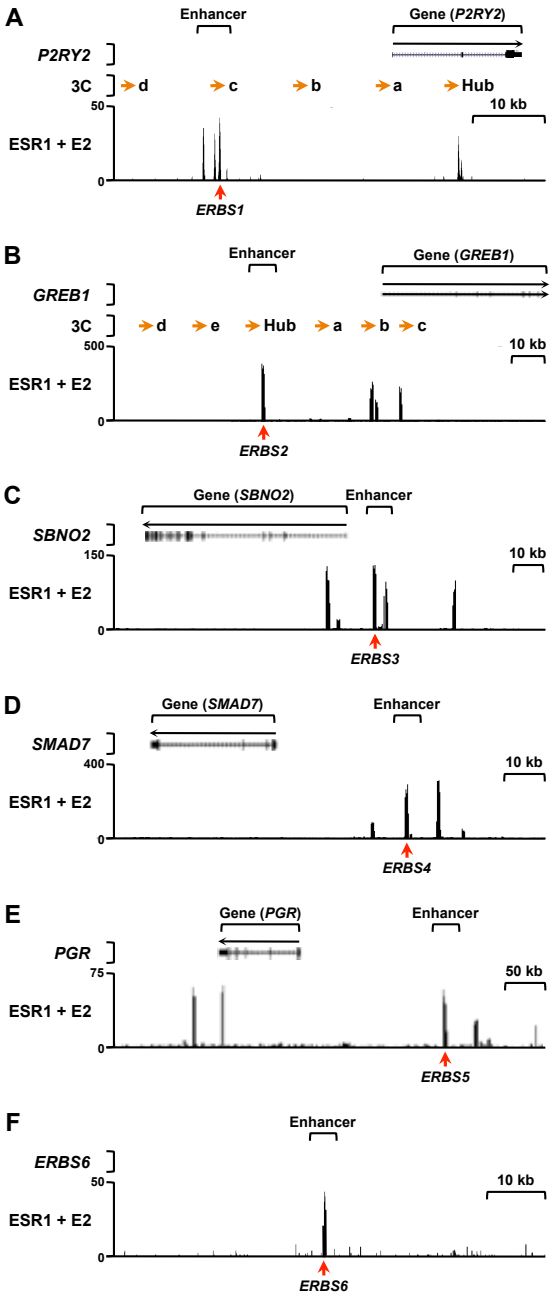
Supplemental Materials

"Enhancer Transcripts Mark Active Estrogen Receptor Binding Sites"

Hah *et al.* (2013)

This document contains the following supplemental information: Page

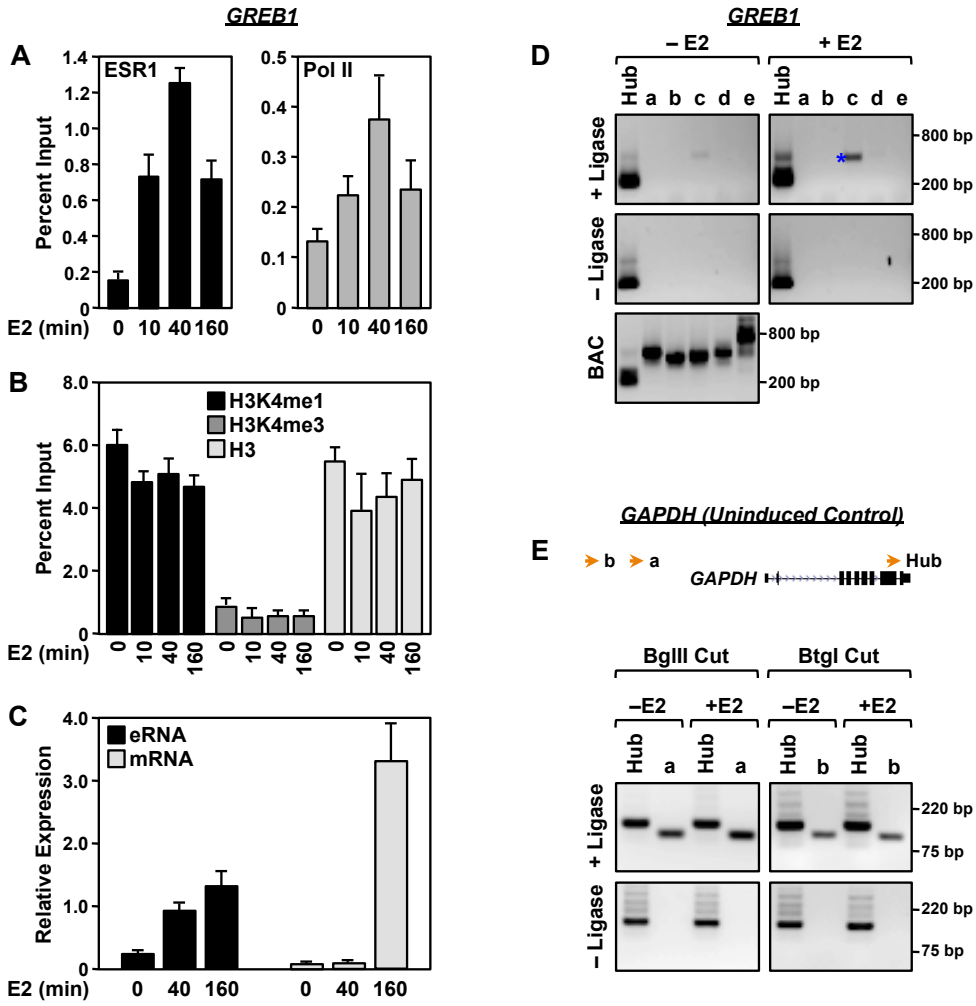
1. Supplemental Figures



Supplemental Figure 1. Schematics of genomic loci examined in detail in this study. Genome browser tracks showing gene annotations and ESR1 ChIP-seq data. Red arrows indicated the ERBSs that were examined by ChIP-qPCR. Orange arrows indicated the location of primer sets that were designed for the 3C analyses. The black arrows above the gene annotations indicate the direction of transcription for the annotated genes. Scale bars show the length of the indicated region. **(A)** *ERBS1* and the *P2RY2* gene. **(B)** *ERBS2* and the *GREB1* gene. **(C)** *ERBS3* and the *SBNO2* gene. **(D)** *ERBS4* and the *SMAD7* gene. **(E)** *ERBS5* and the *PGR* gene. **(F)** *ERBS6*, which was used as a control enhancer that does not produce eRNAs. *[Related to Fig. 1]*

For each of the following panels in Fig. 2,
jump to the page indicated or use the bookmarks to the left:

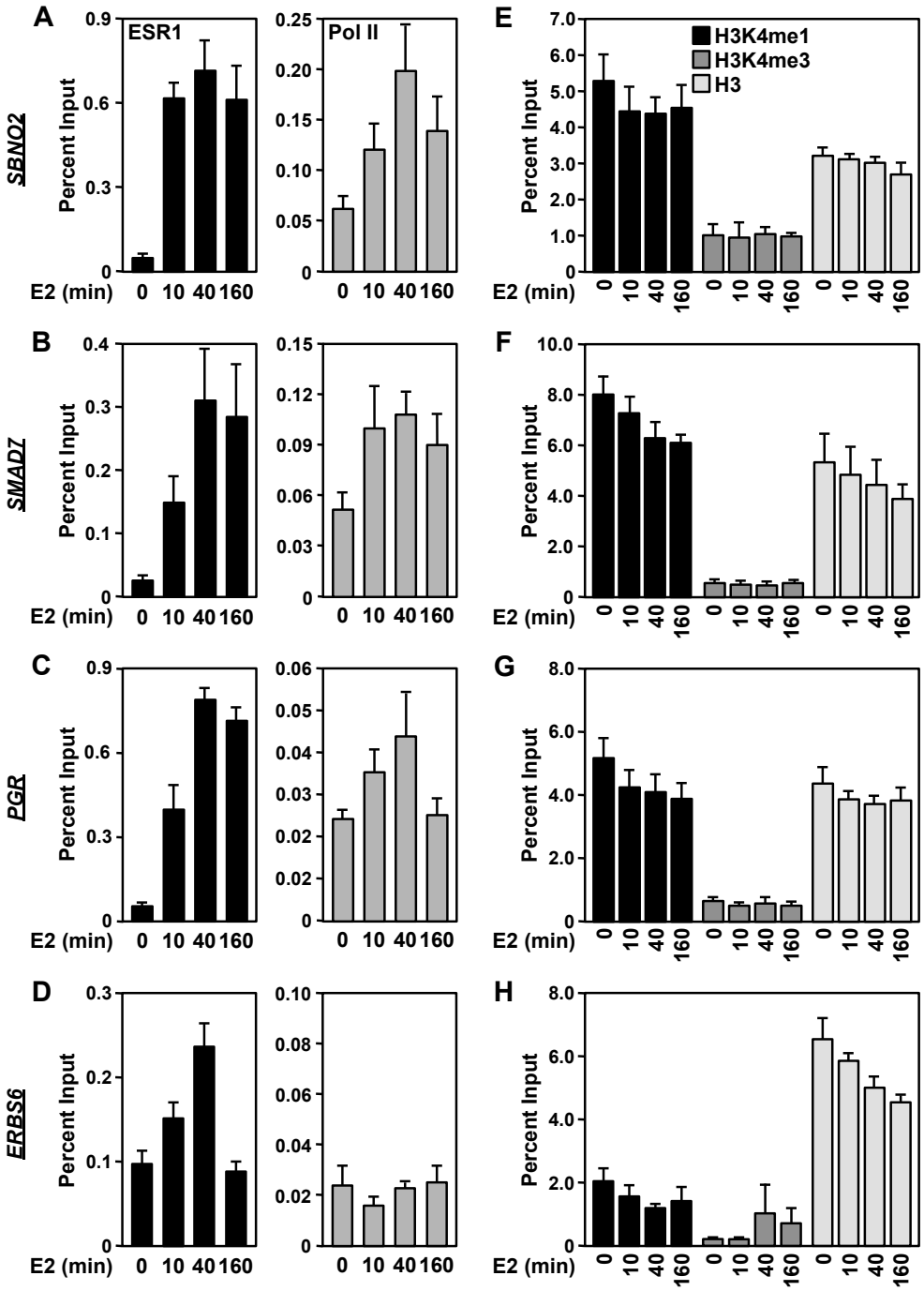
Supplemental Figure 2. Browser tracks showing genomic data for six loci with significant peaks of ESR1 binding.
Each panel shows data from 43 different genomic data sets for the six loci noted in the legend to Supplemental Fig. 1. The data labels shown on the left are descriptive names indicating the data set and conditions (U, untreated; E, E2-treated; 0, 10, 25, 40 min., time course of E2 treatment). The tracks are grouped according to the type of data and are color coded: (1) GRO-seq (GROSEQ), *red* = plus strand, *blue* = minus strand, four time points of E2 treatment; (2) Pol II ChIP-seq (POLII), *maroon*; (3) ESR1 (ER), *black*; (4) pioneer factors FOXA1 (FOXA1) and TFAP2C (AP2GAMMA), *purple*; (5) coactivators EP300, CREBBP, NCOA1, NCOA2, and NCOA3, *orange*; (6) histone modifications H3K4me1, H3K4me2, H3K4me3, H3K27me3, H3K9ac, and H3K14ac, *green*; (7) DNaseI and FAIRE, *gold*, and (8) ESR1 ChIA-PET, *black bars*. The black bars shown for the ChIA-PET data indicate the "heads" and "tails" that make contact in the gene loops, which are indicated by the dotted black lines. **(A)** *ERBS1* and the *P2RY2* gene. **(B)** *ERBS2* and the *GREB1* gene. **(C)** *ERBS3* and the *SBNO2* gene. **(D)** *ERBS4* and the *SMAD7* gene. **(E)** *ERBS5* and the *PGR* gene. **(F)** *ERBS6*, which was used as a control enhancer that does not produce eRNAs.
[Related to Fig. 1]



Supplemental Figure 3. Dynamics of ESR1 enhancer activity for the E2-responsive *GREB1* gene in MCF-7 cells.

Locus-specific molecular assays for E2-responsive enhancers. In panels A through C, each bar represents the mean + the SEM for three or more independent biological replicates. **(A)** ChIP-qPCR analyses for ESR1 and Pol II at a distal enhancer of the *GREB1* gene (*ERBS2*) in response to a time course of E2 treatment. **(B)** ChIP-qPCR analyses for H3K4me1, H3K4me3, and bulk H3 at ERBS2 in response to a time course of E2 treatment. **(C)** RT-qPCR analyses for *ERBS2* eRNA and *GREB1* mRNA in response to a time course of E2 treatment. **(D)** 3C-PCR assays for E2-induced looping between *ERBS2* and the *GREB1* gene. The lower case letters correspond to the primers denoted by orange arrows shown in Supplemental Fig. 1B. The assays were conducted in the presence (experimental) or absence (control) of DNA ligase, as indicated. Digested and ligated bacterial artificial chromosome (BAC) DNA spanning the entire *GREB1* locus was used as a PCR control. The size of the PCR fragments in bp is shown. One representative experiment from three conducted is shown. **(E)** Control 3C-PCR assay showing previously characterized E2-independent looping events at the *GAPDH* locus (Wang et al. 2009). The assay was performed twice each using different restriction enzymes: BglII (used for the *P2RY2* 3C assay) and BtgI (used for the *GREB1* 3C assay). As expected, the chromatin loops at the *GAPDH* locus occur independently of E2 treatment and are not regulated by E2.

