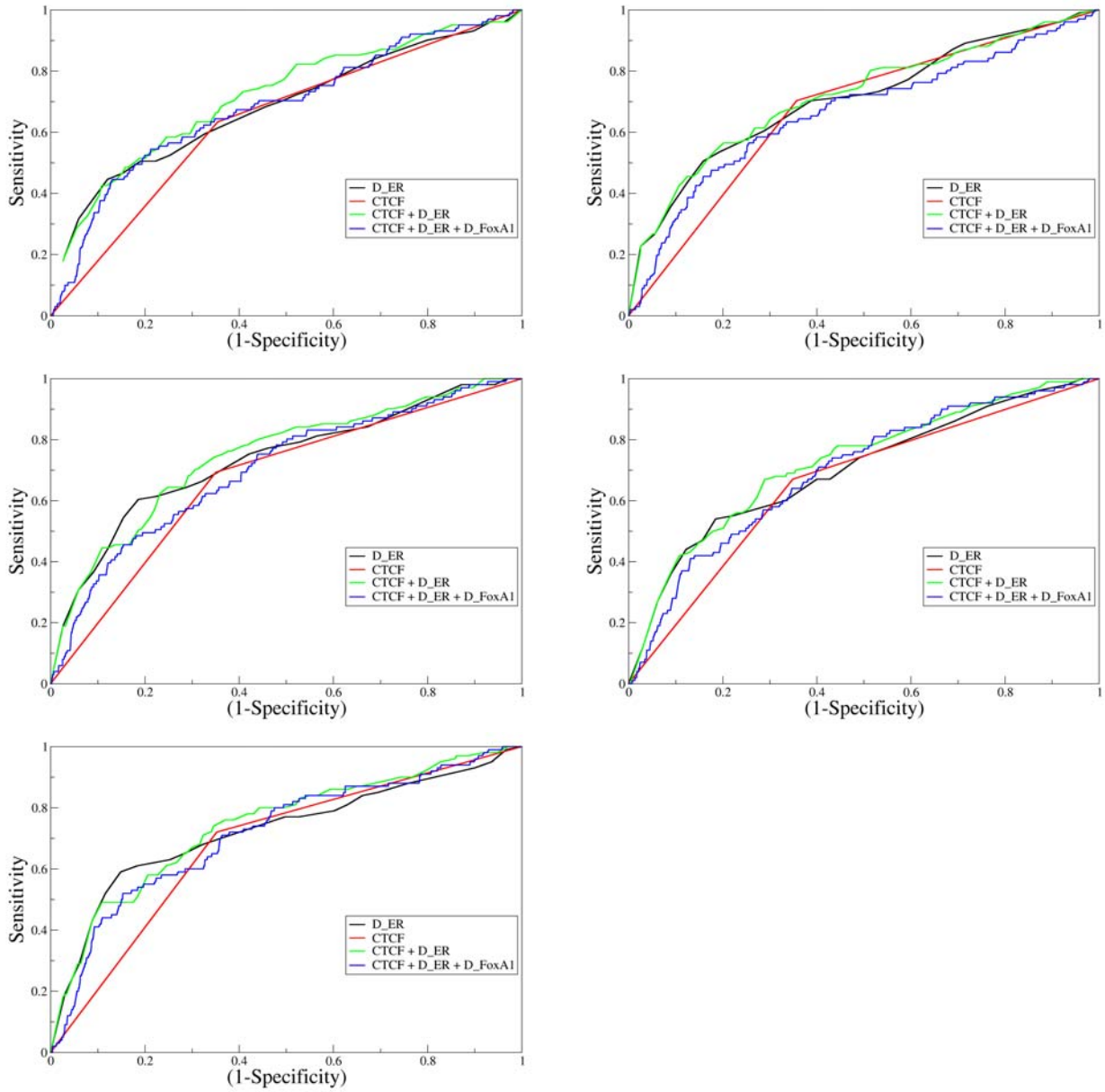
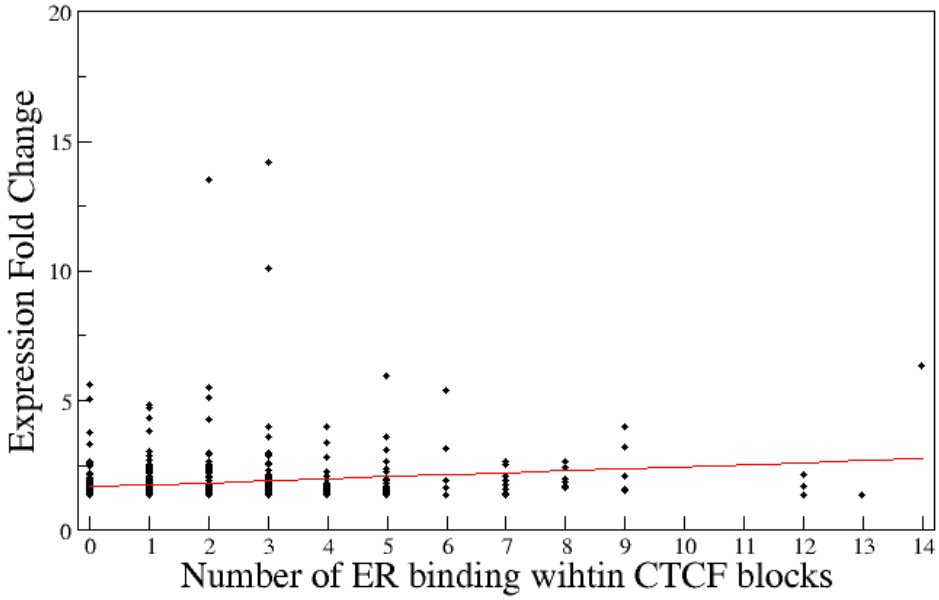


