

Supplemental Materials - Kininis et al. (2007) Molecular and Cellular Biology
"Genomic Analyses of Transcription Factor Binding, Histone Acetylation, and Gene Expression Reveal Mechanistically Distinct Classes of Estrogen-Regulated Promoters"

List of Supplemental Materials

A. Materials included in this document: page

1) Supplemental Materials and Methods.....	2
2) Supplemental Figure Legends	10
3) Supplemental Table 1 Legend.....	12
4) Supplemental References	15
5) Outline of Statistical Analysis Using the Statistical Software R.....	19
6) Description of the ChIP-chip Data Set.....	21
7) Description of the Expression Microarray Data Set.....	22
8) Definitions and Data Used to Derive the Conclusions Presented	23

B. Materials included as separate files:

- 1) Supplemental Figures (*file format: PDF*)
- 2) Supplemental Table 1 (*file format: PDF*)

C. Materials provided upon request:

- 1) Annotated list of promoter sequences included on the array (*file format: Excel*)
- 2) R scripts (*file format: zipped folder containing r files*)
- 3) ChIP-chip data set (*file format: Excel*)
- 4) Expression microarray data set (*file format: Excel*)

1. Supplemental Materials and Methods

ChIP Assays

ChIPs were performed essentially as described previously (1, 40). MCF-7 cells were maintained in phenol red-free MEM supplemented with 5% charcoal-dextran stripped calf serum for at least 3 days before using. After growing to ~80 to 90% confluence, the cells were treated with either ethanol or 100 nM E2 for 45 min at 37°C, and then cross-linked with 1% formaldehyde in PBS at 37°C for 10 min. Crosslinking was quenched with 125 mM glycine in PBS for 5 min on ice. The cells were then rinsed twice with ice-cold PBS, collected into PBS containing a protease inhibitor cocktail (Roche Molecular Biochemicals), and centrifuged for 5 min. The cell pellets were resuspended in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris•HCl, pH 7.9, 1x protease inhibitor cocktail), incubated on ice for 10 minutes, and sonicated with three 11 second bursts (Branson Digital Sonifier, Model 250) to an average DNA fragment size of ~500 bp as determined by agarose gel electrophoresis. After centrifugation for 10 min at full speed in a microcentrifuge, the supernatants were collected and diluted 10-fold in dilution buffer (0.5% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris•HCl, pH 7.9, 1x protease inhibitor cocktail) followed by immunoclearing with protein A-agarose beads (Invitrogen; 1:1 ratio, 20 µL of a 50% slurry in 10 mM Tris•HCl, pH 8.1, and 1 mM EDTA) for 1 hr at 4°C. Immunoprecipitation was performed overnight at 4°C using a panel of specific polyclonal antibodies generated in-house or from commercial sources (see below), as well as no antibody controls. Immune complexes were collected by the addition of 40 µL of protein A-agarose with an additional 2 to 3 hour incubation. The agarose beads were washed four times for 10 min each in wash buffer (0.25% NP-40, 0.05% SDS, 2 mM EDTA, 20 mM Tris•HCl, pH 7.9, 250 mM NaCl, 2 mg/ml leupeptin, 2 mg/ml aprotinin), washed once with TE, resuspended in elution buffer (1% SDS, 0.1 M sodium bicarbonate), and heated at 65°C overnight to reverse the formaldehyde cross-linking and elute the DNA. After digestion with RNase (2 µg for 30 min at 37°C) and proteinase K (10 mg for 30 min at 37°C), the DNA fragments were purified using a Qiagen PCR purification kit. The resulting ChIP DNA material was used to probe the custom promoter array and in gene-specific qPCR analyses as described below.

The antibodies used were as follows: ER α (rabbit polyclonal generated in the Kraus lab), RNA Pol II (Santa Cruz N-20 and C-21; SC-899 and SC-900, respectively), pan-SRC (SRC-1/2/3, rabbit polyclonal generated in the Kraus lab), acetylated histones H3 and H4 (Upstate Biotechnology, Inc. 06-599 and 06-866, respectively), Pol II Ser2-Phos CTD (Covance H5; MMS-129R), c-Fos (rabbit polyclonal generated in the Kraus lab), and NELFA (SantaCruz A20; SC-23599).

Human promoter array design

A custom spotted estrogen-regulated promoter array was designed using approaches described previously (34, 36). A large panel of ~600 estrogen-regulated promoters were selected based on the following: (1) previous expression microarray studies (10, 11, 20, 26, 27, 31), (2) bioinformatic analyses (4, 33), and (3) previous gene-by-gene studies (22, 39). The genes selected for the array included both estrogen up-regulated and down-regulated genes, as well as genes whose regulation by estrogens is cell type-specific. For each gene, the sequence from approximately -800 to +200 bp relative to the TSS was extracted using the Database of Transcription Start Sites (DBTSS) (43) and the University of California at Santa Cruz (UCSC) genome browser (17). PCR primer pairs were either extracted from a ~13,000 promoter database

provided by the Whitehead Institute Center for Microarray Technology (34) or designed using the Primer 3 algorithm (38). Semi-automated PCR in microtiter plates was performed as described previously (34, 36) and the PCR products were verified by size. Our PCR conditions led to a > 90% success rate for the production of full-length products. The verified PCR products were purified and used to prepare spotted DNA microarrays at the Cornell University Microarray Core Facility. Each PCR product was spotted in quadruplicate on the arrays in order to ensure reproducibility within the same experimental condition. In addition to gene promoters, ~1 kb fragments spanning a 12 kb region surrounding the TSS of four well-characterized E2-regulated genes (TFF1, MYC, EBAG9 and CASP7) were also included on the microarray. Furthermore, twelve previously characterized non-promoter ER α -binding regions (6) and ~250 non-E2-regulated promoters and intergenic regions (designated as ‘controls’) were also included on the microarray. An annotated database of the promoter sequences included on the array can be found in the supplemental materials.

Hybridization of ChIP material to the custom promoter array

For microarray hybridization, the purified DNA was blunted using T4 DNA Polymerase (New England Biolabs), ligated to linkers, and amplified using LM-PCR, as described previously (36). The cycles of LM-PCR were empirically determined (23 cycles for ER α , RNA Pol II, and SRC; 21 cycles for acetylated histones H3 and H4) to ensure linear amplification of the IP material (data not shown). The amplified DNA was labeled using the Bioprime random primer labeling kit (Invitrogen) and purified using a PCR purification kit (Qiagen). The immunoenriched DNA was labeled with Cy5 fluorophore and the input DNA was labeled with Cy3 fluorophore (Amersham PA55021 and PA53021, respectively). The labeled DNA was combined (3-5 μ g each of the IP-DNA and input-DNA) and hybridized to the genomic array described above for 18 hours at 42°C using a TECAN HS400 hybridization machine. The arrays were then washed in 6X SSC, 0.005% Triton-X (25°C, 10 min) and 0.1X SSC, 0.005% Triton-X (4°C, 10 min). The washed slides were scanned using a GenePix 4000B scanner (Axon) and the data were collected using the GenePix Pro 6.0 software (Axon).

Several quality controls and improved DNA microarray technology were used to ensure that this protein-DNA interaction study was of high quality and the false positive error rate was minimized. First, we performed controls to determine the IP efficiency, the average size of the sonicated DNA, and the efficiency of DNA amplification and labeling. Second, to control for non-specific binding, preferential amplification, and preferential hybridization, we performed control experiments where a mock IP was hybridized to the promoter microarray. Finally, dye-swapping experiments were performed to control for sample handling and labeling biases (data not shown). Our ChIP-chip experiments had false positive error rates of less than 10 percent, which are similar to, if not lower, than those reported for other ChIP-chip analyses (21, 24, 25, 44).