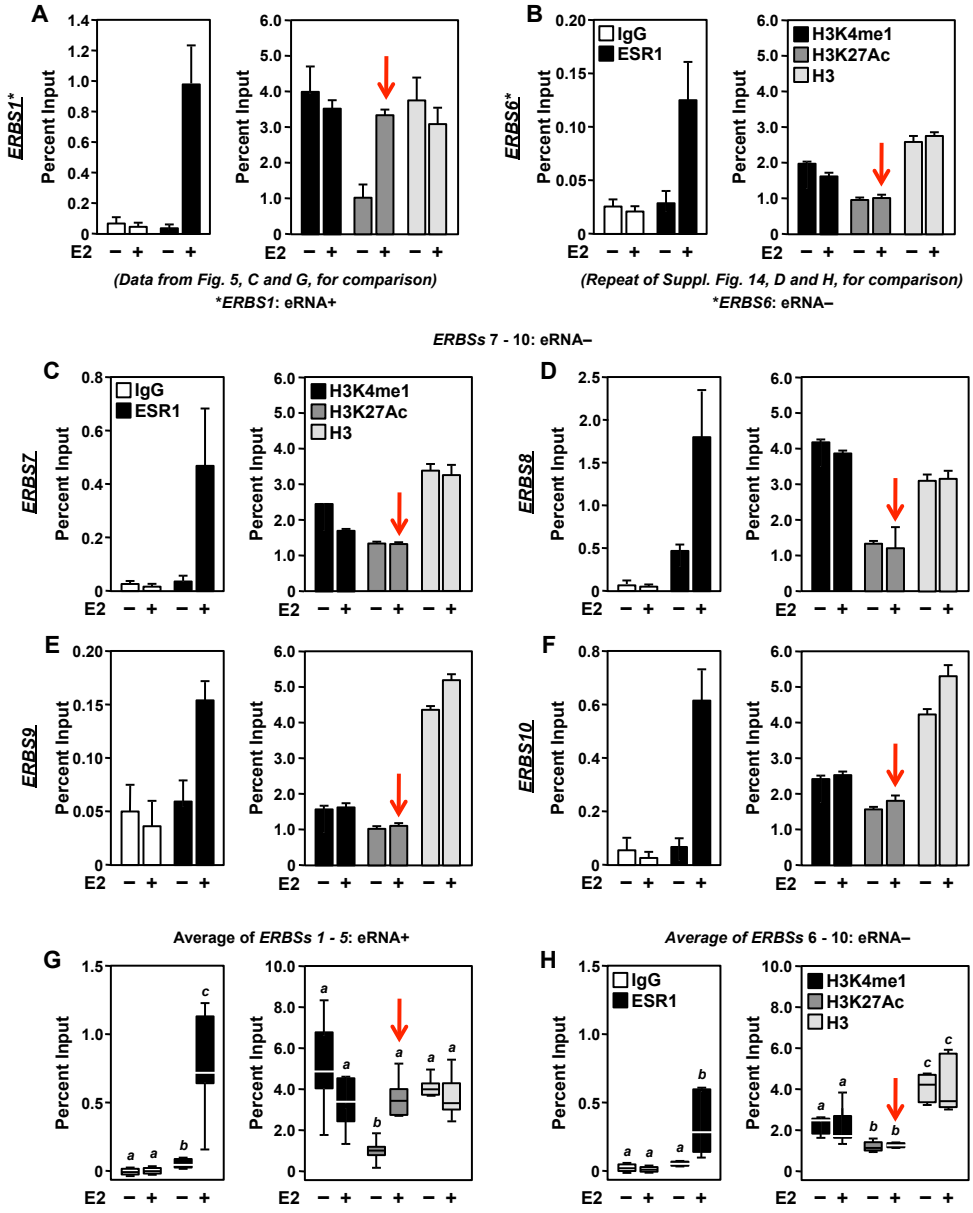
[Related to Fig. 1]



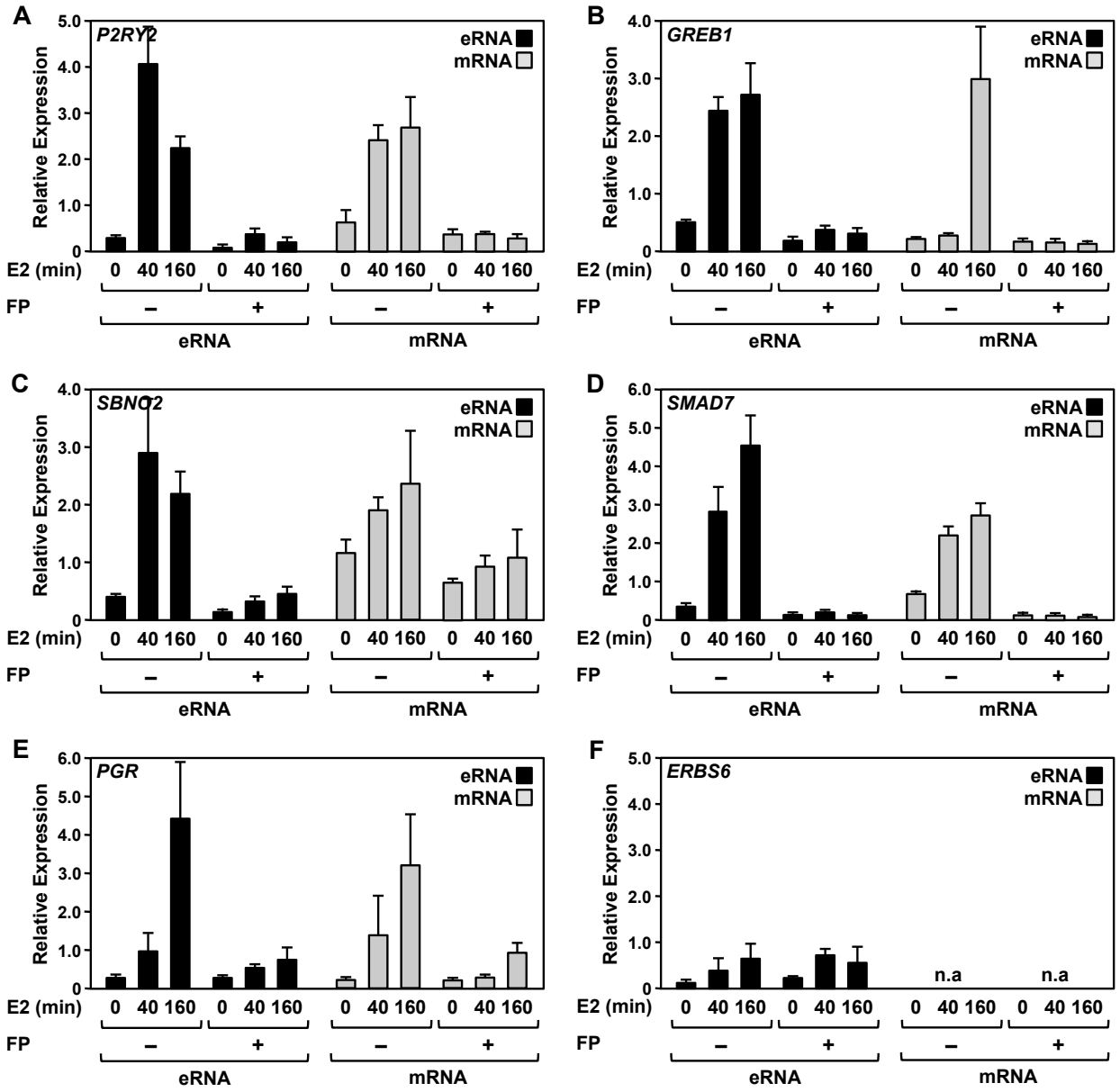
Supplemental Figure 4. Dynamics of ESR1 enhancer activity for three E2-responsive *loci* in MCF-7 cells.

Locus-specific ChIP-qPCR analyses for ESR1 (A-D, *left*), Pol II (A-D, *right*), and H3K4me1, H3K4me3, and bulk H3 (E-H) for four different ESR1 enhancers in response to a time course of E2 treatment. Each bar represents the mean + the SEM for three or more independent biological replicates. **(A, E)** *ERBS3* (distal enhancer of the *SBNO2* gene). **(B, F)** *ERBS4* (distal enhancer of the *SMAD7* gene). **(C, G)** *ERBS5* (distal enhancer of the *PGR* gene). **(D, H)** *ERBS6*, which was used as a control enhancer that does not produce eRNAs.

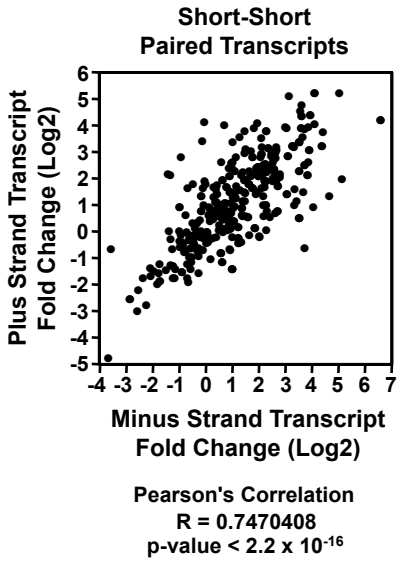
[Related to Fig. 1]



Supplemental Figure 5. H3K4me1 and H3K27ac levels at ERBSs with and without eRNAs. Locus-specific ChIP-qPCR analyses for ESR1, H3K4me1, H3K27ac, and bulk H3 for ERBSs that produce or do not produce eRNAs (as determined by GRO-seq) with or without E2 treatment, as indicated. In panels A though F, each bar represents the mean + the SEM for three or more independent biological replicates. **(A)** ChIP-qPCR analyses of *ERBS1*, an eRNA+ enhancer for the *P2RY2* gene. The data are from Fig. 5, C and G and are shown for comparison. **(B)** ChIP-qPCR analyses of *ERBS6*, an eRNA– ERBS. The experiment from Supplemental Fig. 14, D and H was repeated for comparison. (**C through F**) ChIP-qPCR analyses of *ERBSs 7* through *10*, a set of four eRNA– ERBSs. **(G and H)** Boxplots summarizing ChIP-qPCR analyses for five eRNA+ ERBSs (*ERBSs 1* through *5*) and five eRNA– ERBSs (*ERBSs 7* through *10*). The boxplots were generated from the average values obtained from each individual ERBS. Boxes marked with different letter superscripts are statistically different (Wilcoxon rank sum test; p-value < 0.05).
[Related to Fig. 1]

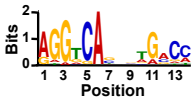
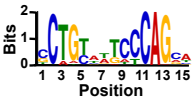


Supplemental Figure 6. Expression of E2-responsive eRNAs and mRNAs in response to a time course of E2 treatment in MCF-7 cells.
RT-qPCR assays for eRNAs and mRNAs from enhancer/gene pairs in response to a time course of E2 treatment in MCF-7 cells with (+) or without (-) a 1 hour pretreatment with flavopiridol (FP). Each bar represents the mean + the SEM for three or more independent biological replicates. **(A)** *ERBS1* and the *P2RY2* gene. **(B)** *ERBS2* and the *GREB1* gene. **(C)** *ERBS3* and the *SBNO2* gene. **(D)** *ERBS4* and the *SMAD7* gene. **(E)** *ERBS5* and the *PGR* gene. **(F)** *ERBS6*, which was used as a control enhancer that does not have eRNA transcripts.
[Related to Fig. 1]



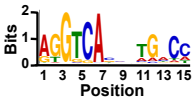
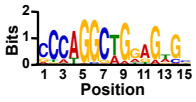
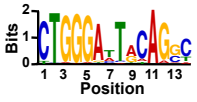
Supplemental Figure 7. Correlation between the E2-regulated transcription of short-short paired eRNAs from the plus and minus strands.
Pearson's correlation of fold changes in the transcription of transcripts originating from the plus and minus strand for short-short paired eRNAs. GRO-seq reads assigned to called eRNA transcripts were compared to their paired transcript.
[Related to Fig. 2]

A *ERBS with eRNA*

Motif	Pattern Found	Predicted Motif	No. of Sites*	% of sites*	MEME E-value
1		ESR1	873 *745*	55% *47%*	4.5×10^{-465}
2		SP1	822 *558*	52% *35%*	4.1×10^{-178}
3		STAT3	303 *n.a.*	19% *n.a.*	7.4×10^{-082}
4		BCD	110 *89*	7% *6%*	7.6×10^{-073}

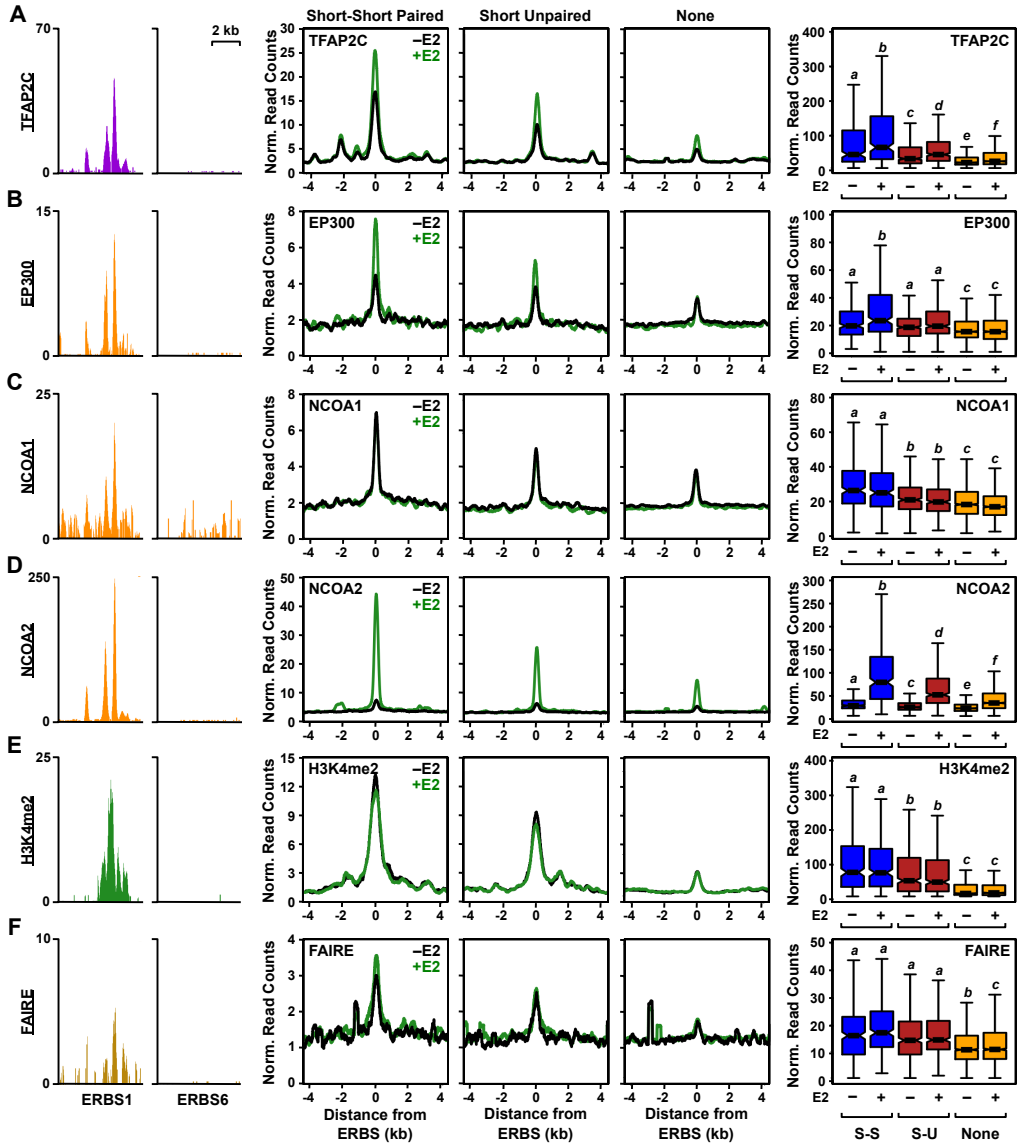
** Searched with the corresponding PWM derived in panel B*

B *ERBS without eRNA*

Motif	Pattern Found	Predicted Motif	No. of Sites*	% of Sites*	MEME E-value
1		ESR1	966 *931*	61% *58%*	3.6×10^{-742}
2		PAX6	139 *n.a.*	9% *n.a.*	4.0×10^{-105}
3		BCD	75 *140*	5% *9%*	3.7×10^{-090}
4		SP1	419 *613*	26% *39%*	4.1×10^{-082}