Figure S9: ROC curves for 5-fold cross validation of 4 Bayesian network models in Figure 4A. The mean area under the curve (AUC) was 0.71 ± 0.01 for D_{ER}, 0.67 ± 0.02 for CTCF, 0.73 ± 0.01 for CTCF+D_{ER} and 0.69 ± 0.02 for CTCF+D_{ER}+D_{FoxA1}.

(A)



(B)

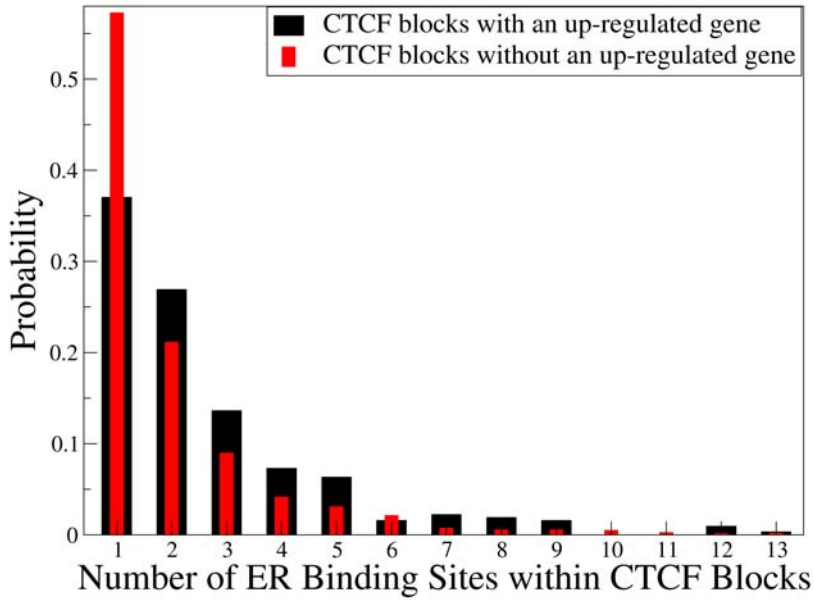


Figure S10: The effect of number of ER binding sites within CTCF blocks on expression fold-change. (A) The graph shows a scatter plot of expression fold-changes for up-regulated genes in Bourdeau *et al.* as a function of ER binding sites. There was only a weak correlation of 0.162 between fold-change and number of ER binding sites (P-value $= 1.706 \times 10^{-4}$) (B) #ER in CTCF blocks with ER and with differentially expressed genes $= 2.7 \pm 2.2$. #ER in CTCF blocks with ER but without differentially expressed genes $= 2.0 \pm 2.0$. p-value for the difference in their distributions is 3.7×10^{-6} using t-test (2.0×10^{-12} using Wilcoxon).

Supplemental Tables

Cluster Size	Number of Cluster	Number of CTCF Blocks
1	1471	1471
2	265	530
3	67	201
4	20	80
5	7	35
6	1	6
7	1	7
8	1	8

Table S1: Clustering of ER containing CTCF blocks. Most ER containing CTCF blocks are isolated. 2 CTCF blocks were considered to be in the same cluster if they are adjacent.

Cluster Size	#Genes in Cluster	#Clusters	Partition Type × #Clusters
1	1	184	(1)×184
1	2	13	(2)×24
2	1	60	(1)×60
2	2	9	(2)×8, (1,1)×1
2	3	2	(3)×2
2	5	2	(2,3)×2
3	1	13	(1)×14
3	2	6	(1,1)×4, (2)×1
4	1	4	(1)×4
4	2	3	(1,1)×2, (2)×1
4	3	1	(1,1,1)×1
5	1	1	(1)×1
5	2	1	(1,1)×1
7	1	1	(1)×1
8	4	1	(1,1,1,1)×1

Table S2: Partition types of up-regulated genes within clusters in Table S1. In a cluster containing N contiguous ER-blocks, we denote the distribution, or partition type, of k genes in the cluster as $(k_1, k_2, \ldots, k_N)$, where k_i = number of genes in i-th ER-block such that $\sum_{i=0}^{N} k_i = k$. To simplify notation, we omit 0 entries in the partition type. For example, $(2, 3)$ in "Cluster Size"=2 means 1 ER-block contained 2 up-regulated genes and 1 ER-block contained 3 up-regulated genes. It is seen that up-regulated genes are not concentrated into single ER-blocks, but they are instead distributed evenly across contiguous ER-blocks.