Statistical analysis for ChIP-chip and ChIP-expression analyses

Statistical analysis was performed using the statistical software R (GNU project, Free Software Foundation) and all the scripts that were used are available in the supplemental material. After filtering out the spots flagged for bad quality, the signal ratio of IP-DNA over Input-DNA was \log_2 transformed and normalized based on a set of control genes, as described in the array design and in the supplemental materials. For each factor studied, the fold change between untreated and E2-treated cells was calculated and \log_2 transformed. Analysis of

variance was performed across all the replicates and a nominal p-value threshold of $p < 0.05$ was used to select promoters for further analyses. In addition to the p-value threshold, a \log_2 fold change threshold of ± 0.37 was used to select target promoters. Promoters that passed both the p-value and fold change thresholds were classified as target sites where the factor occupancy changed upon E2-treatment. Based on standard ChIP-qPCR experiments, our estimated false positive error rate using the thresholds described above was $< 9\%$ (genes tested = 30). To identify target promoters where ER α or RNA Pol II were localized in both untreated and E2-treated cells, we performed a median percentile rank analysis as described previously (5). Hierarchical clustering of the data was performed using the Cluster software and visualized using the Treeview software (9). For the correlation analyses, Spearman rank correlation coefficients were calculated using data that passed the p-value threshold as described above.

ChIP-qPCR

For qPCR analyses of the ChIP material, 2 μ l out of 50 μ l of ChIP DNA, 1x SYBR Green PCR Master Mix, and 250 nM of forward and reverse primer were used in 40 cycles of amplification (95°C for 15 sec, 60°C for 1 min) using a DNA Engine Opticon Detection System (MJ Research) following an initial 10 min incubation at 95 °C. Melting curve analysis was performed to ensure that only the targeted amplicon was amplified. Mock IP (i.e., non-specific ChIP) values were subtracted from the ER α and Pol II IP values to generate the values shown in the graphs. The sequences of the primers used for the ChIP-qPCR are as follows:

BCAS1 forward	5'-TGGACACCAGAAATCAGGTAACC-3'
BCAS1 reverse	5'-GCCCTTTCCAACATCATGGTA-3'
BCL2 forward	5'-GCAACGATCCCATCAATCTTC-3'
BCL2 reverse	5'-CCCGCGACTCCTGATTCA-3'
BLNK forward	5'-CTATTGGGCAGGCAAATTTCTT-3'
BLNK reverse	5'-TGACCAGCCAGAACTTGTCTTA-3'
C21orf88 forward	5'-TTCACGGAGCACCAGGAATT-3'
C21orf88 reverse	5'-GGCCAGACAGAACTCAAAGAAC-3'
CA12 forward	5'-GAAATGAAATTTGGCCTACAGAGAA-3'
CA12 reverse	5'-TGCCTTTGAATCTGGAGGAAA-3'
CAP2 forward	5'-AGTCTTGATGTCTCGCTAAAGCATT-3'
CAP2 reverse	5'-CCAGACCCAATATGAGGATAAAGG-3'
CASP7 forward	5'-AGCTGGGATCCAGGTTTCG-3'
CASP7 reverse	5'-CCTGGGCCTTACATCAACT-3'
CYP1B1-Prom1 forward	5'-CGTGCGGCCTCGATTG-3'
CYP1B1-Prom1 reverse	5'-AGGTGCCACGTTTCCATT-3'
CYP1B1-Prom2 forward	5'-ACGTTTCCATTGTGCGGTAAC-3'
CYP1B1-Prom2 reverse	5'-TTGTACCGAGCGTGGTTCTG-3'
EBAG9 forward	5'-ACCGGCCCGGAAAGC-3'
EBAG9 reverse	5'-GGCAGTAAGGTCACGCAAATTT-3'
ESR1 forward	5'-GCACATAAGGCAGCACATTAGAGA-3'
ESR1 reverse	5'-TGGGCTTAAAATAAACGCGAAA-3'
FOS forward	5'-CCCTTACACAGGATGTCCATATTAGG-3'
FOS reverse	5'-CCCAGGGCTACAGGGAAAG-3'
ID1 forward	5'-TCCGTTCGGGCCTCAAT-3'
ID1 reverse	5'-TTTTTCCCATATTCACTTTCTCACTTC-3'

MGC11242 forward	5'-GGACCACCACCTGGATGTTG-3'
MGC11242 reverse	5'-AGCCTGCCAATGGAAGCA-3'
MMP2 forward	5'-GCACTATACGAGGCCAAGTTAAGG-3'
MMP2 reverse	5'-GACAGTGGAAGGTCCCAGGTT-3'
MYC forward	5'-TTTGCGGTGGGCAGAAA-3'
MYC reverse	5'-TCTGGCTCACACAGGCGATAT-3'
NBPF15 forward	5'-AGCCCCAAATCTCAATATTAACCA-3'
NBPF15 reverse	5'-TGCATCTAAACAGGATTTGCTTTC-3'
NRIP1 forward	5'-GACTTGGCGGCTTTGGAA-3'
NRIP1 reverse	5'-TCCTGTGGGTTAATGCATACTGTT-3'
NRIP-Enh3 forward	5'-AGGAAGACAGAAATGCTTAACAAGTTG-3'
NRIP-Enh3 reverse	5'-GTGCTTGCGGCTTTTCTGA-3'
P2RX7 forward	5'-TGGAAGCTCCCAGTCTTGTGA-3'
P2RX7 reverse	5'-CACTTTTTTGGTCTCATGTCTCTTG-3'
PCP4 forward	5'-CGACCTAACGTGCCGTTCTC-3'
PCP4 reverse	5'-CAAGCCTGGCCGACATCT-3'
PDZK1 forward	5'-CCTCAGTGATTGAGGATACTCCTGTA-3'
PDZK1 reverse	5'-CAACCCCCAATCCTGTTAGGAT-3'
PPARG forward	5'-CGAAGTCGCTGCTTTGTG-3'
PPARG reverse	5'-GTTAGGCTCCCAGGAGGAGT-3'
PRUNE forward	5'-ACATACACATTTGTTTACCGAACGA-3'
PRUNE reverse	5'-TCCGCAATGTCCCTAGCAA-3'
SAMSN1 forward	5'-CCCATGTAAATGTTGAAGTCATATCCT-3'
SAMSN1 reverse	5'-TGCTGTTACAGAAACAACCTGAAAACAG-3'
SERPINA1 forward	5'-TGGAGGAGGAATGAAGAAAGCA-3'
SERPINA1 reverse	5'-AGCAGGACCCCCAAATTCTGA-3'
SLC27A2 forward	5'-CACGCCTGCAATATCTCCTTTAAT-3'
SLC27A2 reverse	5'-CACGGTTTCTTTAAATGGTGATGA-3'
SLC7A4-Enh1 forward	5'-GGCAGGCCAAACACACATG-3'
SLC7A4-Enh1 reverse	5'-GCCCTGGACACAAACTGCAT-3'
SLC7A4-Enh2 forward	5'-CATCTTGAACCTCTCTTGACATGGTAAA-3'
SLC7A4-Enh2 reverse	5'-GGTGACAGGTTTTTGTGTTGCTTT-3'
SLC7A4 forward	5'-TGAGCTCCGGTAACGACCTT-3'
SLC7A4 reverse	5'-GCTGCCACACACCCAGATATTT-3'
SLC9A8 forward	5'-CTGCACAGTCCTGCCAAGTC-3'
SLC9A8 reverse	5'-CACCCGAGCCTTAGTTTCCTT-3'
SMAP forward	5'-CCACACAGCGCGAATTGA-3'
SMAP reverse	5'-CTGGCCCGAATAAAAAAGATGT-3'
TFF1 forward	5'-CCTGGATTAAGGTCAGGTTGGA-3'
TFF1 reverse	5'-TCTTGGCTGAGGGATCTGAGA-3'
TFF1+0.5Kb forward	5'-GAGTCAAAGCCACCTCTCTCATG-3'
TFF1+0.5Kb reverse	5'-CCACTTGACAACACCAGGGAAT-3'
TFF2 forward	5'-TGGCATGACCCAACCTCTCTGT-3'
TFF2 reverse	5'-CGGGCCAGGGTGACTCT-3'
TMRPSS3 forward	5'-CTTGGCATTGTGATAGGTGGAA-3'
TMRPSS3 reverse	5'-CAACGGCTTGCATCAAATAAC-3'