** Searched with the corresponding PWM derived in panel A*

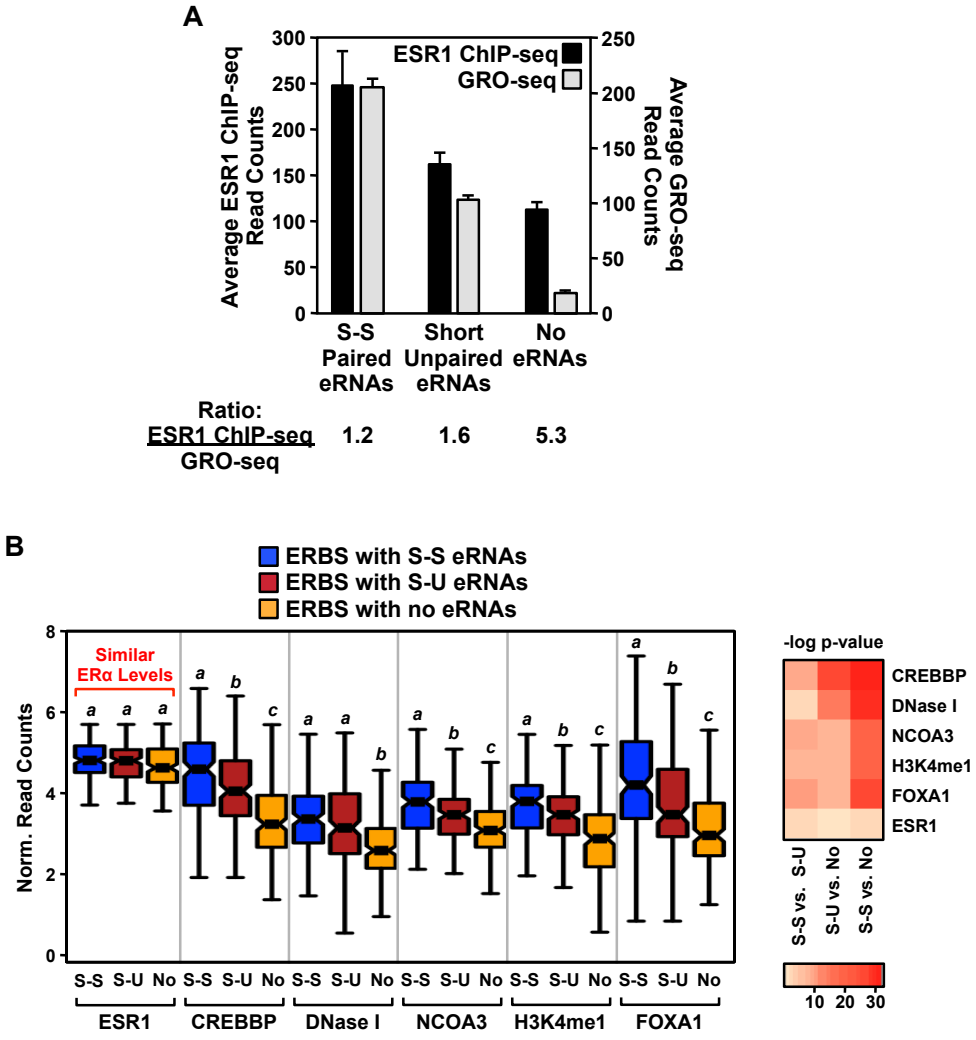
Supplemental Figure 8. Motif analyses of ERBSs associated with eRNAs.
For ERBSs with **(A)** and without **(B)** transcripts (1597 and 1594 respectively; see Fig. 2A), we performed de novo motif analyses on a 200 bp region around the center of the ESR1 peak (± 100 bp) using the command-line version of MEME. The predicted motifs from MEME were matched to known motifs using STAMP. The top four non-repetitive motifs are shown for each condition. The motifs in each condition are ordered based on the MEME E-values. In a subsequent analysis, the position weight matrices (PWMs) derived in panel A for ESR1, SP1, and BCD were used to search ERBSs without eRNA (data in panel B *highlighted in green*) and vice versa (data in panel A *highlighted in blue*). Although FOXA1 motifs were not strongly enriched in this de novo motif search, a directed motif search using three different PWMs for FOXA1 revealed that ~20% of ERBSs with or without eRNA (when using a 100 bp window on either side of the ERBS) or ~45% of ERBSs with or without eRNA (when using a 500 bp window on either side of the ERBS) have FOXA1 motifs.
[Related to Fig. 2]



Supplemental Figure 9. The production of eRNAs from ERBSs positively correlates with the recruitment of coactivators, the levels of histone modifications, and the chromatin state in MCF-7 cells.

Browser tracks, metaplots, and boxplots showing a positive correlation between eRNA production at ERBS with known markers of enhancer function. *(Left two panels)* Browser track representation of coactivator or histone modification ChIP-seq data, or FAIRE-seq data, as indicated on the y-axis for ERBS1 and ERBS6. *(Middle three panels)* Metaplot analyses of ChIP-seq or FAIRE-seq read counts for sets of ERBSs with short-short paired, short unpaired, or no transcripts in the presence *(green line)* or absence (*black line*) of E2 treatment. *(Right panel)* Box plot representations of ChIP-seq or FAIRE-seq data for sets of ERBSs with short-short paired *(blue boxes)*, short unpaired *(maroon boxes)*, or no transcripts *(yellow boxes)* in the presence (+) or absence (-) of E2 treatment. Boxes marked with different letter superscripts are statistically different (Wilcoxon rank sum test; p-value < 0.001). **(A)** TFAP2C ChIP-seq. **(B)** EP300 ChIP-seq. **(C)** NCOA1 ChIP-seq. **(D)** NCOA2 ChIP-seq. **(E)** H3K4me2 ChIP-seq. **(F)** FAIRE-seq.

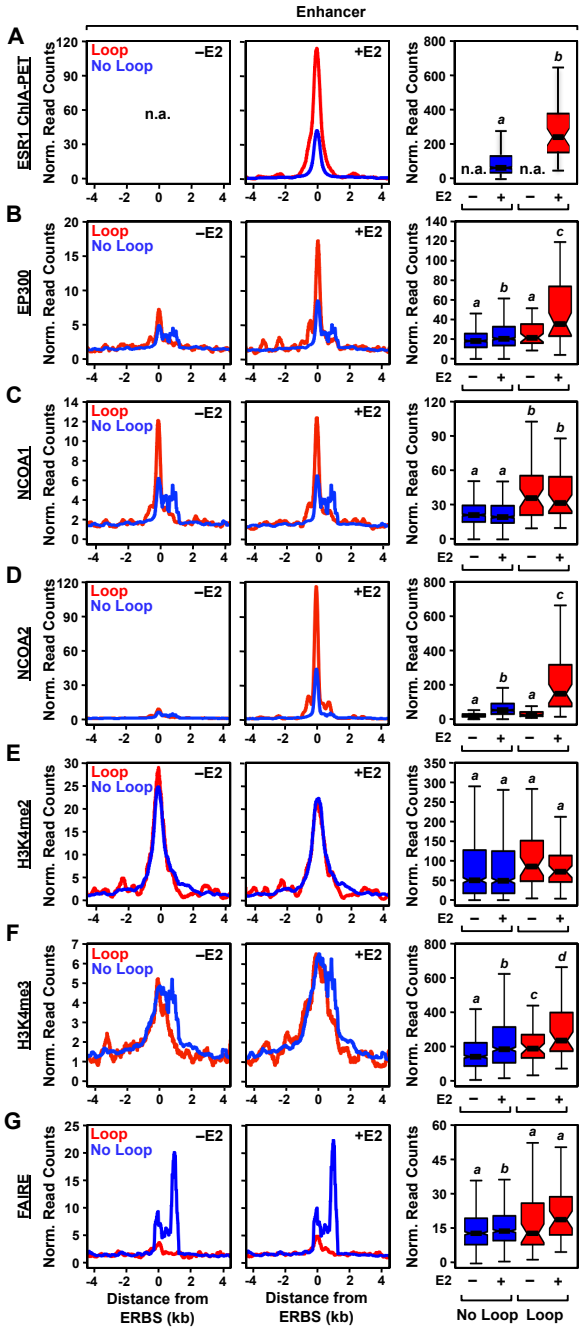
[Related to Fig. 3]



Supplemental Figure 10. The extent of ESR1 binding at enhancers does not account for the reduced GRO-seq signals or less pronounced enhancer features observed at ERBSs without eRNAs .

(A) Comparison of average ESR1 ChIP-seq and GRO-seq read counts for sets of ERBSs with short-short paired, short unpaired, or no transcripts in the presence of E2. **(B)** Data normalized for strength of ESR1 binding reveal major differences in enhancer features at ERBSs with short-short paired, short unpaired, or no transcripts in the presence of E2. ERBSs in the three groups (S-S, S-U and no transcripts) were randomly sampled to identify a subset of ERBSs with similar levels of ESR1 binding. A Kolmogorov-Smirnov test confirmed that the randomly sampled ERBSs in each group had similar distributions (not shown). The sampled ERBSs in each group were assayed for ESR1, CREBBP1, NCOA3, FOXA1, and H3K4me1 levels using ChIP-seq data sets, and DNase I hypersensitivity using a DNase-seq data set. The results are shown in boxplots. Those boxes marked with different letter superscripts are statistically different (Wilcoxon rank sum test; p-value < 0.001).

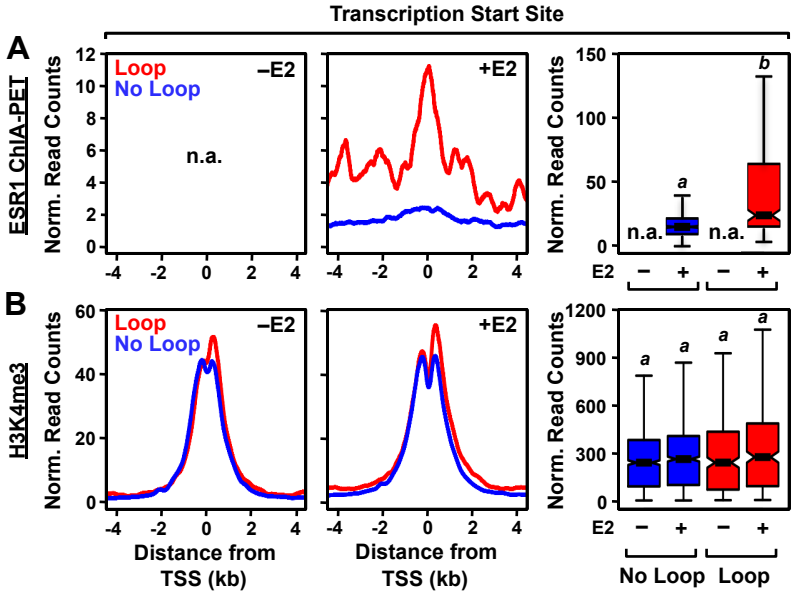
[Related to Fig. 3]



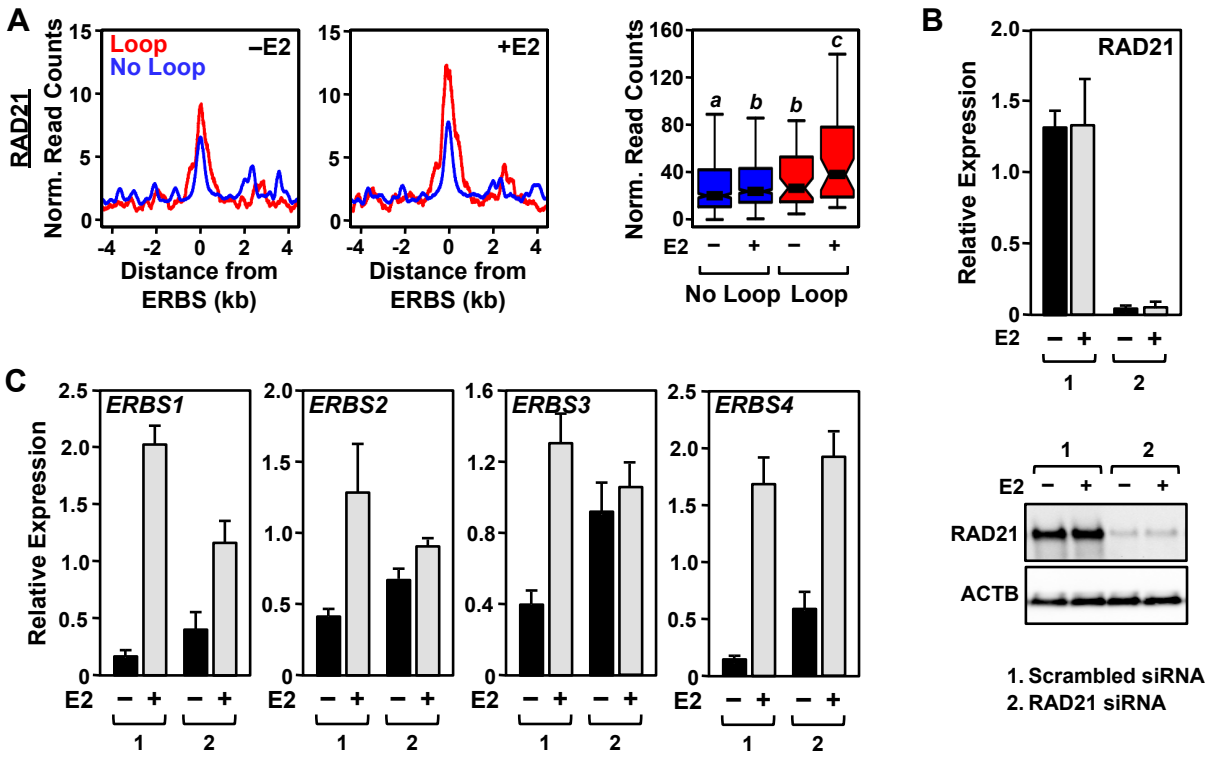
Supplemental Figure 11. The production of eRNAs from ERBSs correlates with enhancer looping to target genes.

Metaplots analyses and boxplots of genomic data, as indicated, for ERBSs that either loop to (Loop, *red lines*) or do not loop to (No Loop, *blue lines*) target gene promoters. ERBSs were queried to determine if they loop to the promoters of RefSeq genes, based on ESR1 ChIA-PET data. Looping was assayed within a 2 kb window (± 1 kb) around ESR1 peak centers and a 10 kb window (± 5 kb) around the TSSs of target genes. Boxes in the boxplots marked with different letter superscripts are statistically different (Wilcoxon rank sum test; p-value < 0.001). **(A)** ChIA-PET. **(B)** EP300. **(C)** NCOA1. **(D)** NCOA2. **(E)** H3K4me2. **(F)** H3K4me3. **(G)** FAIRE.