UGT2B15 forward	5'-TGAAGTGTACACACTAATTGGTGAGTCA-3'
UGT2B15 reverse	5'-TCGTGGTGCAAGTAATGTCTTCTAA-3'
WISP2 forward	5'-TTCCTTATGGGATTGTTGTGCAT-3'
WISP2 reverse	5'-TGCATGGGACATTCACCAAA-3'
XBP1 forward	5'-ACACGCTCCCGCACGTAA-3'
XBP1 reverse	5'-GCAATCCCTGGCCAAAGG-3'

DNA sequence analyses

The genomic sequence for every region on the microarray was derived from the DBTSS and the UCSC genome browser (see Supplemental Materials). Each DNA sequence was scanned for the presence of ERE and AP-1 binding elements. Position weight matrices were obtained from the TRANSFAC database (29) [accession numbers: M00174 (19) and M00517 (33) respectively], and were converted to probability models. Pseudocounts were introduced to avoid over-fitting the motif models, which were based on relatively limited training datasets. The sequence of all spotted DNA regions was analyzed to generate a third-order Markov background model. Log odds scores for all potential motifs, x , of length k were calculated in the standard way (47) using the formula:

$$S = \log \frac{P(x | \Theta_m)}{P(x | \Theta_{bg})}$$

where θ_m represents the motif model and θ_{bg} represents the background model. A threshold score for p -value $< 1.5 \times 10^{-4}$ was used to distinguish genuine binding sites from those present by chance. This threshold was determined empirically by scoring for binding sequences within 1,000,000 base pairs of simulated sequence based on the third-order background model. For each region on the microarray, the highest-scoring binding element was picked, and the binding elements were sorted into classes I-IV, as defined by the ChIP-chip experiments. Relative motif scores (Fig. 3B, *bottom*) were calculated by mean centering the motif scores for each transcription factor binding element studied. The Mann-Whitney test was applied to compare the motif scores across different gene classes and determine whether a motif was enriched in one class versus another one.

Transient transfection reporter gene assays

MCF-7 cells were maintained in phenol red-free MEM supplemented with 5% charcoal-dextran stripped calf serum for at least 3 days before using for the experiments. The cells were plated in 6-well plates for 24 h and reached ~70% confluence prior to transfection using GeneJuice transfection reagent (Novagen). Each well received the following combination of plasmid DNAs: (1) 1 μ g of a luciferase reporter construct containing either the native UGT2B15 promoter (spanning the region from -449bp to +114bp) or the native UGT2B15 promoter with a deletion of the AP-1 binding element identified by our bioinformatics analysis (ATGACTCAC, -418bp to -409bp; see Supplemental Fig. 3), and (2) 600 ng of pCMV β , a constitutive β -galactosidase expression vector used to normalize for transfection efficiency. Twelve hours after transfection, the cells were treated with vehicle or 10 nM E2 for an additional 16 h. Luciferase activity was measured in extracts from the transfected cells using a 96-well plate luminometer (Beckman Coulter LD400) and normalized to β -galactosidase activity measured in the same

extracts using the plate reader. To ensure reproducibility, the experiments were performed in triplicate.

RNA extraction and RT-qPCR

MCF-7 cells were maintained in phenol red-free MEM supplemented with 5% charcoal-dextran stripped calf serum for at least 3 days before using for the experiments. After growing to ~60% confluence (~6 x 10⁵ cells), the cells were treated with ethanol or 100 nM E2 for 1 h or 3 h. Total RNA was isolated using RNeasy columns and treatment with RNase-free DNase I (Qiagen) following the manufacturer's specifications. First strand cDNA synthesis was performed using 300 ng of total RNA, 1 µg oligo(dT), and 400 units of MMLV reverse transcriptase (Promega) in a 50 µL reaction, as described previously (1). The resulting cDNA from each samples was then diluted 1:10 and analyzed by real-time PCR. Each real-time PCR reaction consisted of 2 µL cDNA, 1x SYBR Green PCR Master Mix, and 0.25 µM each of forward and reverse primer for a total reaction volume of 20 µL. Reactions were carried out in duplicates using a 96-well DNA Engine Opticon (MJ Research) or a 384-well Prism 7700 (ABI) real-time PCR thermocycler for 45 cycles (95°C for 15 sec, 60°C for 1 min) following an initial 10-minute incubation at 95°C. The fold change in expression of each gene was calculated using a standard curve of diluted cDNA from untreated samples (1:1, 1:10, 1:100) and normalized against the fold change of β-actin, a well-characterized housekeeping gene used as internal control. Melting curve analysis was performed to ensure that only the targeted amplicon was amplified. The sequences of the primers used for real-time PCR were designed based on exon-exon junctions using AutoPrime software (<http://www.autoprime.de/AutoPrimeWeb>) and are as follows:

ACTB forward	5'-AGCTACGAGCTGCCTGAC-3'
ACTB reverse	5'-AAGGTAGTTTCGTGGATGC-3'
AHSA1 forward	5'-TTACTGATCTGGTCCCTGAG-3'
AHSA1 reverse	5'-GTCGATGAAGGTCAAGGTG-3'
BCAS1 forward	5'-TGGACACGAACTCACTGC-3'
BCAS1 reverse	5'-AACTGGTCCGATGGATACTG-3'
BCL2 reverse	5'-GGGCCAAACTGAGCAGAG-3'
BCL2forward	5'-ATAACGGAGGCTGGGATG-3'
BLNK forward	5'-CAAGAAGCTGTGCAGTCAC-3'
BLNK reverse	5'-ATCCACAGTTGGGTTTCC-3'
BTG2 forward	5'-GCCCCTTAGTGAGGAATCTTCAC-3'
BTG2 reverse	5'-TGAGCAGTTTAGAAATCCAGGAAAG-3'
C11orf8 forward	5'-TTCAAAGAGTGGGCTGTG-3'
C11orf8 reverse	5'-TCGAGGCATTGATGTACG-3'
C18orf24 forward	5'-TCCAGCTTTATGAAGTCAGCAAGT-3'
C18orf24 reverse	5'-GATACAGATGAACAACAGCCACATG-3'
CA12 forward	5'-CACTGCCAGCAACAAGTC-3'
CA12 reverse	5'-ACATGTTGAAGGTGACTGAAG-3'
CAP2 forward	5'-GTGGAAAGACTGGAACGAG-3'
CAP2 reverse	5'-CCATCAGCTTGTCAAAGG-3'
CASP7 forward	5'-GGTATGGGCGTTCGAAACG-3'
CASP7 reverse	5'-AGGCTTCGGAAGCACTTGAA-3'
CYP1B1 forward	5'-AACGTACCGGCCACTATC-3'

CYP1B1 reverse	5'-CACGACCTGATCCAATTC-3'
EBAG9 forward	5'-CGGAAATTAAGTGGAGACC-3'
EBAG9 reverse	5'-TCCTTCGATCTTTACACTGG-3'
ESR1 forward	5'-CACCCAGGGAAGCTACTG-3'
ESR1 reverse	5'-CGGAACCGAGATGATGTAG-3'
FOS forward	5'-TTGTGAAGACCATGACAGG-3'
FOS reverse	5'-GCTGCAGCCATCTTATTC-3'
HOXC5 forward	5'-GACCCAGCAAGTGGTCCTA-3'
HOXC5 reverse	5'-GGGTCAGGAGGGCACAGA-3'
ID1 forward	5'-GAATCCGAAGTTGGAACC-3'
ID1 reverse	5'-CTTCAGCGACACAAGATG-3'
MAP2K6 forward	5'-CTCGAGATTTAGACTCCAAGG-3'
MAP2K6 reverse	5'-TCGTCCCAGTTCCATTATAG-3'
MYC forward	5'-GGATTTTTTTTCGGGTAGTGGAA-3'
MYC reverse	5'-TCCTGTTGGTGAAGCTAACGTT-3'
NBPF15 forward	5'-GTGGCTGAGAAAGTGCAG-3'
NBPF15 reverse	5'-GCTGGTTGGAGTCATAAGG-3'
NRIP1 forward	5'-CCAGAAAGCGATTCCTTC-3'
NRIP1 reverse	5'-CCATTACGGAGAGATCCAC-3'
P2RX7 forward	5'-CCTGTGTGGTCAACGAATAC-3'
P2RX7 reverse	5'-TGGTTCACCATCCTAATGTG-3'
PCM1 forward	5'-TGAAGATCATGGCTCACC-3'
PCM1 reverse	5'-TTGGGACCATCATAACACC-3'
PDZK1 forward	5'-GTGGCTATGAGAGCTGGAG-3'
PDZK1 reverse	5'-CCAGCAGGAACATGACAC-3'
PRUNE forward	5'-AAGCCTGTGATTTGGACTC-3'
PRUNE reverse	5'-CACCTCGCAGAGGTAGTTC-3'
RARG forward	5'-TGGATGACACCGAGACAG-3'
RARG reverse	5'-GTACAGCCTCAGGGCTTC-3'
RENT1 forward	5'-CCAAAGCTGACTCAGTGG-3'
RENT1 reverse	5'-ATCTTGACCAGCCAGGAC-3'
SERPINA1 forward	5'-ACAGGACGCTGTGGTTTC-3'
SERPINA1 reverse	5'-GTCCTCGTCCGTATTTAAGC-3'
SLC27A2 forward	5'-TATCGGTGAACTGCTTCG-3'
SLC27A2 reverse	5'-TCCACACATCTCCTCGTAAG-3'
SLC9A8 forward	5'-AAGGACGTCAACCTCAGC-3'
SLC9A8 reverse	5'-ATGTAGTGGGCCTCGTACTC-3'
SMAP forward	5'-CTGCCAGAGAGTCTCACC-3'
SMAP reverse	5'-TTCCTGCACCCATAAGTC-3'
TFF1 forward	5'-TGCTTCTATCCTAATACCATCG-3'
TFF1 reverse	5'-AGATCCCTGCAGAAGTGTC-3'
WISP2 forward	5'-GAGAGGCACACCGAAGAC-3'
WISP2 reverse	5'-GGCAGGTACATGGTGTCG-3'

Expression Microarray Analyses

MCF-7 cells were grown as described above and treated with ethanol or 100 nM E2 for 3h. Total RNA was prepared using Trizol Reagent according to the manufacturer's instructions

(Invitrogen) and was further purified using RNeasy columns (Qiagen). Seven μg of total RNA were subjected to One-Cycle Target Labeling Assay (Affymetrix) to generate biotinylated cRNA targets for hybridization to Affymetrix U133A 2.0 microarray. The raw data were processed by Affymetrix GCOS software to obtain detection calls and signal values, and then normalized by scaling. Only probe sets having "present" calls on at least two of the three arrays were included for further analysis; those signals were \log_2 -transformed and median-centered. The t-test was applied to the normalized data matrix to identify differential genes between the E2-treated and untreated control samples. A significance level cutoff of 0.05 was applied to select each differential gene set and gene regulation was considered significant if the observed change was greater than 2-fold. Three independent experiments were performed for the microarray analyses. Hierarchical clustering of the data was performed using the Cluster software and visualized using the Treeview software (9).

2. Supplemental Figure Legends

Figure S1. Factor occupancy at the MYC and CASP7 genes in the absence or presence of E2.

ChIP-chip tiling for the MYC (class IIB, panels A-C) and CASP7 (class IIA; panels D-F) genes in MCF-7 cells. Occupancy, expressed as ChIP-enrichment ratios (IP/input), for ER α (red), RNA Pol II (black), acetylated histones (blue), and SRC (green) throughout the indicated genomic region is shown in the absence (dotted line) or presence (solid line) of a 45 min. treatment with E2. The X-axis represents the location along the gene expressed as distance from the transcription start site (TSS, "0") in kb.

Figure S2. Gene-by-gene validation of RNA Pol II and ER α promoter occupancy by qPCR.

Validation by quantitative real-time PCR of the ChIP-chip results for selected promoters from the estrogen-regulated promoter array. Results for ER α (panel A) and RNA Pol II (panel B) are shown. Mock IP (i.e., non-specific ChIP) values were subtracted from the ER α and Pol II IP values to generate the values shown. The empty and filled bars represent ChIP-enrichment for untreated (U) and E2-treated (E) cells, respectively. The class IIA, IIB, IIIA, IVA, and IVB promoters are as described in Fig. 2C. Each bar represents the mean + SEM for at least two separate determinations. False positive results (i.e., DNA sites where ChIP-chip showed factor binding, but ChIP-qPCR did not confirm it) are labeled with an asterisk (*). False negative results (i.e., DNA sites where ChIP-chip predicted no factor binding, but ChIP-qPCR showed factor binding) are labeled with two asterisks (**). Overall, our false positive rates from the ChIP-chip experiments were less than 10%, which are similar to, if not lower, than those reported for other ChIP-chip analyses (21, 24, 25, 44). The false negative rates were about 30%, but were only determined for a very small set of genes (~10).

Figure S3. Bioinformatic sequence analysis for selected ER α -binding promoters.

Schematics of DNA probes for EBAG9, an "ERE only" (i.e., ERE-positive/AP-1 site-negative) promoter, and UGT2B15, an "AP-1 only" (i.e., ERE-negative/AP-1 site-positive) promoter, as defined in Figs. 3B, 3C, and 3D. The sequences and locations of the ERE and AP-1 site relative to the TSS are shown. "-" and "+" indicate the regions upstream and downstream of the TSS, respectively. The primers used for ChIP-qPCR in Fig. 3D are also shown (green arrows under the probes).

Figure S4. Gene-by-gene validation of E2-dependent changes in Pol II occupancy and acetylated histone levels by qPCR.

Validation by quantitative real-time PCR of the ChIP-chip results for selected promoters from the estrogen-regulated promoter array. The graphs show the fold change in Pol II occupancy (panel A) and AcH levels (panel B) upon treatment with E2 for a selected subset of genes. The fold changes are log₂ transformed: log₂ fold > 0 represents increased occupancy/levels after E2 treatment, whereas log₂ fold < 0 represents decreased occupancy/levels after E2 treatment. Most (25 out of 30; 83%) of the regions tested showed the same trend for Pol II and AcH. The regions

showing different trends are labeled with an asterisk (*). Classes IIA, IIB, IIIA, IVA, IVB and IA are as described in Fig. 2C. Each bar represents the mean + SEM for at least two separate determinations.

Figure S5. Gene-by-gene validation of E2-dependent changes in ER α and SRC occupancy by qPCR.

Validation by quantitative real-time PCR of the ChIP-chip results for selected promoters from the estrogen-regulated promoter array. The graphs show the fold change in ER α (panel A) and SRC (panel B) occupancy upon treatment with E2 for a selected subset of genes. The fold changes are log₂ transformed: log₂ fold > 0 represents increased occupancy after E2 treatment, whereas log₂ fold < 0 represents decreased occupancy after E2 treatment. Most (26 out of 30; 87%) of the regions tested showed the same trend for ER α and SRC. The regions showing different trends are labeled with an asterisk (*). Classes IIA, IIB, IIIA, IVA, IVB and IA are as described in Fig. 2C. Each bar represents the mean + SEM for at least two separate determinations.

Figure S6. E2-dependent gene expression correlates with Pol II recruitment and acetylation of histones H3 and H4 at gene promoters.

(A) Correlation analysis of E2-dependent changes in AcH levels at promoters and expression of the associated genes in MCF-7 cells. Increased AcH levels correlate with increased expression of the associated genes, and vice versa (correlation coefficient = 0.43, p = 0.0001). All promoters showing significant changes (p < 0.05) for RNA expression and AcH levels were included in the analysis. The fold changes are log₂ transformed: log₂ fold > 0 represents an increase upon E2 treatment and log₂ fold < 0 represents a decrease upon E2 treatment.