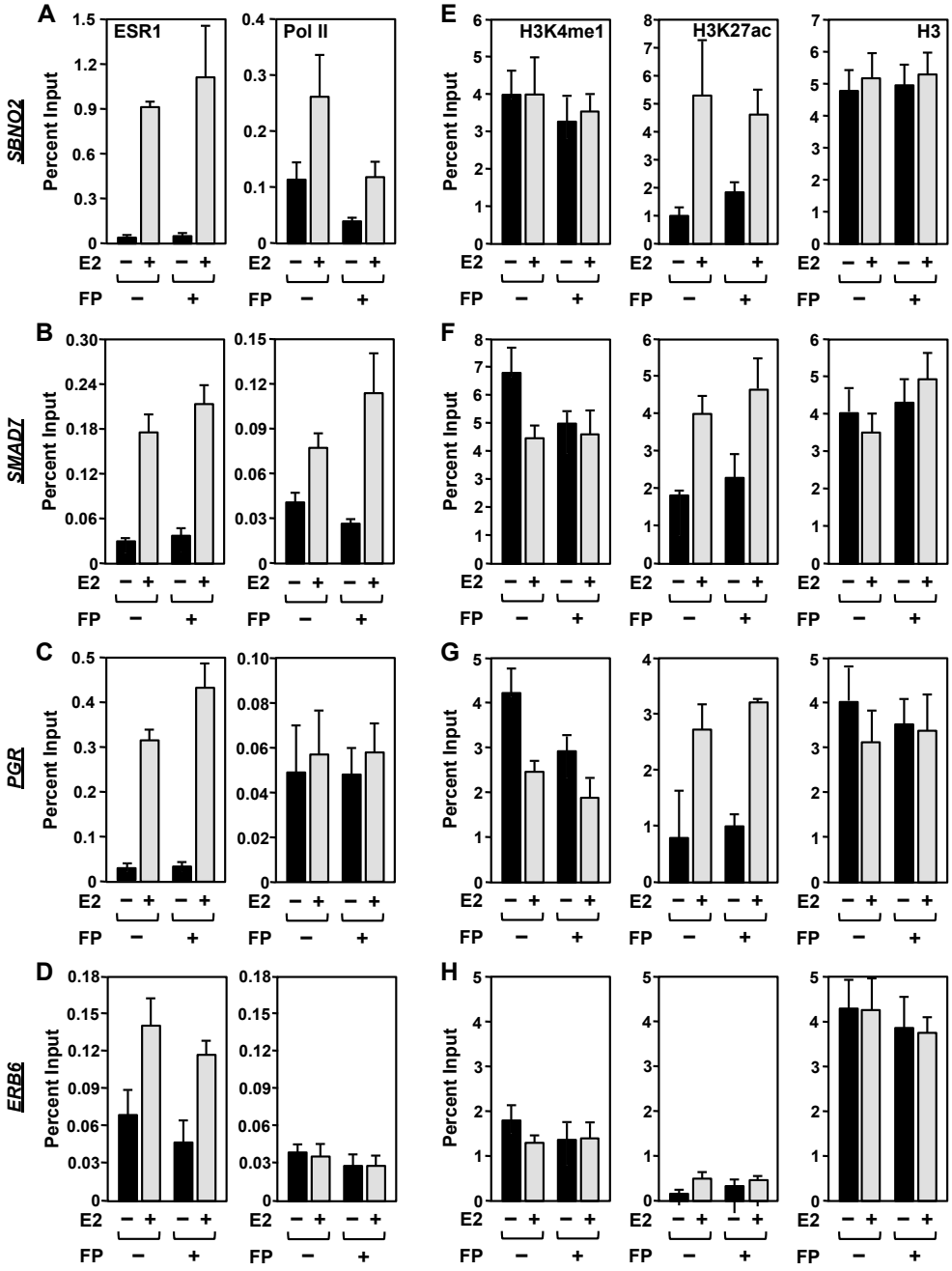
[Related to Fig. 4]



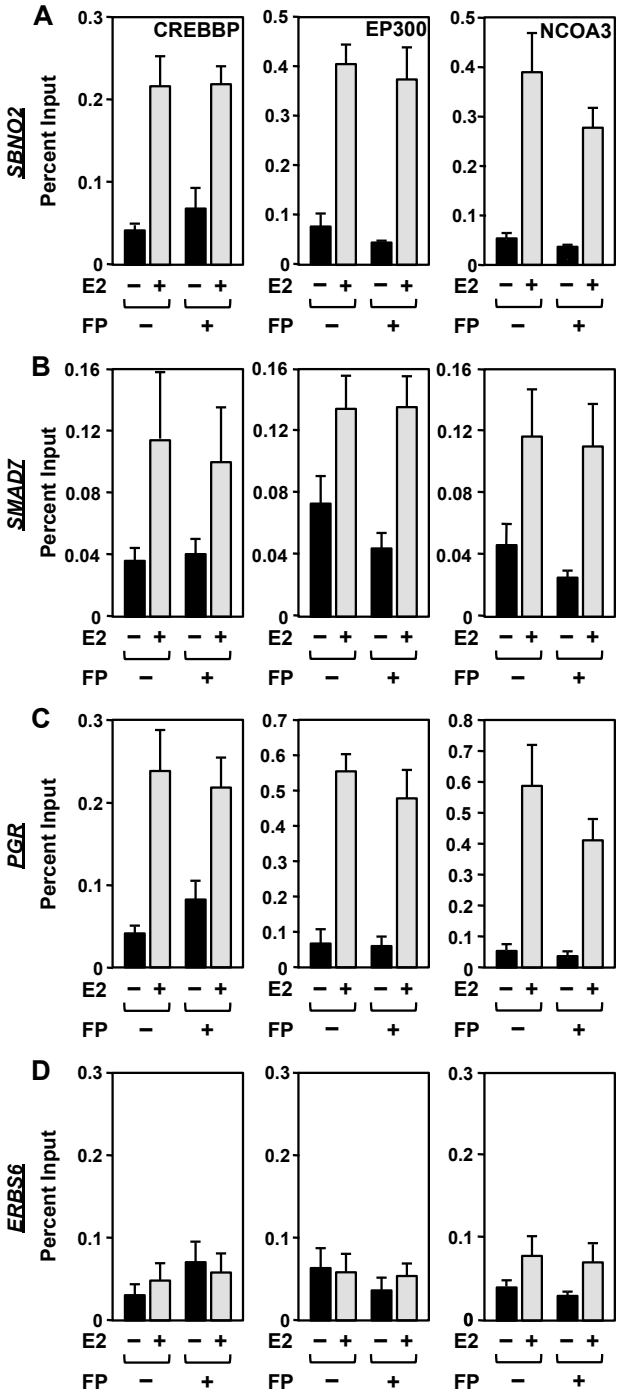
Supplemental Figure 12. ChIA-PET and H3K4me3 levels at target gene promoters that are either looped to or are not looped to from distal ERBSs.
Metaplot analyses and boxplots of genomic data, as indicated, for target gene promoters that are either looped to (Loop, *red lines*) or are not looped to (No Loop, *blue lines*) from distal ERBSs. Promoters were queried to determine if they are looped to from distal ERBSs, based on ESR1 ChIA-PET data. Looping was assayed within a 2 kb window (± 1 kb) around ESR1 peak centers and a 10 kb window (± 5 kb) around the TSSs of target genes. Boxes in the boxplots marked with different letter superscripts are statistically different (Wilcoxon rank sum test; p-value < 0.001). **(A)** ChIA-PET. **(B)** H3K4me3.
[Related to Fig. 4]



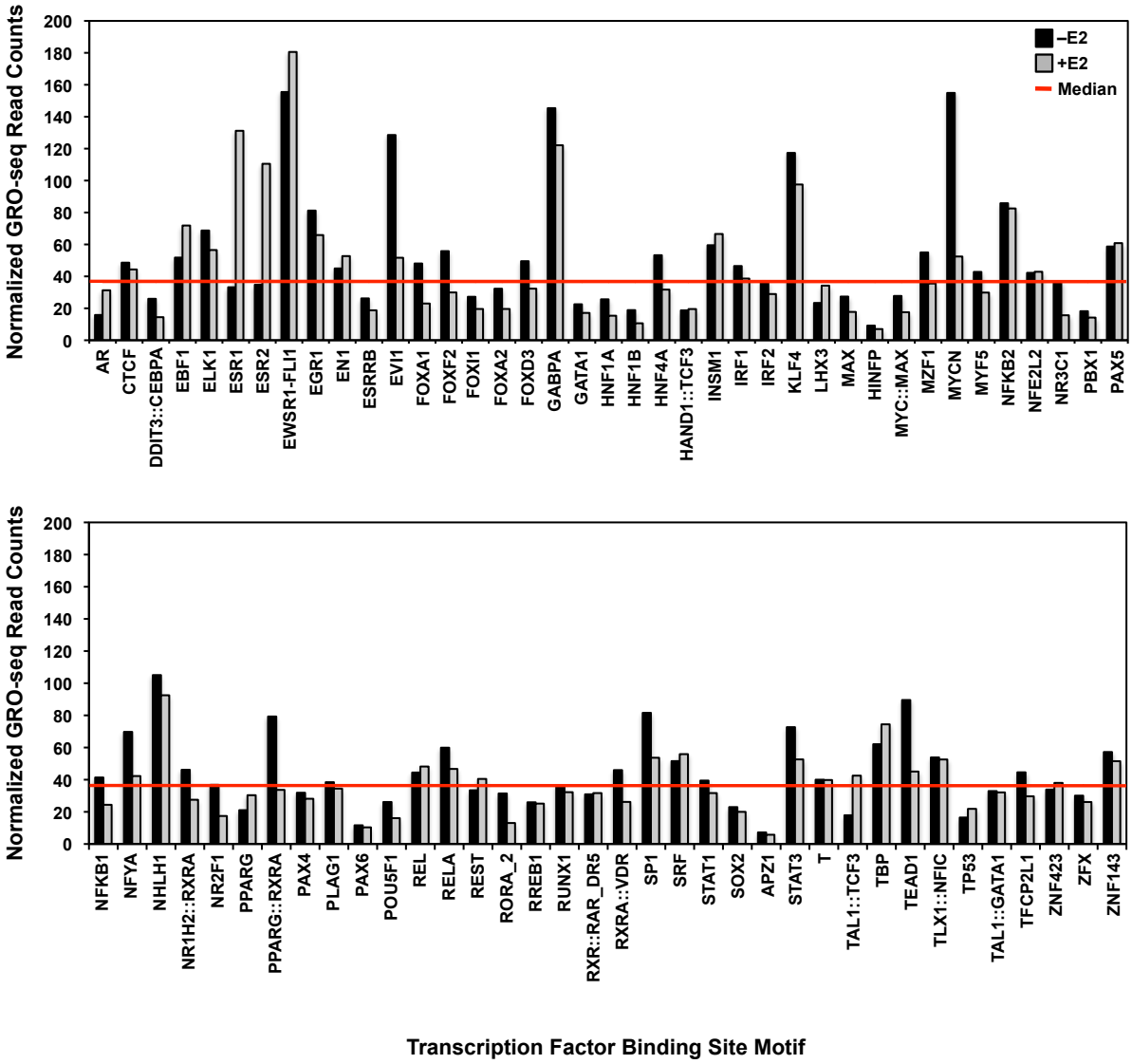
Supplemental Figure 13. RAD21 is enriched at ERBSs that loop to target gene promoters, and suppresses basal and enhances E2-induced eRNA production at some ERBSs.
(A) Metaplot analyses and boxplots of RAD21 ChIP-seq data for ERBSs that either loop to (Loop, *red lines*) or do not loop to (No Loop, *blue lines*) target gene promoters. ERBSs were queried to determine if they are looped to from distal ERBSs, based on ESR1 ChIA-PET data. Looping was assayed within a 2 kb window (± 1 kb) around ESR1 peak centers and a 10 kb window (± 5 kb) around the TSSs of target genes. Boxes in the boxplots marked with different letter superscripts are statistically different (Wilcoxon rank sum test; p-value < 0.001). **(B)** siRNA-mediated knockdown of RAD21. MCF-7 cells were transfected with an siRNA pool directed against RAD21 or a control ("scrambled") siRNA pool. The cells were collected and assayed for *RAD21* mRNA by RT-qPCR (*top*) and RAD21 protein by Western blotting (*bottom*). Each bar represents the mean + SEM from three or more independent biological replicates. **(C)** Effect of RAD21 knockdown on eRNA production from four ERBSs (*ERBS1* for *P2RY2*; *ERBS2* for *GREB1*; *ERBS3* for *SBNO2*; *ERBS4* for *SMAD7*) as assayed by RT-qPCR. Each bar represents the mean + SEM.
[Related to Fig. 4]



Supplemental Figure 14. Inhibition of eRNA production by flavopiridol does not inhibit ESR1 or Pol II binding, or alter H3K4me1 or H3K27ac levels, at ERBSs.
Locus-specific ChIP-qPCR analyses for ESR1 (A-D, *left*), Pol II (A-D, *right*), and H3K4me1, H3K4me3, and bulk H3 (E-H) for four different ESR1 enhancers in the absence or presence of E2 and flavopiridol (FP), as indicated. Each bar represents the mean + the SEM for three or more independent biological replicates. **(A, E)** *ERBS3* (distal enhancer of the *SBNO2* gene). **(B, F)** *ERBS4* (distal enhancer of the *SMAD7* gene). **(C, G)** *ERBS5* (distal enhancer of the *PGR* gene). **(D, H)** *ERBS6*, which was used as a control enhancer that does not have eRNA transcripts.
[Related to Fig. 5]



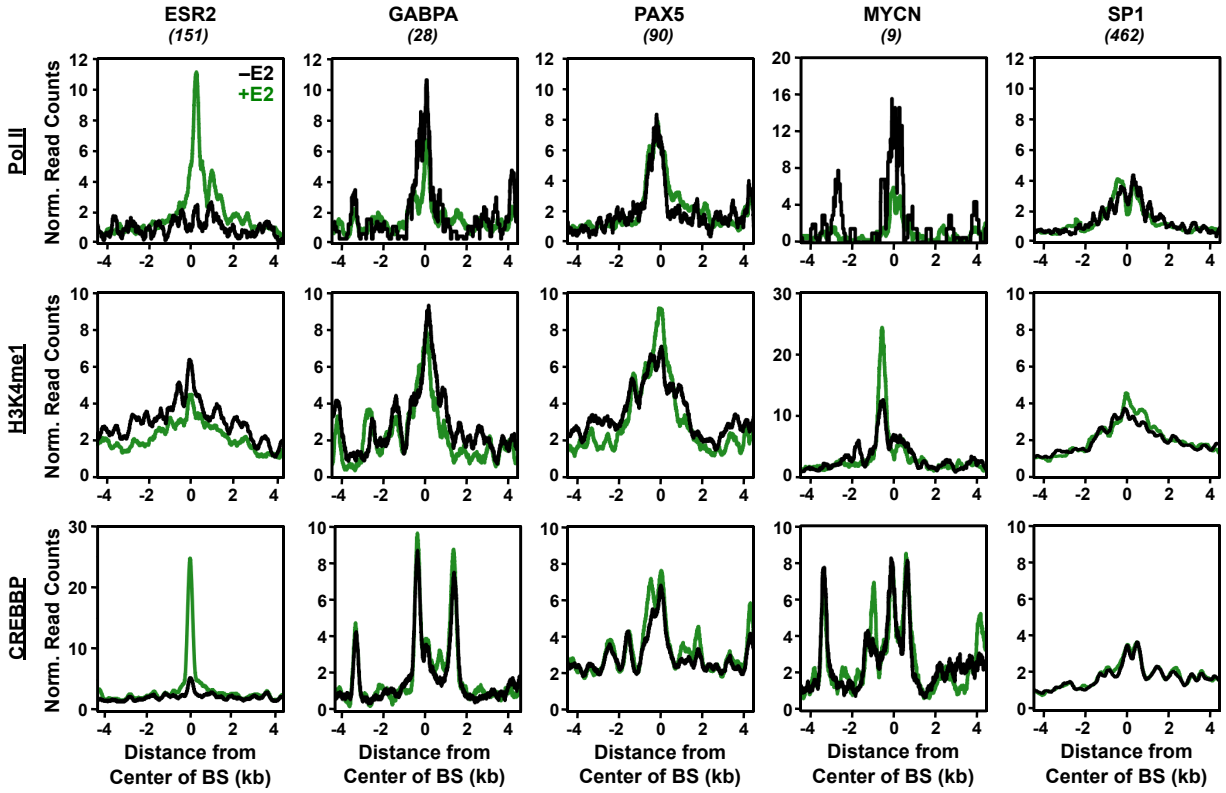
Supplemental Figure 15. Inhibition of eRNA production by flavopiridol does not inhibit CREBBP, EP300, or NCOA3 binding at ERBSs.
Locus-specific ChIP-qPCR analyses for CREBBP *(left)*, EP300 *(middle)*, and NCOA3 *(right)* for four different ESR1 enhancers in the absence or presence of E2 and flavopiridol (FP), as indicated. Each bar represents the mean + the SEM for three or more independent biological replicates. **(A)** *ERBS3* (distal enhancer of the *SBNO2* gene). **(B)** *ERBS4* (distal enhancer of the *SMAD7* gene). **(C)** *ERBS5* (distal enhancer of the *PGR* gene). **(D)** *ERBS6*, which was used as a control enhancer that does not have eRNA transcripts.
[Related to Fig. 5]