(B) Correlation analysis of gene expression, as measured by qPCR, and factor recruitment to promoters upon E2 treatment in MCF-7 cells. Spearman rank correlation coefficients and associated p-values were calculated for the fold changes in RNA expression (after a 1h or 3h treatment with E2) occupancy by ER α , Pol II, SRC, and AcH. All genes with significant data (p < 0.05) for RNA expression and factor occupancy were included in the analysis. The top line in each row shows the correlation coefficient for each condition, while the bottom line in each row shows the p-value (*italics*). The strongest correlations were between Pol II recruitment and RNA expression after 3h of E2 treatment, as well as between acetylated histone levels and RNA expression after 3h of E2 treatment.

(C) Gene-by-gene confirmation of RNA expression for a subset of target genes in MCF-7 cells. MCF-7 cells were treated with vehicle control or E2 for 1h or 3h. RNA was isolated, reverse-transcribed, and analyzed by qPCR for expression of genes in classes IA, IIA, IIB, IIIA, and IV (as defined in Fig. 2C). The graph shows the fold change in RNA expression after 1h (white bars) or 3h (black bars) of E2-treatment. The fold changes are log₂ transformed: log₂ fold > 0 represents an increase upon E2 treatment and log₂ fold < 0 represents a decrease upon E2 treatment. Each bar represents the mean + SEM for at least two separate determinations.

3. Supplemental Table 1 Legend

Supplemental Table 1. Comparison of factor occupancy and gene expression data between Kininis et al., 2007 and other published ChIP-chip studies.

- Factor occupancy and gene expression data for the 58 direct E2 target genes that we identified are shown.
- Direct E2 target genes are defined as genes that show E2-dependent changes in Pol II promoter occupancy and gene expression. For more details, see the paragraph entitled “Definitions and Data Used to Derive the Conclusions Presented” in Section 8 of the Supplemental Materials.
- For each direct E2 target gene, the following information is shown:
 - Column A: Probe ID (as assigned in the file “List of promoter sequences.xls”, which is available upon request)
 - Column B: Gene Name
 - Column C: Gene Symbol
 - Column D: RefSeq accession number
 - Column E: Class (as assigned in Figure 2C)
- Available data for: (1) occupancy by RNA polymerase II (“Pol II”), ER α (“ER”), SRC, and acetylated histones (“AcH”) in the presence and absence of 17 β -estradiol (E2) and (2) gene expression in response to E2 are presented for the following studies: Kininis et al., 2007 (current study), Laganieri et al., 2005 (24), Carroll et al., 2005 and 2006 (6, 7), Cheng et al., 2006 (8), Frasor et al., 2003(10), and various gene-specific (*i.e.*, not microarray-based) studies (2, 3, 12-16, 18, 23, 28, 30, 32, 35, 37, 41, 42, 45, 46, 48).

1. Columns G through O: Kininis et al., 2007 (current study)

The binding of Pol II, ER α , SRC, and AcH at the indicated promoter regions in MCF-7 cells was determined by ChIP-chip and ChIP-qPCR in the absence (untreated; “U”) or presence (E2-treated; “E”) of a 45 min. treatment with E2. “+” indicates the presence of a factor; “-” indicates the absence of a factor, as described in the Supplemental Materials and Methods (Section 1). For the gene expression data, “up” indicates a ≥ 1.5 -fold increase following E2 treatment, “dn” indicates ≥ 1.5 -fold decrease following E2 treatment, “n.d.” indicates that the gene was not included on the microarray used, and a blank cell indicates <1.5 -fold change following E2 treatment.

2. Columns Q through Y: Laganieri et al., 2005 (24)

The binding of ER α at the indicated promoter regions in MCF-7 cells was determined by ChIP-chip in the presence, but not the absence, of a 45 min. treatment with E2 (“E”). ChIP-qPCR was used to provide gene-specific data for ER α , Pol and SRC3 (a.k.a. AIB1) in the absence (“U”) or presence of E2. “+” indicates the presence of a factor; “-” indicates the absence of a factor, and “n.d.” indicates that factor occupancy or gene expression were not determined in the study.

3. Columns AA through AI: Carroll et al., 2005 and 2006 (6, 7)

The binding of Pol II and ER α at the indicated promoter regions in MCF-7 cells was determined

by ChIP-chip in the presence, but not the absence, of a 45 min. treatment with E2 (“E”). ChIP-qPCR was used to provide gene-specific data for ER α , Pol II and SRC3 (a.k.a. AIB1) in the absence (“U”) or presence of E2. Data from both Carroll *et al.*, 2005 (6) and Carroll *et al.*, 2006 (7) are combined. “+” indicates the presence of the a factor within 10 Kb of the designated promoter, “-” indicates the absence of a factor within 10 Kb of the designated promoter, and “n.d.” indicates that occupancy of the factor was not determined in the study. For the gene expression data, “up” indicates an increase following E2 treatment, “dn” indicates a decrease following E2 treatment, and blank cells indicate that there was no regulation by E2 or the associated genes were not included on the microarray used.

4. Columns AK through AS: Cheng *et al.* (8)

The binding ER α and the presence of Ac-H3K9 at the indicated promoter regions in MCF-7 cells were determined by ChIP-chip in the absence (“U”) and presence (“E”) of a 3 hour treatment with E2. ChIP-qPCR was used to provide gene-specific data for ER α , Pol II, SRC3, Ac-H3K9 in the absence and presence of E2. “+” indicates the presence of a factor at the designated promoter, “-” indicates the absence of a factor at the designated promoter, “n.d.” indicates that factor occupancy or regulation of gene expression were not determined in this study, and blank cells indicate either the absence of a factor or that the designated promoters were not included on the CpG island microarray used.

5. Column AU: Frasor *et al.* (10)

The change in gene expression of the indicated genes upon a 4 hour treatment with E2 was determined by gene expression microarrays. “up” indicates a ≥ 1.5 -fold increase following E2 treatment, “dn” indicates ≥ 1.5 -fold decrease following E2 treatment, “n.d.” indicates that the gene was not included on the microarray used, and blank cells indicate <1.5 -fold change following E2 treatment.

6. Column AW: Gene-specific studies

Gene-specific (i.e., not microarray-based) gene expression studies showing regulation of the indicated genes by E2 or ER α are listed (2, 3, 12-16, 18, 23, 28, 30, 32, 35, 37, 41, 42, 45, 46, 48).

7. Conclusions from this comparison

As this tabular comparison reveals, our ER α ChIP-chip results after a 45 min. treatment with E2 generally agree with the results from Laganier *et al.*, 2005 (24) and Carroll *et al.*, 2005 and 2006 (6, 7) (81% and 65% agreement, respectively). In addition, our ChIP-chip results for Pol II after a 45 min. treatment with E2 show 53% agreement with the Pol II ChIP-chip results from Carroll *et al.*, 2005 and 2006 (6, 7), as well as 48% agreement with the microarray expression results after a 4 h treatment with E2 from Frasor *et al.*, 2003 (10). The good agreement between our results and those of Laganier *et al.*, 2005 (24) is likely due to the similarities in the microarray platforms used. Likewise, the differences between our results and those of Carroll *et al.*, 2005 and 2006 (6, 7) and Frasor *et al.*, 2003 (10) are likely due to differences in the microarray platform used and the length of E2 treatment, respectively. Direct comparisons of the Cheng *et al.*, 2006 (8) study to the other ChIP-chip studies could not be made because many of the designated promoters were not included on their CpG island microarray and because they

used a 3 hour treatment with E2. As indicated, some of the direct E2 target genes that we defined in our studies have been examined in previous gene-specific studies.

This tabular comparison highlights the fact that no previous ChIP-chip study has determined the localization of Pol II and ER α at E2 target promoters in both the absence and presence of a short (45 min.) E2 treatment. Comparisons between the -E2 and +E2 conditions in our current study allowed us to identify a set of promoters constitutively bound by ER α , a set of E2-repressed genes with low occupancy by SRC, and a set of E2-regulated genes with constitutively bound and possibly promoter-proximal paused Pol IIs. Finally, this tabular comparison also highlights the fact that no previous ChIP-chip study has determined the global localization of SRC at E2 target promoters either in the presence or absence of E2 treatment. To our knowledge, our current study is the first to examine the ligand-dependent localization of a non-DNA binding nuclear receptor cofactor on a global scale.

4. Supplemental References

1. **Acevedo, M. L., K. C. Lee, J. D. Stender, B. S. Katzenellenbogen, and W. L. Kraus.** 2004. Selective recognition of distinct classes of coactivators by a ligand-inducible activation domain. *Mol Cell* **13**:725-38.
2. **Alkayed, N. J., S. Goto, N. Sugo, H. D. Joh, J. Klaus, B. J. Crain, O. Bernard, R. J. Traystman, and P. D. Hurn.** 2001. Estrogen and Bcl-2: gene induction and effect of transgene in experimental stroke. *J Neurosci* **21**:7543-50.
3. **Banerjee, S., N. Saxena, K. Sengupta, O. Tawfik, M. S. Mayo, and S. K. Banerjee.** 2003. WISP-2 gene in human breast cancer: estrogen and progesterone inducible expression and regulation of tumor cell proliferation. *Neoplasia* **5**:63-73.
4. **Bourdeau, V., J. Deschenes, R. Metivier, Y. Nagai, D. Nguyen, N. Bretschneider, F. Gannon, J. H. White, and S. Mader.** 2004. Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol Endocrinol* **18**:1411-27.
5. **Buck, M. J., and J. D. Lieb.** 2004. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**:349-60.
6. **Carroll, J. S., X. S. Liu, A. S. Brodsky, W. Li, C. A. Meyer, A. J. Szary, J. Eeckhoute, W. Shao, E. V. Hestermann, T. R. Geistlinger, E. A. Fox, P. A. Silver, and M. Brown.** 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**:33-43.
7. **Carroll, J. S., C. A. Meyer, J. Song, W. Li, T. R. Geistlinger, J. Eeckhoute, A. S. Brodsky, E. K. Keeton, K. C. Fertuck, G. F. Hall, Q. Wang, S. Bekiranov, V. Sementchenko, E. A. Fox, P. A. Silver, T. R. Gingeras, X. S. Liu, and M. Brown.** 2006. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet.*
8. **Cheng, A. S., V. X. Jin, M. Fan, L. T. Smith, S. Liyanarachchi, P. S. Yan, Y. W. Leu, M. W. Chan, C. Plass, K. P. Nephew, R. V. Davuluri, and T. H. Huang.** 2006. Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters. *Mol Cell* **21**:393-404.
9. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**:14863-8.
10. **Frasor, J., J. M. Danes, B. Komm, K. C. Chang, C. R. Lyttle, and B. S. Katzenellenbogen.** 2003. Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology* **144**:4562-74.
11. **Frasor, J., F. Stossi, J. M. Danes, B. Komm, C. R. Lyttle, and B. S. Katzenellenbogen.** 2004. Selective estrogen receptor modulators: discrimination of agonistic versus antagonistic activities by gene expression profiling in breast cancer cells. *Cancer Res* **64**:1522-33.
12. **Frazier-Jessen, M. R., and E. J. Kovacs.** 1995. Estrogen modulation of JE/monocyte chemoattractant protein-1 mRNA expression in murine macrophages. *J Immunol* **154**:1838-45.
13. **Ghosh, M. G., D. A. Thompson, and R. J. Weigel.** 2000. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. *Cancer Res* **60**:6367-75.

14. **Harrington, W. R., S. Sengupta, and B. S. Katzenellenbogen.** 2006. Estrogen regulation of the glucuronidation enzyme UGT2B15 in estrogen receptor-positive breast cancer cells. *Endocrinology* **147**:3843-50.
15. **Jakowlew, S. B., R. Breathnach, J. M. Jeltsch, P. Masiakowski, and P. Chambon.** 1984. Sequence of the pS2 mRNA induced by estrogen in the human breast cancer cell line MCF-7. *Nucleic Acids Res* **12**:2861-78.
16. **Kahlert, S., S. Nuedling, M. van Eickels, H. Vetter, R. Meyer, and C. Grohe.** 2000. Estrogen receptor alpha rapidly activates the IGF-1 receptor pathway. *J Biol Chem* **275**:18447-53.
17. **Karolchik, D., R. Baertsch, M. Diekhans, T. S. Furey, A. Hinrichs, Y. T. Lu, K. M. Roskin, M. Schwartz, C. W. Sugnet, D. J. Thomas, R. J. Weber, D. Haussler, and W. J. Kent.** 2003. The UCSC Genome Browser Database. *Nucleic Acids Res* **31**:51-4.
18. **Kawakubo, H., J. L. Carey, E. Brachtel, V. Gupta, J. E. Green, P. D. Walden, and S. Maheswaran.** 2004. Expression of the NF-kappaB-responsive gene BTG2 is aberrantly regulated in breast cancer. *Oncogene* **23**:8310-9.
19. **Kel, A., O. Kel-Margoulis, V. Babenko, and E. Wingender.** 1999. Recognition of NFATp/AP-1 composite elements within genes induced upon the activation of immune cells. *J Mol Biol* **288**:353-76.
20. **Kian Tee, M., I. Rogatsky, C. Tzagarakis-Foster, A. Cvorov, J. An, R. J. Christy, K. R. Yamamoto, and D. C. Leitman.** 2004. Estradiol and selective estrogen receptor modulators differentially regulate target genes with estrogen receptors alpha and beta. *Mol Biol Cell* **15**:1262-72.
21. **Kirmizis, A., S. M. Bartley, A. Kuzmichev, R. Margueron, D. Reinberg, R. Green, and P. J. Farnham.** 2004. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev* **18**:1592-605.
22. **Klinge, C. M.** 2001. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res* **29**:2905-19.
23. **Labhart, P., S. Karmakar, E. M. Salicru, B. S. Egan, V. Alexiadis, B. W. O'Malley, and C. L. Smith.** 2005. Identification of target genes in breast cancer cells directly regulated by the SRC-3/AIB1 coactivator. *Proc Natl Acad Sci U S A* **102**:1339-44.
24. **Laganier, J., G. Deblois, C. Lefebvre, A. R. Bataille, F. Robert, and V. Giguere.** 2005. Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc Natl Acad Sci U S A* **102**:11651-6.
25. **Lee, T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young.** 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**:799-804.
26. **Levenson, A. S., K. M. Svoboda, K. M. Pease, S. A. Kaiser, B. Chen, L. A. Simons, B. D. Jovanovic, P. A. Dyck, and V. C. Jordan.** 2002. Gene expression profiles with activation of the estrogen receptor alpha-selective estrogen receptor modulator complex in breast cancer cells expressing wild-type estrogen receptor. *Cancer Res* **62**:4419-26.
27. **Lin, C. Y., A. Strom, V. B. Vega, S. L. Kong, A. L. Yeo, J. S. Thomsen, W. C. Chan, B. Doray, D. K. Bangarusamy, A. Ramasamy, L. A. Vergara, S. Tang, A. Chong, V. B.**