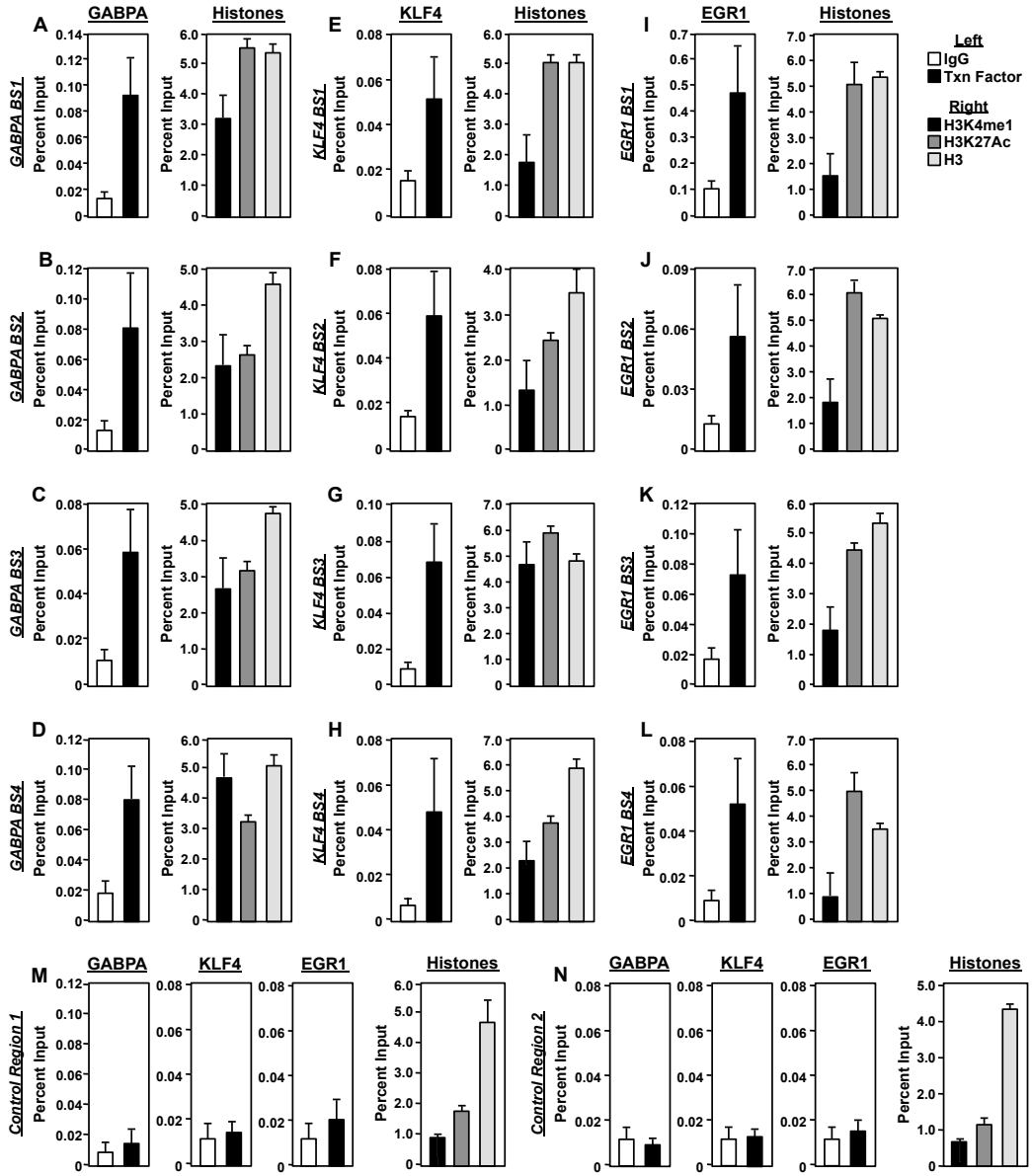
Supplemental Figure 16. Directed search for ESR1 and non-ESR1 enhancers in MCF-7 cells using GRO-seq data.

Bar graph showing the normalized GRO-seq read count density per occurrence for 75 motifs from the JASPAR database ±E2. The motifs were mapped to the human genome using FIMO. For all intergenic motifs (>10 kb from RefSeq genes) with eRNAs (either short-short paired or short unpaired) originating within a 2 kb window around the center of the motif (i.e., ± 1 kb relative to the motif), we collected the GRO-seq reads within 1 kb window around the center of the motif (i.e., ± 0.5 kb relative to the motif) and normalized them to the total number of occurrences of the motif.

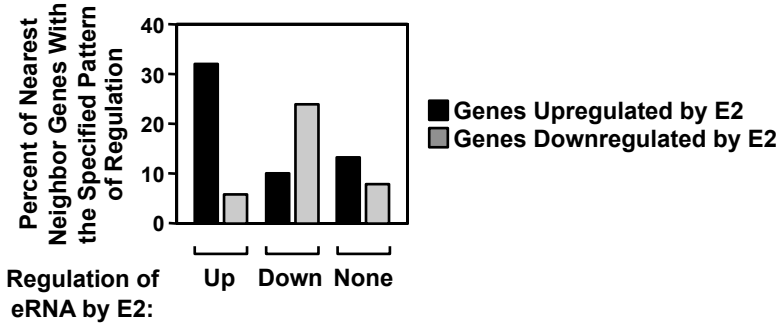
[Related to Fig. 6]



Supplemental Figure 17. Metaplot analyses of Pol II, H3K4me1, and CREBBP ChIP-seq data for five putative enhancer types selected based on a directed motif search.
Metaplot analyses for the ChIP-seq tag counts for Pol II, H3K4me1, and CREBBP for five putative enhancer types selected from the analyses shown in Fig. 6 and Supplemental Fig. 16. *[Related to Fig. 6]*



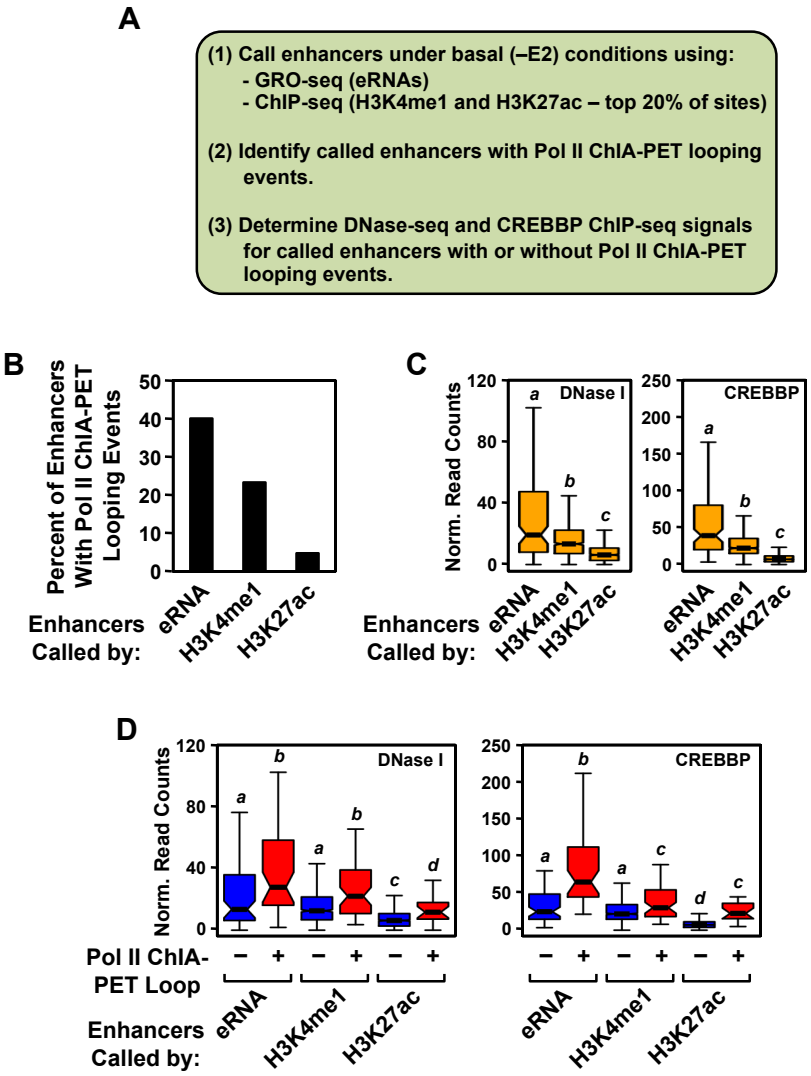
Supplemental Figure 18. GABPA, KLF4, and EGR1 bind to their cognate predicted enhancers, but not control regions lacking predicted binding sites.
ChIP-qPCR assays of GABPA, KLF4, and EGR1 binding (*left* in each panel) and enhancer-associated histone modifications (i.e., H3K4me1 and H3K27ac; *right* in each panel) at cognate predicted binding sites, as well as two control regions. Each bar is the mean ± SEM for three or more independent determinations. **(A through D)** ChIP-qPCR assays of GABPA, H3K4me1, and H3K27ac at predicted GABPA binding sites (*GABPA BSs*) 1 through 4. **(E through H)** ChIP-qPCR assays of KLF4, H3K4me1, and H3K27ac at predicted KLF4 binding sites (*KLF4 BSs*) 1 through 4. The data in panel E, which are also presented in Fig. 6E, are shown for comparison. **(I through L)** ChIP-qPCR assays of EGR1, H3K4me1, and H3K27ac at predicted EGR1 binding sites (*EGR1 BSs*) 1 through 4. The data in panel I, which are also presented in Fig. 6F, are shown for comparison. **(M and N)** ChIP-qPCR assays of GABPA, KLF4, EGR1, H3K4me1, and H3K27ac at two control regions (1 and 2).
[Related to Fig. 6]



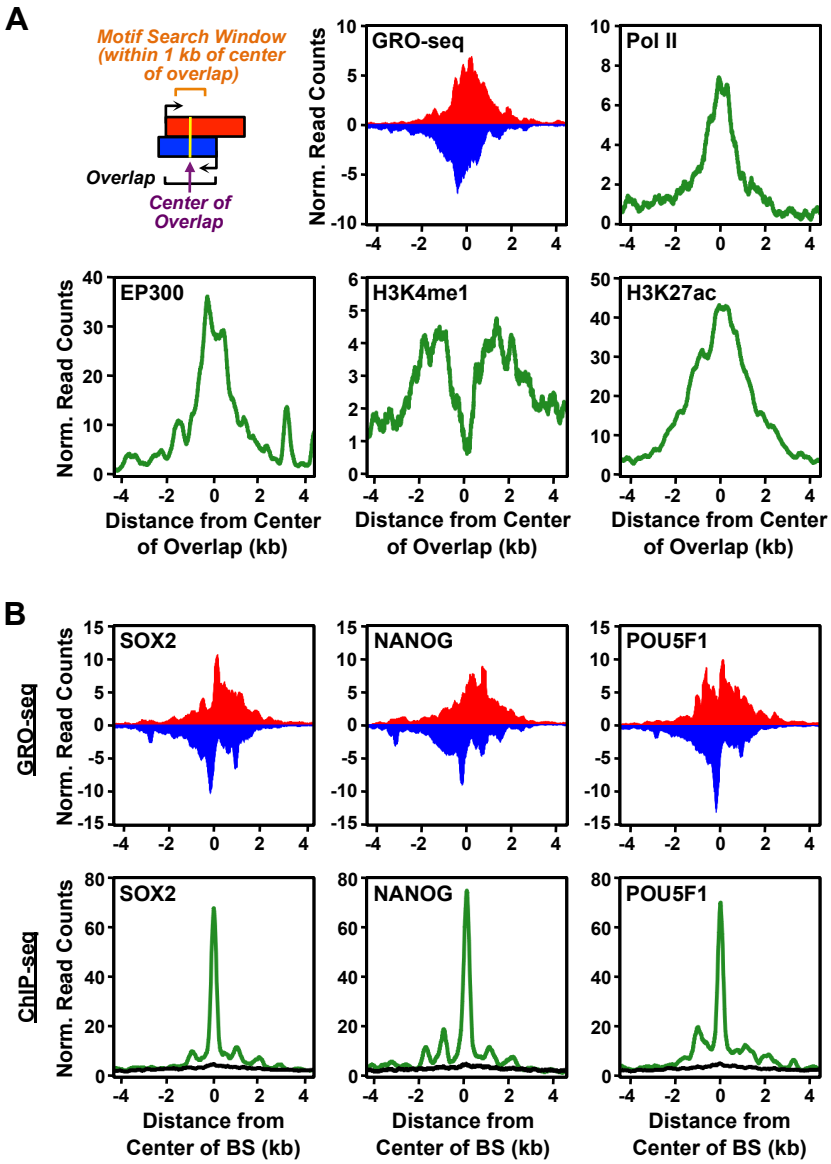
Supplemental Figure 19. eRNAs produced from predicted enhancers show similar patterns of E2-dependent transcriptional regulation as mRNAs produced from putative (nearest neighbor) target genes.

The nearest neighbor gene for each predicted enhancer (i.e., intergenic S-S paired eRNA-producing; 569 total) was identified. The analysis was limited to the nearest neighboring genes showing regulation of mRNA production by E2. The data were expressed as the percent of nearest neighbor genes with the specified pattern of regulation by E2 [i.e., up, down, unregulated ("none")]. For each pattern of regulation, the strength of association between the regulation of the enhancer and the regulation of nearest neighbor gene is statistically significant (Fisher's exact test; $p < 0.0001$).

[Related to Fig. 7]



Supplemental Figure 20. Enhancers called by GRO-seq have a greater frequency of Pol II-mediated gene looping events than enhancers called by H3K4me1 and H3K27ac ChIP-seq. **(A)** Outline of the analysis comparing enhancer prediction using GRO-seq versus H3K4me1 and H3K27ac ChIP-seq. We predicted enhancers in MCF-7 cells grown under basal (i.e., -E2) conditions using GRO-seq, as described above, or H3K4me1 and H3K27ac ChIP-seq. For the ChIP-seq data, enhancers were predicted by peak calling using the MACS algorithm with a p-value threshold of 1 x 10^{-10} and peaks >10 kb from the 5' or 3' ends of annotated genes. Only the top 20% of the enhancers based on the ChIP-seq calls were used for further analysis. The enhancers predicted by GRO-seq (i.e., eRNA +) and histone marks were then assayed for looping events using Pol II ChIA-PET data. **(B)** Enhancers called by GRO-seq have a greater frequency of Pol II-mediated gene looping events than enhancers called by H3K4me1 and H3K27ac ChIP-seq. **(C)** Enhancers called by GRO-seq show greater DNase I hypersensitivity and have a greater enrichment of CREBBP than enhancers called by H3K4me1 and H3K27ac ChIP-seq. **(D)** Pol II-dependent looped enhancers called by GRO-seq show greater DNase I hypersensitivity and have a greater enrichment of CREBBP than Pol II-dependent looped enhancers called by H3K4me1 and H3K27ac ChIP-seq.
[Related to Fig. 6]



Supplemental Figure 21. Unbiased identification of enhancers in mouse ES cells using GRO-seq data.

(A) *(Upper left panel)* Schematic representation of the unbiased enhancer search pipeline for mouse ES cells. Transcripts were called from GRO-seq data and all intergenic short-short paired eRNAs were identified. *(All other panels)* GRO-seq data and Pol II, EP300, H3K4me1, and H3K27ac ChIP-seq data were collected, mapped relative to the center of the plus and minus strand overlap of the short-short paired eRNAs, and expressed as metaplots. **(B)** GRO-seq and ChIP-seq metaplot analyses for SOX2, NANOG, POU5F1 binding sites overlapping enhancers predicted using GRO-seq. NANOG, POU5F1, and SOX2 peaks were called from ChIP-seq data using MACS. ChIP-seq peaks within a 2 kb window (± 1 kb) around the center of the plus and minus strand overlap of the predicted SS paired eRNAs were identified and metaplots of the ChIP-seq reads were plotted.

[Related to Fig. 7]

2. Supplemental Tables

Supplemental Table 1. Existing genomic data sets used for data mining.

A. ChIP-seq and other deep sequencing genomic data sets from MCF-7 cells.
see: (Fullwood et al. 2009; Welboren et al. 2009; Joseph et al. 2010; Schmidt et al. 2010; Kong et al. 2011; Tan et al. 2011; Yamashita et al. 2011; Zwart et al. 2011; Frietze et al. 2012; He et al. 2012; Li et al. 2012)

Data Set	NCBI GEO Accession numbers (or other access*)
• POLR2A (Pol II) (-E2, +E2)	GSM365929, GSM365930
• ESR1 (-E2, +E2)	GSM365925, GSM365926
• FOXA1 (-E2, +E2)	GSM588929, GSM588930
• TFAP2C (-E2, +E2)	GSM588928, GSM588927
• EP300 (-E2, +E2)	GSM720425, GSM720424
• CREBBP (-E2, +E2)	ERR045723, ERR045724
• NCOA1 (-E2, +E2)	ERR045725, ERR045726, ERR045729, ERR045730
• NCOA2 (-E2, +E2)	ERR045717, ERR045718, ERR045722, ERR045721
• NCOA3 (-E2, +E2)	ERR045716, ERR045715, ERR045719, ERR045720
• H3K4me1 (-E2, +E2)	GSM588569, GSM588568
• H3K4me2 (-E2, +E2)	GSM822391, GSM822392
• H3K4me3 (-E2, +E2)	GSM588571, GSM588570
• H3K27me3 (-E2, +E2)	GSM588565, GSM588564
• H3K9ac (-E2, +E2)	GSM588573, GSM588572
• H3K14ac (-E2, +E2)	GSM588575, GSM588574
• H3K27ac (-E2)	GSM946850
• CTCF (+E2)	GSM614614, GSM614615
• RAD21 (Cohesin)	GSM614618, GSM614619, GSM614612, GSM614613
• ESR1 ChIA-PET (+E2)	ChIA-PET visualization browser (http://cms1.gis.a-star.edu.sg)*
• Pol II ChIA-PET (-E2)	GSM832458, GSM832459
• DNase1 (-E2, +E2)	GSM822389, GSM822390
• FAIRE (-E2, +E2)	GSM58856, GSM588562
• TSS	SRA003625 (GenBank/DDBJ)*

B. ChIP-seq data sets from mouse embryonic stem cells.
see: (Marson et al. 2008; Shen et al. 2012)

Data Set	NCBI GEO Accession numbers
• NANOG	GSE11724
• SOX2	GSE11724
• POU5F1	GSE11724
• POLR2A (Pol II)	GSE29218
• EP300	GSE29218
• H3K4me1	GSE29218
• H3K27ac	GSE29218

Supplemental Table 2. Primers used for PCR-based assays.

A. Primers used for RT-qPCR

ERBS/Gene	Name	Sequence (5' to 3')
• *ERBS1/P2RY2*	eRNA-Fwd	AGGCAAATCCATTGTCATCC
	eRNA-Rev	AACTGGCTGGATCTTGAAGC
	mRNA-Fwd	CGGTGGACTTAGCTCTGAGG
	mRNA-Rev	GCCTCCAGATGGGTCTATGA
• *ERBS2/GREB1*	eRNA-Fwd	GGGAATAGAGCCCTGAGCTT
	eRNA-Rev	TTGATCTGCTCTTGCCTGAA
	mRNA-Fwd	CCTATTTTGGAATAAAAACTGACC
	mRNA-Rev	GGGGAGAATGACACAAAAGC
• *ERBS3/SBNO2*	eRNA-Fwd	CCTGTATTCTGGGGGCACTA
	eRNA-Rev	CTCACCCCATCCAGTACACC
	mRNA-Fwd	GACTGGGCACCCACAAGGGC
	mRNA-Rev	GGAAGGGCTGGGGGAGGGAG
• *ERBS4/SMAD7*	eRNA-Fwd	GGCATAGCTAGGACCTCACC
	eRNA-Rev	GAGGGAGGAAAGTGGCTTCT
	mRNA-Fwd	AAGAGAAGCATTCTCATTGGAAA
	mRNA-Rev	TCAGGAGTCCTTTCTCTCTCAAA
• *ERBS5/PGR*	eRNA-Fwd	ATGCAGAGCCATTGCAAAAT
	eRNA-Rev	ATCAGCAAGATGCAAACACG
	mRNA-Fwd	TTGCCAAGAAGGTGAAACTG
	mRNA-Rev	CTTTGCATTGTCACCCCATC
• *ERBS6*	ERBS6-Fwd	TGTGGAAGCTGCATTCTTTG
	ERBS6-Rev	TCAGAACCATGCAGAACCTG
• *RAD21*	mRNA-Fwd	CAATGCCAACCATGACTGAT
	mRNA-Rev	CGGTGTAAGACAGCGTGTAAA
• *SRC3*	mRNA-Fwd	TTCAGGAAAGGTTGTCAATATAGATACA
	mRNA-Rev	AATACACCTTCGGATTATATCTTCAAA
• *GABPA*	mRNA-Fwd	TTAAACCTGCGGACACTGTTG
	mRNA-Rev	TTAAACCTGCGGACACTGTTG
• *ACTB*	mRNA-Fwd	AGCTACGAGCTGCCTGAC
	mRNA-Rev	AAGGTAGTTTCGTGGATGC

B. Primers used for ChIP-qPCR

Gene/ERBS	Name	Sequence (5' to 3')
• *ERBS1/P2RY2*	ERBS1-Fwd	CCATCAAAGCTGTTGCTTCT
	ERBS1-Rev	CCAGGATAGTGCCAGTGAAC
• *ERBS2/GREB1*	ERBS2-Fwd	TAGGCTTCAAGAGGACCACA
	ERBS2-Rev	AGCAGCAAAACTGCATAGGA
• *ERBS3/SBNO2*	ERBS3-Fwd	GGGAGGATAAACAGGGAGAA
	ERBS3-Rev	TCCCAGTCCATCTATCCTCA
• *ERBS4/SMAD7*	ERBS4-Fwd	TGGGTCCAAGGACAGATGTA
	ERBS4-Rev	ACTCTCTGCATTGGTGAAGC

Supplemental Table 2 ***(continued)***

• *ERBS5/PGR*	ERBS5-Fwd	TTCCCAGAGGTTTTCACAGA
	ERBS5-Rev	TTACACAGGCAGGACGACTT
• *ERBS6*	ERBS6-Fwd	AAGGAAGCAGTGACCAGGTT
	ERBS6-Rev	CCCTTTACCTCCAGTCTTGG
• *ERBS7*	ERBS7-Fwd	GCTGGCACTGCTTTTGTTAT
	ERBS7-Rev	GAAATTCCAGCTGCTCAAGA
• *ERBS8*	ERBS8-Fwd	AGCCAAGCATTCAACTGAGA
	ERBS8-Rev	AATCTTGGCTTCTGGTCACA
• *ERBS9*	ERBS9-Fwd	ACAAGGGCAAGAAAAAGACC
	ERBS9-Rev	GTCGCCAGTGTTAGAAGTGC
• *ERBS10*	ERBS10-Fwd	GATTTAGGGCACAGGGAGAT
	ERBS10-Rev	TGCAGTGTCTAGCCATCAGA
• *EGRBS1*	EGR1-BS1-Fwd	CTGAGTCTGGTGGGAGTGTC
	EGR1-BS1-Rev	GTCCAGCTCACCTCCGTTA
• *EGRBS2*	EGR1-BS2-Fwd	ACTGGATGGCAATGTTTGTC
	EGR1-BS2-Rev	CCTCGCCACACTCTTCTCTA
• *EGRBS3*	EGR1-BS3-Fwd	TGCGGATGATCCTGATAGTT
	EGR1-BS3-Rev	AGAAAGCACGTGGAAAACAG
• *EGRBS4*	EGR1-BS4-Fwd	GTCACCAAACCACAGAGAGG
	EGR1-BS4-Rev	GGTGGGGATGTTTAAAGGAA
• *GABPABS1*	GABPA-BS1-Fwd	ACACCTGGAGGGTTTACTGG
	GABPA-BS1-Rev	GCACAACCCCTTTCTTTCTC
• *GABPABS2*	GABPA-BS2-Fwd	TCAGAGCTGACGTCCAGTG
	GABPA-BS2-Rev	GATCTGTTCGCTCGAGTCC
• *GABPABS3*	GABPA-BS3-Fwd	GAACGTGTCCAGAGACTGCT
	GABPA-BS3-Rev	AGAGTGGCTTGCTGAGTTTG
• *GABPABS4*	GABPA-BS4-Fwd	CCCAGCATATTTCACTGTCC
	GABPA-BS4-Rev	CTTCCTCTTGACCACAGTGC
• *KLF4BS1*	KLF4-BS1-Fwd	CTGAGTCTGGTGGGAGTGTC
	KLF4-BS1-Rev	GTCCAGCTCACCTCCGTTA
• *KLF4BS2*	KLF4-BS2-Fwd	GTGTGCGAGGTAGTGCTCTT
	KLF4-BS2-Rev	GACGTAAGCCCTGCTCAGTA
• *KLF4BS3*	KLF4-BS3-Fwd	GCAAAAGCAACAAGTGTGGT
	KLF4-BS3-Rev	TATGCCCTTCCATCTCCTTT
• *KLF4BS4*	KLF4-BS4-Fwd	CTGACGTAATGCCAGTTTCC
	KLF4-BS4-Rev	GATGATCTTGGCGTTCAATC
• *Control Region 1*	CR1-Fwd	CCTGCTTGCTGTCTGAGC
	CR1-Rev	TGTCGCCATCAGGATTTC
• *Control Region 1*	CR2-Fwd	TGGCAGGCAGTAGATGCT
	CR2-Rev	AGCTCTGCAAAGGGGAAG

Supplemental Table 2 ***(continued)***

C. Primers used for 3C-PCR

Gene	Name	Outer Primer Sequence (5' to 3')	Inner Primer Sequence (5' to 3')
• *GREB1*	H Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	AGACCCCTTACACAGTCA	GACATGTCTTTGATGTTTTC
	a Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	GTTCAAGCAGTCCGAGTA	AGGTGATCTGCCTATCTCT
	b Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	CATGATTTGTTTTATCTTCC	GGAATTGTTCATCTTCTTTC
	c Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	ACCTGAACCTTCTAAGTAGC	CACAGCCAGTTAATTTTTAT
	d Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	AACAGTAGATGCTCTGTGAG	CGAGTAGCTGGGATTACA
	e Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	GACTCATTTGAGGTTCGT	GAATCTTCCTTTTCCTCTC
• *P2RY2*	H Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	AGCAAATGTTTACTCAGAAG	GGAGATGCTTATGTGGTG
	a Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	AGGACAGTTAAGCCTCTG	GGTAGAAAGGGTCAGTCA
	b Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	CAGAAATGTTGTGAGAACTAA	ACATACACAGAGTGCTGTTC
	c Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	CTGGTTTACCAACAATGATA	ATAGCAACCAGAACAGAGA
	d Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	GACTAAGCTCCAGAGTGTTT	CTCCACCTCCCTTATCTAC
• *GAPDH*	H Fwd	CCCACTCCTCCACCTTTGAC	TGGTGGCTGGCTCAGAAAAA
	Rev	GAGATTCAGTGTGGTGGGGG	TCCTCTTGTGCTCTTGCTGG
	a Fwd	GTAAATGTCACCGGGAGGAT	CCTTCTCCCCATTCCGTCTT
	Rev	CAGAAGGGCCTCTCCTGTA	TGTGCGGTGTGGGATTGTC
	b Fwd	CTACATGGTGAGCCCCAAAG	CGTATTCCCCCAGGTTTACA
	Rev	TTATCCCTCTCTCCCCACAA	TGGAATGATCCGGTATGGAG

D. siRNAs used to knockdown RAD21 *(Dharmacon SMART Pool)*

Gene	Sequence (5' to 3')
• *RAD21*	GGAAGAAGCAUUUGCAUUG
	GAACAGAGCACCAGCAAUC
	GAGCCCAACUUAGUGAUUA
	GGGAGUAGUUCGAAUCUAU

3. Supplemental Methods

Cell culture and treatments

MCF-7 human breast adenocarcinoma cells were kindly provided by Dr. Benita Katzenellenbogen (University of Illinois, Urbana-Champaign). The cells were maintained in minimal essential medium (MEM) supplemented with Hank's salts (Sigma) and 5% calf serum. The cells were plated for experiments in phenol red-free MEM (Sigma) supplemented with 5% charcoal-dextran treated calf serum (CDCS) for at least three days prior to hormone or drug treatment. As indicated for the different experiments, the cells were treated with 100 nM E2 for the times specified. For the transcription inhibition experiments, the cells were pretreated with or without 1 μM flavopiridol (Sigma) for 1 hour prior to treatment with E2.

Antibodies

The antibodies used for chromatin immunoprecipitation (ChIP) assays are as follows: ESR1 (rabbit polyclonal generated in Kraus lab); POLR2A (Pol II; sc-899), NCOA2 (sc-343), NCOA3 (sc-9119), EP300 (sc-585), CREBBP (sc-369), GABPA (sc-22810), ACTB (sc-69879), and non-immune IgG (rabbit polyclonal; all from Santa Cruz Biotech); RAD21 (Cohesin; ab992), KLF4 (ab151733), H3K4me1 (ab8895), H3K4me3 (ab8580), H3K27ac (ab4729), and H3 (ab1791) (rabbit polyclonal; all from Abcam).

siRNA-mediated knockdown of RAD21

SMART pool siRNAs (Dharmacon, Thermo Scientific, Inc.) were used to specifically knockdown RAD21 versus pool of non-targeting siRNAs used as a negative control. MCF-7 cells were plated at a density of 3×10^5 cells per well in 6-well plates and transfected using DharmaFECT transfection reagent according to the manufacturer's instructions. siRNA-transfected cells were grown for 24 hours in serum-free media followed by complete medium with charcoal-dextran stripped serum for three days prior to estrogen treatment. The knockdown efficiency of RAD21 was assessed by both RT-qPCR and Western blotting.

Chromatin immunoprecipitation-quantitative PCR (ChIP-qPCR)

MCF-7 cells were grown in estrogen free medium to ~80% confluence and then treated with 100 nM E2 for the indicated time, with or without 1 μM Flavopiridol (Sigma) for 1 h prior to E2 treatment. ChIP analyses were conducted as described previously (Kininis et al. 2007), with a few modifications. Treated MCF-7 cells were crosslinked with 1% formaldehyde for 10 min. at 37°C, followed by quenching with 125 mM glycine for 5 min. at 4°C. The crosslinked cells were washed with PBS, harvested, lysed with lysis buffer [Tris•HCl (pH 7.9), 0.5% SDS, 10 mM EDTA, 1 mM DTT, 1x protease inhibitor cocktail (Roche)] and subjected to sonication for 7 cycles of 20 seconds each at setting high using a Diagenode Bioruptor to obtain ~500 bp DNA fragments. The lysate was incubated with the antibodies indicated along with a rabbit IgG control after input material was removed, followed by incubation with protein A-agarose beads for 1.5 hours. The immunoprecipitates were collected and washed with wash buffer [20 mM Tris•HCl (pH 7.9), 0.25% NP-40, 0.05% SDS, 2 mM EDTA, 250 mM NaCl, 1x protease inhibitor cocktail (Roche)] at 4°C and eluted by incubating overnight in elution buffer [100 mM $NaHCO_3$, 1 % SDS] at 65°C to reverse the crosslinks, followed by digestion with proteinase K. The ChIP'ed DNA was subjected to phenol:chloroform extraction and analyzed by qPCR using gene-specific primers (Supplemental Table 2) and a 384-well real-time PCR thermocycler with

SYBR Green detection. Each experiment was performed a minimum of three times with independent biological samples to ensure reproducibility.