- Bajic, L. D. Miller, J. A. Gustafsson, and E. T. Liu.** 2004. Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome Biol* **5**:R66.
28. **Loose-Mitchell, D. S., C. Chiappetta, and G. M. Stancel.** 1988. Estrogen regulation of c-fos messenger ribonucleic acid. *Mol Endocrinol* **2**:946-51.
29. **Matys, V., E. Fricke, R. Geffers, E. Gossling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, D. U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Munch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, and E. Wingender.** 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* **31**:374-8.
30. **Miyagawa, S., Y. Katsu, H. Watanabe, and T. Iguchi.** 2004. Estrogen-independent activation of erbBs signaling and estrogen receptor alpha in the mouse vagina exposed neonatally to diethylstilbestrol. *Oncogene* **23**:340-9.
31. **Monroe, D. G., B. J. Getz, S. A. Johnsen, B. L. Riggs, S. Khosla, and T. C. Spelsberg.** 2003. Estrogen receptor isoform-specific regulation of endogenous gene expression in human osteoblastic cell lines expressing either ERalpha or ERbeta. *J Cell Biochem* **90**:315-26.
32. **Murphy, L. J., L. C. Murphy, and H. G. Friesen.** 1987. Estrogen induction of N-myc and c-myc proto-oncogene expression in the rat uterus. *Endocrinology* **120**:1882-8.
33. **O'Lone, R., M. C. Frith, E. K. Karlsson, and U. Hansen.** 2004. Genomic targets of nuclear estrogen receptors. *Mol Endocrinol* **18**:1859-75.
34. **Odom, D. T., N. Zizlsperger, D. B. Gordon, G. W. Bell, N. J. Rinaldi, H. L. Murray, T. L. Volkert, J. Schreiber, P. A. Rolfe, D. K. Gifford, E. Fraenkel, G. I. Bell, and R. A. Young.** 2004. Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**:1378-81.
35. **Pratt, M. A., T. E. Bishop, D. White, G. Yasvinski, M. Menard, M. Y. Niu, and R. Clarke.** 2003. Estrogen withdrawal-induced NF-kappaB activity and bcl-3 expression in breast cancer cells: roles in growth and hormone independence. *Mol Cell Biol* **23**:6887-900.
36. **Ren, B., H. Cam, Y. Takahashi, T. Volkert, J. Terragni, R. A. Young, and B. D. Dynlacht.** 2002. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev* **16**:245-56.
37. **Romano, G. J., A. Krust, and D. W. Pfaff.** 1989. Expression and estrogen regulation of progesterone receptor mRNA in neurons of the mediobasal hypothalamus: an in situ hybridization study. *Mol Endocrinol* **3**:1295-300.
38. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**:365-86.
39. **Shang, Y., and M. Brown.** 2002. Molecular determinants for the tissue specificity of SERMs. *Science* **295**:2465-8.
40. **Shang, Y., X. Hu, J. DiRenzo, M. A. Lazar, and M. Brown.** 2000. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**:843-52.
41. **Stevens, T. A., and R. Meech.** 2006. BARX2 and estrogen receptor-alpha (ESR1) coordinately regulate the production of alternatively spliced ESR1 isoforms and control breast cancer cell growth and invasion. *Oncogene* **25**:5426-35.

42. **Stossi, F., V. S. Likhite, J. A. Katzenellenbogen, and B. S. Katzenellenbogen.** 2006. Estrogen-occupied estrogen receptor represses cyclin G2 gene expression and recruits a repressor complex at the cyclin G2 promoter. *J Biol Chem* **281**:16272-8.
43. **Suzuki, Y., R. Yamashita, S. Sugano, and K. Nakai.** 2004. DBTSS, DataBase of Transcriptional Start Sites: progress report 2004. *Nucleic Acids Res* **32**:D78-81.
44. **Thibaud-Nissen, F., H. Wu, T. Richmond, J. C. Redman, C. Johnson, R. Green, J. Arias, and C. D. Town.** 2006. Development of Arabidopsis whole-genome microarrays and their application to the discovery of binding sites for the TGA2 transcription factor in salicylic acid-treated plants. *Plant J* **47**:152-62.
45. **Tozlu, S., I. Girault, S. Vacher, J. Vendrell, C. Andrieu, F. Spyrtos, P. Cohen, R. Lidereau, and I. Bieche.** 2006. Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. *Endocr Relat Cancer* **13**:1109-20.
46. **Tsuchiya, Y., M. Nakajima, S. Kyo, T. Kanaya, M. Inoue, and T. Yokoi.** 2004. Human CYP1B1 is regulated by estradiol via estrogen receptor. *Cancer Res* **64**:3119-25.
47. **Wasserman, W. W., and A. Sandelin.** 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* **5**:276-87.
48. **Watanabe, T., S. Inoue, H. Hiroi, A. Orimo, H. Kawashima, and M. Muramatsu.** 1998. Isolation of estrogen-responsive genes with a CpG island library. *Mol Cell Biol* **18**:442-9.

5. Outline of Statistical Analysis Using the Statistical Software R

Note: Each microarray slide contained four replicates of the same spot in two sectors: the lower and the upper sector, each having two identical spots. Every ChIP-chip experiment was performed on two or three independent biological replicates on a different day and from independent cell cultures (e.g., for the ER α ChIP-chip, we analyzed a total of 12 spots for each gene in 3 independent experiments with 4 spots per microarray slide).

Steps before data analysis:

1. Performed ChIP-chip under two conditions: U = untreated, and E = E2-treated cells.
2. Collected data using GenePix Pro 6.0. The data from the upper and lower sectors on the same slide were presented separately. F635 (*i.e.*, Cy5 labeled IP-DNA) and F532 (*i.e.*, Cy3 labeled input DNA) signals were also presented separately.

Data analysis:

1. BACKHISTO R file

- a) Subtracted background (B) from signal (F) for each spot.
- b) Calculated the ratio of (F635/F532), and \log_2 transformed it.
- c) Flagged a spot as “bad” if F532-B532 < 80.
- d) Labeled the control genes as "isg" and "iskeep", as described below:
 - "isg" were all control (non-E2 regulated) genes and intergenic regions.
 - "iskeep" were all control genes for which the E/U fold fell above the 80% percentile.
- e) Normalized all genes by subtracting the mean of the \log_2 (F635/F532) ratio of "isg&!iskeep" (isg AND NOT iskeep) genes from the \log_2 (F635/F532) ratio of each gene.
- f) Wrote the output table where data from the upper and lower sectors on the same slide were presented separately.

2. POOLER.EXE

Pooled data from upper and lower backhisto files (**step 1**).

3. MEANPVAL R file

Calculated the means of only the “good” spots from **step 2**, so that the output table contained the mean of the data from all the four replicate spots on each slide. Each replicate experiment on a different day was presented separately. This file was the one that was used for the U and E occupancy analysis in **step 6**.

4. ANOVA R file

- a) Subtracted the U \log_2 (F635/F532) ratios from the E \log_2 (F635/F532) ratios from **step 2**. Each spot replicate was treated separately at this step resulting in 12 $\log_2E - \log_2U$ folds (3 independent biological replicates * 4 spots per slide) or 8 $\log_2E - \log_2U$ folds (2 independent biological replicates * 4 spots per slide).
- b) Calculated the mean of the $\log_2E - \log_2U$ folds from all the replicates.
- c) Performed ANOVA for the $\log_2E - \log_2U$ folds across all the replicates and calculated the associated p-value.
- d) Wrote a table with the mean of the $\log_2E - \log_2U$ folds and their associated P-values.

5. Identification of genes that show changes in factor occupancy between U and E conditions

Imported the ANOVA data from **step 4** in Excel and used a p-value threshold of 0.05 and a log₂ fold threshold of ± 0.37 to select target promoters where there is a significant change in factor occupancy between untreated and E2-treated cells. To reduce false positive error results, spots that had p-value < 0.05 for only one of the four factors tested were eliminated as spots showing high overall variability. Based on standard ChIP-qPCR experiments, our estimated false positive error rate using the thresholds described above was < 9% (genes tested = 30, **Figure S2**).

6. Analysis of Factor Occupancy in U and E separately

- a) Imported the output of the meanpval R file (**step 3**) in Excel.
- b) Reversed the log₂ of the data.
- c) Normalized the data so that every set of U and E biological replicates had equal total signal.
- g) Log₂ transformed the normalized U and E values.

This step resulted in a table that contained the normalized values for U and E separately. Every independent biological replicate was presented separately (replicates were not combined).

7. Identification of promoters where a factor is localized in both the U and E conditions

- a) Calculated the mean of the U and E signals for all the independent biological replicates calculated in **step 6**. Only genes with signals in both U and E conditions were included in the analysis.
- b) Assigned percentiles to the U and E data separately.
- d) Performed median percentile analysis as described previously (5) to determine genes that show factor enrichment in U or E separately.
- e) Eliminated all the genes that showed high enrichment in ChIP-chip experiments where the microarray was hybridized with a mock ChIP sample.
- f) All the genes that showed significant factor occupancy at their promoter in both U and E (as determined by **step 7e**) and did not show significant change in factor occupancy between U and E (as determined by **step 5**), were identified as genes with high factor occupancy at both U and E conditions. We tested five of the promoters that fell into this category using standard ChIP-qPCR experiments (**Figure S2**, class IIIA for RNA Pol II). All of them showed high RNA Pol II occupancy at both U and E conditions, confirming our analysis.

6. Description of the ChIP-chip Data Set

The excel file contains 9 worksheets.

Sheets 1-7

All probes in classes IIA, IIB, IIIA, IIIB, IVA, IVB and IA are shown in separate excel sheets. For each probe, the following information is provided:

- (i) Probe ID (as assigned in the file “List of promoter sequences.xls”)
- (ii) Gene Name
- (iii) Gene Symbol
- (iv) RefSeq accession number.

Sheet 8: Data – Fold change

All the data for the filtered spots on the array are shown. The file contains:

- (i) Probe ID (as assigned in the file “List of promoter sequences.xls”)
- (ii) ER log₂ fold(E/U): The log₂ transformed ratio of the ER occupancy after E2 treatment over the ER occupancy before E2 treatment. The values that pass the fold threshold (log₂ fold ≥ 0.37 or log₂ fold ≤ -0.37) are shown in red
- (iii) POL log₂ fold(E/U): The log₂ transformed ratio of the Pol II occupancy after E2 treatment over the Pol II occupancy before E2 treatment. The values that pass the fold threshold (log₂ fold ≥ 0.37 or log₂ fold ≤ -0.37) are shown in dark blue
- (iv) SRC log₂ fold(E/U): The log₂ transformed ratio of the SRC occupancy after E2 treatment over the SRC occupancy before E2 treatment. The values that pass the fold threshold (log₂ fold ≥ 0.37 or log₂ fold ≤ -0.37) are shown in green
- (v) AcH log₂ fold(E/U): The log₂ transformed ratio of the acetylated histone levels after E2 treatment over the acetylated histone levels before E2 treatment. The values that pass the fold threshold (log₂ fold ≥ 0.37 or log₂ fold ≤ -0.37) are shown in purple
- (vi) ER p-val, POL p-val, SRC p-val and AcH p-val: p-values derived by ANOVA tests comparing the fold changes across all the experimental replicates (see “Outline of statistical analysis” for more information). P-values ≤ 0.05 are shown in blue. A fold change is considered significant if: (a) log₂ fold(E/U) ≥ 0.37 (up-regulated) or log₂ fold(E/U) ≤ -0.37 (down-regulated), and (b) p-value ≤ 0.05 .

Sheet 9: Data – Occupancy

All the data for the filtered spots on the array are shown. The file contains:

- (i) Probe ID (as assigned in the file “List of promoter sequences.xls”)
- (ii) UER: The log₂ transformed ratio of the ER ChIP signal over the input signal before E2 treatment
- (iii) UER%: The percentile of UER among all the spots shown
- (iv) EER: The log₂ transformed ratio of the ER ChIP signal over the input signal after E2 treatment
- (v) EER%: The percentile of EER among all the spots shown
- (vi) UPOL, UPOL%, EPOL, EPOL%, USRC, USRC%, ESRC, ESRC%, UAcH, UAcH%, EAcH and EAcH%: Same nomenclature as the one described above.

7. Description of the Expression Microarray Data Set

The t-test was applied to the normalized data matrix to identify differential genes between the E2-treated and untreated control samples. A significance level cutoff of p-value ≤ 0.05 was applied to select each differential gene set and gene regulation was considered significant if the observed change was greater than 2-fold.

This excel file contains three worksheets.

Sheet 1: p < 0.05

Contains all Affymetrix probes for which p-value ≤ 0.05 . Log2 fold (E/U): The log2 transformed ratio of the transcript level after E2 treatment over the transcript level before E2 treatment. Fold (E/U): The ratio of the transcript level after E2 treatment over the transcript level before E2 treatment. Fold (E/U) p-value: P-value derived by a t-test applied to the normalized data matrix to identify differential genes between the E2-treated and untreated control samples. Fold (E/U) t-test: T-test value derived by the t-test described above

Sheet 2: E2 fold > 2

Contains all Affymetrix probes with p-value ≤ 0.05 whose expression is up-regulated more than 2 fold upon treatment with E2.

Sheet 3: E2 fold < 2

Contains all Affymetrix probes with p-value ≤ 0.05 whose expression is down-regulated more than 2 fold upon treatment with E2.

8. Definitions and Data Used to Derive the Conclusions Presented

1) Direct E2 target genes:

Genes that show E2-dependent changes in Pol promoter occupancy and gene expression.

These include:

- all class II promoters (IIA: 15, IIB: 25),
- all class IV promoters (IVA: 4, IVB: 12), and
- two of the IIIA promoters (CYP1B1 and EBAG9, the other two do not show E2-regulated gene expression).

TOTAL: $(15+25) + (4+12) + 2 = 40 + 16 + 2 = 58$.

Conclusion: We identify 58 E2 direct target genes (mentioned in the first paragraph of the discussion).

Note: The E2 direct target genes do not include the IA and two of the IIIA promoters.

2) Promoters that recruit ER α upon E2 treatment:

All class A promoters.

These include:

- all class IA promoters (IA: 14)
- all class IIA promoters (IIA: 15)
- all class IIIA promoters (IIIA: 4)
- all class IVA promoters (IVA: 4)

TOTAL: $14 + 15 + 4 + 4 = 37$.

Conclusion: We identify 37 promoters that recruit ER α upon E2 treatment (mentioned in the results, Fig. 2A associated text).

3) Direct E2 target genes that recruit ER α at the promoter upon E2 treatment:

Promoters that belong to the E2 direct target genes, as defined in (1), and recruit ER α upon E2, as defined in (2).

These include:

- all class IIA promoters (IIA: 15),
- all class IVA promoters (IVA: 4), and
- two of the IIIA promoters (CYP1B1 and EBAG9).

TOTAL: $15 + 4 + 2 = 21$.

Conclusion: 21 out of the 58 E2 direct target genes (36%) recruit ER α at their promoter upon E2 (mentioned in the discussion).

4) Promoters that recruit ER α upon E2 treatment and contain an ERE-like sequence:

All ER α -recruiting promoters, as defined in (2), that contain an ERE-like sequence.

These include:

- seven class IA promoters (IA-ERE: 7)
 - seven class IIA promoters (IIA-ERE: 7)
 - three class IIIA promoters (IIIA-ERE: 3, both CYP1B1 and EBAG9 belong here)
 - one class IVA promoter (IVA-ERE: 1)
- TOTAL: $7 + 7 + 3 + 1 = 18$.

Conclusion: 18 out of the 37 ER α -recruiting promoters (49%) contain an ERE-like sequence (mentioned in the results, Fig. 3B associated text).

5) Direct E2 target genes that recruit ER α and contain an ERE-like sequence:

Promoters that belong to the E2 direct target genes, as defined in (1), recruit ER α and contain an ERE-like sequence, as defined in (4).

These include:

- all class IIA-ERE promoters (IIA-ERE: 7)
 - all class IVA-ERE promoters (IVA-ERE: 1)
 - two of the class IIIA-ERE promoters (IIIA-ERE: 2, CYP1B1 and EBAG9)
- TOTAL: $7 + 1 + 2 = 10$.

Conclusion #1: 10 out of the 58 E2 direct target genes (17%) recruit ER α and contain an ERE-like sequence (mentioned in the results, Fig. 3B associated text).

Conclusion #2: 10 out of the 21 ER α -recruiting E2 direct target genes (48%) contain an ERE-like sequence (mentioned in the discussion and agrees with the conclusion in (4) for all ER α -recruiting promoters).