Analysis of eRNAs and mRNAs by reverse transcription-quantitative PCR (RT-qPCR).

MCF-7 cells were grown in estrogen free medium to ~80% confluence and then treated with 100 nM E2 for the indicated time, with or without 1 μM Flavopiridol (Sigma) for 1 h prior to E2 treatment. RT-qPCR detection of eRNAs and mRNAs were performed as described previously (Sun et al.), with some minor modifications. Total RNA was isolated from the treated cells using TRIzol reagent (Invitrogen) and subjected to RT using random hexamers and M-MLV Reverse transcriptase (Promega). The cDNA was then subjected to qPCR analysis using a Roche LightCycler 480 system with SYBR Green detection and gene-specific primers (Supplemental Table 2). Each experiment was performed a minimum of three times with independent biological samples to ensure reproducibility.

Chromosome conformation capture (3C)

Chromosome conformation capture was conducted as previously described (Pan et al. 2008), with the following modifications.

Treatment and crosslinking of cells. MCF-7 cells were grown in estrogen free medium to ~80% confluence and then treated with 100 nM E2 for 40 min., with or without 1 μM Flavopiridol (Sigma) for 1 h prior to E2 treatment. They were then fixed with 1% formaldehyde for 10 min. After quenching with 200 mM glycine for 5 min, the cells were lysed by douncing in lysis buffer [10 mM Tris•HCl (pH 7.5), 10 mM NaCl, 0.2% Triton-X, 1x protease inhibitor cocktail (Roche)] and incubating with gentle mixing at 4°C for 30 min. The nuclei were collected by centrifugation.

DNA digestion. The isolated nuclei were dispersed in 1.2x restriction enzyme buffer supplied by the manufacturer (NEB) with 0.3% SDS, followed by gentle mixing at 37°C for one hour. Triton-X was then added at 2% final concentration and incubated with gentle mixing at 37°C for one hour. The DNA was digested with gentle mixing at 37°C overnight with 400 U of restriction enzyme (NEB). The enzymes used were *BglII* for *P2RY2* and *BtgI* for *GREB1*.

DNA ligation. The digestion reaction was terminated by the addition of 1.6% SDS with incubation at 65°C for 20 min. with gentle mixing. The digested nuclei were then transferred to a 50 mL conical tube containing 6.125 mL ligation buffer [1.15x T4 DNA ligase buffer (NEB) and 1% Triton-X100] and incubated at 37°C for one hour with gentle mixing. Ligation was performed by the addition of 2000 U of T4 DNA Ligase (NEB) with incubation at 16°C for 4 hours, followed by incubation at 25°C for 30 min. As a control, similar samples were incubated without ligase. After digestion with Proteinase K, the DNA was de-crosslinked incubating at 65°C overnight and then purified using phenol-chloroform extraction, followed by ethanol precipitation.

Loop detection. Nested PCR to detect chromatin interactions was performed using Taq DNA polymerase (NEB). Each primer set (Supplemental Table 2) was designed unidirectionally upstream of the restriction enzyme digestion sites such that PCR amplifies any DNA resulting from ligation between the hub and a test site, as described previously (Miele et al. 2006; Hagege et al. 2007). Digested and ligated bacterial artificial chromosome (BAC) DNA spanning the entire locus for each gene analyzed was used as a PCR control. Each experiment was performed three times with independent biological samples to ensure reproducibility.

Global run-on sequencing (GRO-seq)

GRO-seq was performed using nuclei isolated from two biological replicates of E2-treated MCF-7 cells, as previously described (Hah et al. 2011). The 0, 10, and 40 min. E2 treatment libraries that we generated and analyzed previously (Hah et al. 2011) were re-sequenced to a greater depth (~30 million mappable reads per condition), combined with the previous data, and used for the genomic analyses described herein (a total of ~45 million mappable reads per condition). The original and resequenced GRO-seq data sets are available from the NCBI's Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) using accession number GSE27463. For some gene-specific analyses showing genome browser tracks, we also used data from libraries generated from other E2 treatment time points (e.g., 25 min; GEO accession number GSE41324) using a circular ligation-based method (Ingolia et al. 2009).

Analysis of GRO-seq data

GRO-seq data were analyzed using software described previously (Hah et al. 2011) and the approaches described below. Software, scripts, and other information can be obtained by contacting W. Lee Kraus.

Read alignment. GRO-seq reads were aligned to human reference genome (hg18), including autosomes, X chromosome, and one complete copy of an rDNA repeat (GenBank ID: U13369.1). The SOAP.2.21 software package (Li et al. 2009) was used to align the reads using the following parameters: (1) all n mappings were removed (-r 0); (2) three mismatches were allowed in each mapped read (-v 3); (3) low-quality reads with more than 10 ambiguous bases were removed (-N 10); and (4) for reads failing to align over the entire length of the read, the first 32 bp was used (-l 32).

Transcript calling. Transcript calling was performed using a two-state hidden Markov model using the GRO-seq data analysis package described in Hah et al. (2011). We used a shape setting parameter of 5 and -log transition probability of 200 to predict the transcription units. The predicted transcripts were assigned into six classes as described in Hah et al. (2011) using annotations from RefSeq, ENSEMBL, and UCSC Known Gene databases.

Defining Classes of ESR1 enhancer transcripts (eRNAs). The repertoire of genomic ESR1 binding sites (ERBSs) was extracted from ChIP-seq data provided in Welboren et al., (2009) (GEO accession number GSM365926). Those ERBSs >10 kb away from the 5' or 3' ends of annotated genes were defined as "Intergenic ERBSs." They were divided into three classes based on the presence, location, and orientation of GRO-seq-defined transcripts: (1) those overlapping transcripts originating from both strands of DNA, running in opposite directions as a divergent pair; (2) those overlapping a transcript originating from one strand of DNA only ("Unpaired); and (3) those not overlapping a transcript. The transcripts in classes 1 and 2 were further categorized based on the length of the transcript unit/primary transcript as 'short' (length < 9 kb; eRNAs, by our definition) and 'long' (length >9 kb; which likely represent other classes of non-coding RNAs, such as lncRNAs).

Transcript maps. Individual eRNAs were visualized in the transcript map shown in Fig. 2C at genomic positions relative to the associated intergenic ERBS using custom PERL scripts and R. The transcript maps were centered on the ERBSs and the relative positions of each eRNA with respect to the corresponding ERBS was plotted (shorter and longer eRNAs in the pair: blue and red, respectively; unpaired eRNA: red). The transcript maps were ordered based on the length of the shorter eRNA in the pair or on the length of the unpaired eRNA.

Determining estrogen regulation of transcripts and generating heat maps. The effects of E2 treatment on the expression of the eRNAs were analyzed using edgeR (Robinson et al. 2010), as described previously (Hah et al. 2011). We used the 10 min., 40 min., and 160 min. E2 treatment time points (two biological replicates for each time point) to determine the E2-dependent regulation of eRNAs. The results were plotted as a heat map using Java TreeView (Saldanha 2004), ordered based on the magnitude of expression at the 40 min. time point.

Analysis of ChIP-seq data

ChIP-Seq datasets from MCF-7 cells were obtained from the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and the ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) on-line databases. The data sets are listed in Supplemental Table 1. The raw files were aligned to hg18 using BOWTIE (Langmead et al. 2009). Uniquely mappable reads were converted into bigWig files using BEDTools (Quinlan and Hall 2010) for visualization in the UCSC genome browser.

Motif searching under ERBSs. De novo motif searching was performed on a 200 bp region around the center of the ERBSs (± 100 bp) using the command-line version of MEME (Bailey et al. 2006). The following parameters were used for motif prediction: (1) zero or one occurrence per sequence (-mod zoops); (2) number of motifs (-nmotifs 10); (3) minimum, maximum width of the motif (-minw 8, -maxw 15); and (4) search for motif in given strand and reverse complement strand (-revcomp). The predicted motifs from MEME were matched to known motifs using STAMP (Parks and Beiko 2010). In a subsequent analysis, the position weight matrices (PWMs) of the motifs predicted by MEME were taken and scanned across the 200 bp regions (± 100 bp) around the center of ERBS using a command-line version of MotifScanner (Aerts et al. 2003) with a prior value threshold of 0.3.

Analysis of ESR1 ChIA-PET data

ESR1-dependent intra-chromosomal loops defined by Fullwood et al. (2009) were obtained from the ChIA-PET visualization browser (http://cms1.gis.a-star.edu.sg). Using this data, we examined if the ERBSs defined by Welboren et al. (2009) are involved in loops and if they loop to the promoters of 19,008 unique RefSeq genes. We then used the center of the head and tail of each loop defined by Fullwood et al. (2009) to identify looping events between an ERBS and a target gene promoter. Specifically, we assayed for loops originating within a 2 kb window around the peak center of ERBS with transcripts (i.e., either S-S paired or S-U) and looping to a 10 kb window centered around the TSSs of target genes. We then used GRO-seq data to determine the amount of transcription at the ERBSs and the target genes. Various metaplot representations of GRO-seq and ChIP-seq data were used to compare the properties of ERBSs and genes with or without looping events. The raw hg19 aligned files obtained from Fullwood et al. (2009) were lifted over to hg18 and used to plot metaplots.

Genomic data analysis and visualization

We used the following approaches to summarize and visualize genomic data from GRO-seq, ChIP-seq, and ChIA-PET.

Metaplots. Metaplots were used to illustrate the distribution of GRO-seq, ChIP-seq, and ChIA-PET reads around ESR1 peak maxima (or other genomic features) using the metaplot function in our GRO-seq package (Hah et al. 2011).

Boxplots. Boxplot representations were used to minimize the bias caused by outliers in the data, which can overly influence metaplot representations. The boxplots allowed for accurate comparisons across ERBSs with S-S paired eRNAs, S-unpaired eRNAs, and without eRNA. The read distribution in a 2 kb window (± 1 kb) around the ERBSs was calculated and plotted using the boxplot function in R. The inter-quartile regions (IQRs) of the boxplots were used to plot metaplots centered on the ERBSs. All the metaplots and boxplots were scaled to a library size of 15 million reads to normalize against different read densities. A Wilcoxon rank sum test was performed to determine the statistical significance of all comparisons.

Random sampling to select ERBSs with similar levels of ESR1 binding. Random sampling was performed using a custom PERL script to identify a subset of ERBSs with similar levels of ESR1 binding from each population of sites (S-S, S-U, and no transcripts). A Kolmogorov-Smirnov (K-S) test was performed to confirm that the randomly selected subsets of ERBSs in each group had similar distributions of ESR1 binding strength.

Predicting enhancers based on GRO-seq data

As shown in Figs. 6 and 7, and Supplemental Figs. 15 through 17, we used GRO-seq data combined with DNA-binding transcription factor motif information to predict active enhancers. We used directed and unbiased approaches, as described below.

Directed search for enhancers based on GRO-seq data. We performed a genome-wide supervised motif search on the hg18 genome using the command-line version of Find Individual Motif Occurrences (FIMO) in the MEME suite (Grant et al. 2011). The motifs for all curated, non-redundant, vertebrate transcription factors (130 total) in the JASPAR database (Bryne et al. 2008) were searched using FIMO with a p-value threshold of 1×10^{-6}; 78 out of 130 motifs were identified and mapped in hg18 at this p-value. From the total set of identified motifs, we selected only intergenic motifs (>10 kb from RefSeq genes) that have eRNAs [either "short-short paired" (S-S) or "short unpaired" (S-U)] originating within a 2 kb window around the center of the motif (i.e., ± 1 kb relative to the motif). For each occurrence of a specific intergenic motif with an overlapping eRNA that we selected, we collected all GRO-seq reads within a 1 kb window around the center of the motif (i.e., ± 0.5 kb relative to the motif) and normalized them to the total number of occurrences of the motif.

Unbiased search for enhancers based on GRO-seq data. We searched our MCF-7 GRO-seq data sets for sites in the genome expressing intergenic (>10 kb away from the start or end of an annotated RefSeq gene) short-short paired eRNAs using the following parameters: primary transcript length shorter than 9 kb and an average overlap of 3 kb. All occurrences that fit these criteria were collected and subjected to motif analysis. *De novo* motif searching was performed on a 1 kb region around the center of the plus and minus strand overlap (± 500 bp) using the command-line version of MEME (Bailey and Elkan 1994). The following parameters were used for motif prediction: (1) zero or one occurrence per sequence (-mod zoops); (2) number of motifs (-nmotifs 10); (3) minimum, maximum width of the motif (-minw 8, -maxw 15); and (4) search for motif in given strand and reverse complement strand (-revcomp). The predicted motifs from MEME were matched to known motifs using STAMP (Parks and Beiko 2010).

Nearest neighbor gene analysis. The nearest neighbor gene for each predicted enhancer (i.e., intergenic S-S paired eRNA-producing; 569 total) was identified. The analysis was limited to the nearest neighboring genes showing regulation of mRNA production by E2. The data were expressed as the percent of nearest neighbor genes with the specified pattern of regulation by E2

[i.e., up, down, unregulated ("none")]. For each pattern of regulation, the strength of association between the regulation of the enhancer and the regulation of nearest neighbor gene is statistically significant (Fisher's exact test; $p < 0.0001$).

Comparing enhancer prediction using GRO-seq/eRNAs with enhancer prediction using ChIP-seq/histone marks

We predicted enhancers in MCF-7 cells grown under basal (i.e., -E2) growth conditions by using GRO-seq data as described above, as well as by using H3K4me1 and H3K27ac ChIP-seq data. The ChIP-seq data were aligned to the genome as described above. Peak calling was performed using the MACS algorithm with a p-value threshold of 1×10^{-10} and peaks >10 kb from the 5' or 3' ends of annotated genes were selected as enhancers. We limited our analysis to the top 20% of enhancers based on ChIP-seq signal. The enhancers predicted by GRO-seq (eRNAs) and histone marks were then assayed for looping events using a Pol II ChIA-PET dataset (Li et al. 2012).

Testing the enhancer prediction pipeline in mouse embryonic stem cells

In order to test our enhancer prediction method, we subjected a GRO-seq data set from mouse ES cells (Min et al. 2011) to the analysis pipeline noted above. The raw GRO-seq data files were obtained from NCBI/GEO (accession number GSE29218) and aligned to the mm9 build of the mouse genome. Transcripts were called as described previously (Hah et al.) using an HMM-based algorithm with the following parameters to call transcription units: shape setting (V 25) and –log transition probability (B 150). Short-short paired eRNAs were identified in the set of all called transcripts using the enhancer prediction pipeline described above. The results were aligned to ChIP-seq data from the following data sets in NCBI/GEO: Pol II, EP300, H3K4me1, H3K27ac (accession number GSE29218) (Shen et al. 2012); NANOG, POU5F1 and SOX2 (accession number GSE11724) (Marson et al. 2008). For the NANOG, POU5F1, and SOX2 ChIP-seq datasets, peak calling was performed using the MACS pipeline (Zhang et al. 2008). ChIP-seq peaks within a 2 kb window (± 1 kb) around the center of the plus and minus strand overlap of the predicted SS paired eRNAs were identified and metaplots were plotted using the metaplot function in our GRO-seq package (Hah et al. 2011).

4. Supplemental Discussion

Integration of genomic data sets to study ESR1 enhancer eRNAs

In the studies described herein, we integrated and analyzed a large number of genomic data sets using a novel computational pipeline to provide a comprehensive and global view of ESR1 enhancers in the MCF-7 human breast cancer cell line. The data sets that we analyzed included (1) GRO-seq from a short time course of E2 treatment, which allowed us to monitor active transcription at ERBSs, (2) ChIP-seq ± E2, which allowed us to monitor histone modifications (e.g., H3K4 me1, me2, me3), the binding of pioneer factors (e.g., FOXA1 and TFAP2C), and the binding of coregulators (e.g., CREBBP, EP300, NCOA1, NCOA2, and NCOA3), (3) ChIA-PET + E2, which allowed us to monitor enhancer looping, and (4) DNase-seq and FAIRE-seq ± E2, which allowed us to monitor the chromatin state at ERBS. Our analyses reveal key facets of ERα function at its enhancers, as well as new information about the production and features of estrogen-dependent eRNAs.

Properties and features of ESR1 enhancer eRNAs

ESR1 enhancer eRNAs have an average transcription unit (i.e., primary transcript) length of ~3 to 5 kb. In addition to being bidirectionally transcribed, upregulated by estrogen, and minimally polyadenylated, many ESR1 enhancer eRNAs are also 5' 7-methylguanosine capped based on TSS-seq (Yamashita et al. 2011) and locus-specific capping assays (data not shown). These features of ESR1 enhancer eRNAs are consistent with previous reports of enhancer transcription and eRNAs (De Santa et al. 2010; Kim et al. 2010; Hah et al. 2011; Wang et al. 2011; Djebali et al. 2012).

Our results indicate that the production of eRNAs at ERBSs strongly correlates with the enrichment of H3K4me1, H3K27ac, CREBBP, EP300, RNA pol II, and an open chromatin architecture. In addition, we found that the E2-dependent production of eRNAs at ERBSs also strongly correlates with the binding of the p160 steroid receptor coregulators NCOA2 and NCOA3. NCOA proteins interact with ESR1 in an E2-dependent manner through a hydrophobic cleft on the ligand binding domain of the receptor (Heery et al. 1997; Torchia et al. 1997). In addition, NCOA proteins interact directly with CREBBP and EP300, thus allowing them to promote the indirect binding of CREBBP and EP300 to DNA-bound ESR1 (Torchia et al. 1997; Kim et al. 2001). Thus, E2-dependent formation of the DNA-ESR1-NCOA-CREBBP/EP300 complex is likely to represent an initial step in the formation of an active enhancer at ERBSs. Interestingly, we observed that NCOA1 is enriched at ERBSs before and after E2 treatment, which contradicts previous models for the ligand-dependent functions of this coregulator (Acevedo and Kraus 2004). Nonetheless, NCOA1 enrichment strongly correlates with the E2-dependent production of eRNAs at ERBSs.

Collectively, our results demonstrate that active transcription at ERBSs, as determined by GRO-seq, tracks with genomic features thought to be marks of active enhancers. From these results, we can begin to understand the order of events that lead to the assembly of an active enhancer complex at ERBSs. The process is initiated by E2-bound ESR1, which binds to direct (i.e., ERE-mediated) or indirect (i.e., tethered) sites across the genome and then nucleates the formation of enhancer complexes containing coregulators, some of which function as histone-modifying enzymes, as well as RNA pol II and perhaps looping factors, such as Mediator and cohesin (Kagey et al. 2010). Histone modification, looping, and eRNA production then follow.

Predicting enhancers using genomic features

Histone modifications (e.g., H3K4me1, H3K27ac), histone variants (e.g., H2A.Z), coactivators (e.g., EP300, CREBBP, Mediator), and an open chromatin architecture (e.g., DNase I hypersensitivity) have been identified as genomic features that mark or identify enhancers (Melgar et al. 2011; Natoli and Andrau 2012). eRNA production, as well as the aforementioned enhancer features, have been used in a number of studies to identify or predict enhancers on a genome-wide basis (Pennacchio et al. 2007; Won et al. 2008; Visel et al. 2009; Melgar et al. 2011; Fernandez and Miranda-Saavedra 2012; Maston et al. 2012). Our results show that estrogen-regulated eRNAs can be used to predict enhancers.

Interestingly, our data indicate that enhancer transcription correlates with more enhancer features than H3K4me1, which has been used previously to predict enhancers (Heintzman et al. 2007; Maston et al. 2012). For example, enrichment of H3K4me1 does not track with enhancer looping; it is enriched at transcribed ERBS whether or not they loop to a target gene promoter. In this regard, transcription may be a more accurate mark of active enhancers than histone modifications, such as H3K4me1. Thus, analysis of enhancer transcription by GRO-seq provides additional information about enhancer activity and function, which cannot be obtained with ChIP-seq for transcription factors, RNA pol II, or histone modifications. In the end, a comprehensive and integrated approach that includes multiple types of information, as we have described here, is the best approach for defining and understanding enhancers.

5. Supplemental References

Acevedo ML, Kraus WL. 2004. Transcriptional activation by nuclear receptors. *Essays Biochem* **40**: 73-88.

Aerts S, Thijs G, Coessens B, Staes M, Moreau Y, De Moor B. 2003. Toucan: deciphering the cis-regulatory logic of coregulated genes. *Nucleic Acids Res* **31**(6): 1753-1764.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.

Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* **34**(Web Server issue): W369-373.

Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res* **36**(Database issue): D102-106.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* **8**(5): e1000384.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F et al. 2012. Landscape of transcription in human cells. *Nature* **489**(7414): 101-108.

Fernandez M, Miranda-Saavedra D. 2012. Genome-wide enhancer prediction from epigenetic signatures using genetic algorithm-optimized support vector machines. *Nucleic Acids Res* **40**(10): e77.

Frietze S, Wang R, Yao L, Tak YG, Ye Z, Gaddis M, Witt H, Farnham PJ, Jin VX. 2012. Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol* **13**(9): R52.

Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH et al. 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**(7269): 58-64.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**(7): 1017-1018.

Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2**(7): 1722-1733.

Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. 2011. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**(4): 622-634.

He HH, Meyer CA, Chen MW, Jordan VC, Brown M, Liu XS. 2012. Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Res* **22**(6): 1015-1025.

Heery DM, Kalkhoven E, Hoare S, Parker MG. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**(6634): 733-736.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**(3): 311-318.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**(5924): 218-223.

Joseph R, Orlov YL, Huss M, Sun W, Kong SL, Ukil L, Pan YF, Li G, Lim M, Thomsen JS et al. 2010. Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. *Mol Syst Biol* **6**: 456.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**(7314): 430-435.

Kim MY, Hsiao SJ, Kraus WL. 2001. A role for coactivators and histone acetylation in estrogen receptor alpha-mediated transcription initiation. *EMBO J* **20**(21): 6084-6094.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**(7295): 182-187.

Kininis M, Chen BS, Diehl AG, Isaacs GD, Zhang T, Siepel AC, Clark AG, Kraus WL. 2007. Genomic analyses of transcription factor binding, histone acetylation, and gene expression reveal mechanistically distinct classes of estrogen-regulated promoters. *Mol Cell Biol* **27**(14): 5090-5104.

Kong SL, Li G, Loh SL, Sung WK, Liu ET. 2011. Cellular reprogramming by the conjoint action of ERalpha, FOXA1, and GATA3 to a ligand-inducible growth state. *Mol Syst Biol* **7**: 526.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3): R25.

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J et al. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**(1-2): 84-98.

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**(15): 1966-1967.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**(3): 521-533.

Maston GA, Landt SG, Snyder M, Green MR. 2012. Characterization of enhancer function from genome-wide analyses. *Annu Rev Genomics Hum Genet*.

Melgar MF, Collins FS, Sethupathy P. 2011. Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol* **12**(11): R113.

Miele A, Gheldof N, Tabuchi TM, Dostie J, Dekker J. 2006. Mapping chromatin interactions by chromosome conformation capture. *Curr Protoc Mol Biol* **Chapter 21**: Unit 21 11.

Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT. 2011. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev* **25**(7): 742-754.

Natoli G, Andrau JC. 2012. Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet*.

Pan YF, Wansa KD, Liu MH, Zhao B, Hong SZ, Tan PY, Lim KS, Bourque G, Liu ET, Cheung E. 2008. Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements. *J Biol Chem* **283**(47): 32977-32988.

Parks DH, Beiko RG. 2010. Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**(6): 715-721.

Pennacchio LA, Loots GG, Nobrega MA, Ovcharenko I. 2007. Predicting tissue-specific enhancers in the human genome. *Genome Res* **17**(2): 201-211.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1): 139-140.

Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**(17): 3246-3248.

Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* **20**(5): 578-588.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV et al. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**(7409): 116-120.

Tan SK, Lin ZH, Chang CW, Varang V, Chng KR, Pan YF, Yong EL, Sung WK, Cheung E. 2011. AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. *EMBO J* **30**(13): 2569-2581.

Torchia J, Rose DW, Inostroza J, Kamei Y, Westin S, Glass CK, Rosenfeld MG. 1997. The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature* **387**(6634): 677-684.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**(7231): 854-858.

Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA et al. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**(7351): 390-394.

Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, Chen Z, Beroukhim R, Wang H, Lupien M et al. 2009. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**(2): 245-256.

Welboren WJ, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN, Stunnenberg HG. 2009. ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J* **28**(10): 1418-1428.

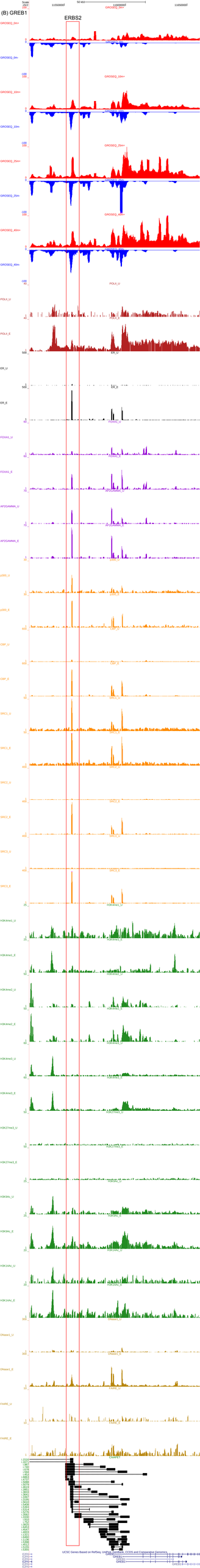
Won KJ, Chepelev I, Ren B, Wang W. 2008. Prediction of regulatory elements in mammalian genomes using chromatin signatures. *BMC Bioinformatics* **9**: 547.

Yamashita R, Sathira NP, Kanai A, Tanimoto K, Arauchi T, Tanaka Y, Hashimoto S, Sugano S, Nakai K, Suzuki Y. 2011. Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis. *Genome Res* **21**(5): 775-789.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9): R137.

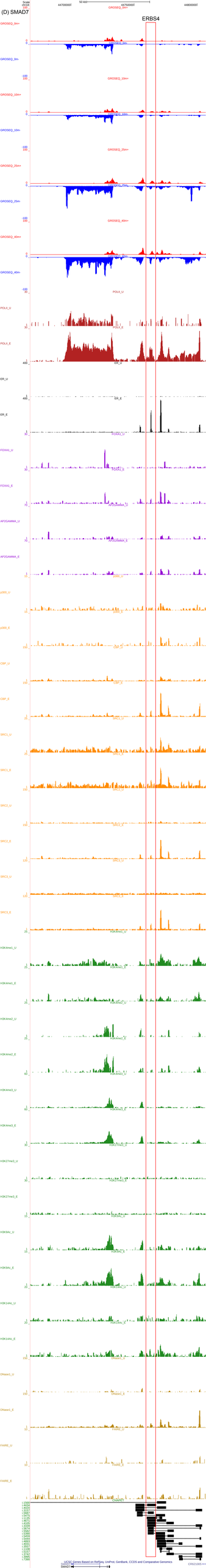
Zwart W, Theodorou V, Kok M, Canisius S, Linn S, Carroll JS. 2011. Oestrogen receptor-co-factor-chromatin specificity in the transcriptional regulation of breast cancer. *EMBO J* **30**(23): 4764-4776.

(A) P2RY2

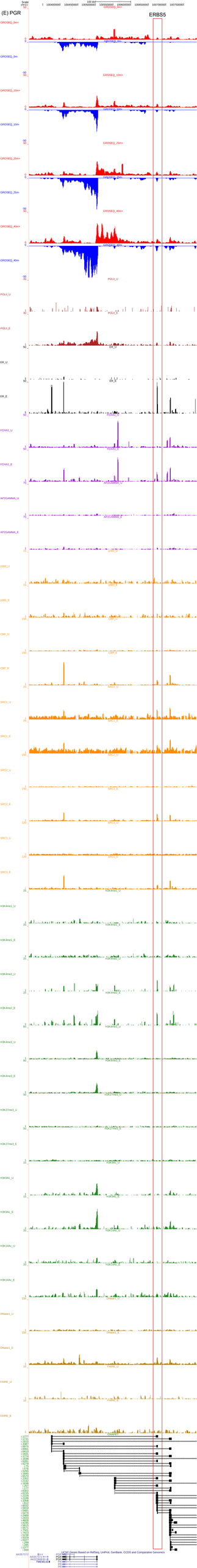

(B) GREB1
ERBS2
GROSEQ_0m+
GROSEQ_0m-
GROSEQ_10m+
GROSEQ_10m-
GROSEQ_20m+
GROSEQ_20m-
GROSEQ_40m+
GROSEQ_40m-
POLII_U
POLII_E
ER_U
ER_E
FOXA1_U
FOXA1_E
AP2AMMA_U
AP2AMMA_E
p300_U
p300_E
CBP_U
CBP_E
SRC1_U
SRC1_E
SRC2_U
SRC2_E
SRC3_U
SRC3_E
H3K4me1_U
H3K4me1_E
H3K4me2_U
H3K4me2_E
H3K4me3_U
H3K4me3_E
H3K27me3_U
H3K27me3_E
H3K9ac_U
H3K9ac_E
H3K14Ac_U
H3K14Ac_E
DNase1_U
DNase1_E
FAIRE_U
FAIRE_E
ChIAPET
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
GREB1

(C) SBNO2

ERBS3
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics



UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics


(E) PGR
ERBS5
GROSEQ_0m+
GROSEQ_0m-
GROSEQ_10m+
GROSEQ_10m-
GROSEQ_25m+
GROSEQ_25m-
GROSEQ_40m+
GROSEQ_40m-
POLII_U
POLII_E
ER_U
ER_E
FOXA1_U
FOXA1_E
AP2GAMMA_U
AP2GAMMA_E
p300_U
p300_E
CBP_U
CBP_E
SRC1_U
SRC1_E
SRC2_U
SRC2_E
SRC3_U
SRC3_E
H3K4me1_U
H3K4me1_E
H3K4me2_U
H3K4me2_E
H3K4me3_U
H3K4me3_E
H3K27me3_U
H3K27me3_E
H3K9AC_U
H3K9AC_E
H3K4AC_U
H3K4AC_E
DNase1_U
DNase1_E
FAIRE_U
FAIRE_E
TMEM133
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics

(F) ERBS6