Supplementary Information

Contents:

I. Supplementary methods

siRNA oligonucleotide sequences used in BrdU proliferation assays

Gene specific siRNA oligonucleotides were siGENOME SMARTpool reagents (Dharmacon, Lafayette, CO). The catalog numbers for siRNAs and controls used in this study are as follows: ERα (M-003401-02), CCND1 (M-003210-02), RAN (M-010353-00), ACCN1 (M-006105-00), HMGB1 (M-018981-00), FOXA1 (M-010319-00), PCNA (M-003289-02), GATA1 (M-009610-00), GATA3 (M-003781-01), SVIL (M-011398-00), PLK1 (M-003290-01), CCNE1 (M-003213-02), BUB1 (M-004102-00), NCOA4 (M-010321-00), RPS6KB1 (M-003616-02), TFDP1 (M-003327-03), CAV1 (M-003467-01), SAFB (M-005150-00), UBE2I (M-004910-00), NCOA3 (M-003759-02), MYC (M-003282-04), ASF1A (M-020222-00), CDK4 (M-003238-02), CHAF1B (M-019937-00), CDC20 (M-003225-03), H2A.Z (M-011683-01), CDK9 (M-003243-02), MeCP2 (M-013094-01), BAF57 (M-017522-00), CDC2 (M-003224-03), MBD3 (M-013616-01), STK6 (M-003545-09), BUB1B (M-004101-00), PLK4 (M-005036-01), HAT1 (M-011490-01), CoCoA (M-007038-00), ASF1B (M-020553-00), SMAD4 (M-003902-01), RAP80 (M-006995-01), PRMT1 (M-010102-00), BLOS1 (M-012580-00), and RNAi control (si*GLO* RISC-Free siRNA, D-001600-01).

Microarray gene expression profiling experiments

Gene expression profiling was performed as previously described (Rifkin *et al*, 2003). Briefly, total RNA was isolated using Trizol (Invitrogen) per the manufacturer's instructions, Poly-A mRNA was purified from total RNA and cDNAs were prepared from 2 mg Poly-A RNA and labeled using the Powerscript fluorescent labeling kit (BD Biosciences) and monofunctional Cy3 and Cy5 dyes (Amersham /Pharmacia) per manufacturer's protocol. The OHU21K human oligonucleotide array (Yale's Keck Facility), was hybridized at 64°C in a hybridization buffer comprised of 2.83 x SSC, Polyadenylic acid 0.5 mg/ul (Sigma), and 0.18% SDS. The arrays were washed and scanned in an Axon 4000 array scanner as described (Rifkin *et al*, 2003). Each hormone treatment time was performed at least three times and hybridizations included dye swaps.

The analysis of gene expression profiles

Intensity values for cDNAs were generated with GenePix (v.4.0), background intensities were subtracted, and values were multiplied by a correction factor to normalize for labeling differential between the channels. Data were ratio normalized and NET Signal Noise filtered (Yale Microarray Database), then corrected for intensity-dependent bias with a Lowess function in MAANOVA (R/maanova, Version 0.98.8). Bayesian analysis of gene expression levels (BAGEL) (Townsend and Hartl, 2002) generated expression estimates for all time points by running WinABAGEL v.3.6 at default settings on all microarray datasets, where significant ($p < 0.01$) probes were defined relative to the control samples in at least one time point. Gene expression data from all hybridizations can be found at the GEO database under accession number GSE10618.

ChIP-chip experiments

MCF7 Cells were treated with 10 nM E2 (45 minutes and 2 hours for mapping ERα and MYC binding sites, respectively) at 80% confluence. ~$5x10^6$ cells per ChIP were cross-linked with 1% formaldehyde for 10 minutes at 37°C then quenched with 125 mM glycine. The cells were washed with cold PBS and scraped into PBS with protease inhibitors (Roche). Cell pellets were resuspended in ChIP lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl [pH 8.1] and sonicated (Fisher Sonic Dysmembrinator). The sheared chromatin was submitted to a clarification spin and the supernatant then used for ChIP or reserved as "Input". Antibodies used were anti-ERα (Ab-1, Ab-3, and AB-10, Lab Vision), anti-ERα (MC-20), anti-MYC (N-262), normal rabbit IgG (sc-2027), (all Santa Cruz) and mouse IgG (#12-371)(Upstate). For ChIP-PCR assays, forward and reverse primer sequences in the H2A.Z promoter were 5'-GCTACATACCGAGGAGACTTCA-3' and 5'-AGGGAAGAAACAGAGCGAGCTA-3'. For ChIP-on-chip, both ChIP DNA and Input DNA were subjected to the linker-mediated amplification and ChIP and Input DNA samples were further fragmented with DNase I then end-labeled with biotin. The resulting samples were hybridized to Affymetrix GeneChip® Human Tiling 2.0R Arrays per the Affymetrix® ChIP protocol. Independent biological triplicates were performed for each transcription factor and for the control (Input).

Analysis of ChIP-chip tiling array data

Tiling array data were normalized and analyzed as described (Cawley *et al*, 2004). Briefly, raw data were normalized within ChIP and control groups and for each genomic position a local dataset composed of intensities for all adjacent probes within a window of ±250 bp was generated. A one-tailed Wilcoxon Rank Sum test was then applied to compare control and ChIP experiments and was performed in a sliding window across all tiled genomic regions. Significantly enriched probes ($p < 1e-5$) were locally extended by merging adjacent enriched probes within 100 bp and these merged regions were defined as transcription factor bound regions. Data from all ChIP-chip hybridizations can be found at the GEO database under accession number GSE10800 .

Visualization and Analysis of Protein-Protein Interaction Network of Estrogen Signaling Targets

Pairwise protein-protein interaction data were derived from Human Protein Reference Database (HPRD) (Peri *et al*, 2003) and the BioGRID database (Stark *et al*, 2006). Only proteins that are E2-regulated, or harbor ERα binding sites within 50 Kb of their TSSs, or harbor MYC binding sites within 20 Kb of their TSSs were selected for visualization in the interaction network using Cytoscape (Shannon *et al*, 2003) (See Figure 5 and data available in Supplementary Table 5). Node size correlates with the degree of the node connectivity. Modules, or highly connected subnetworks, were further identified using MCODE (Bader and Hogue, 2003). Representative modules included the one involved in nucleic acid and protein metabolism (Figure 5).

Tissue Array Design, Immunohistochemistry, and Data Analysis

Specimens and clinical information were collected under the guidelines and approval of a Yale University Institutional Review Board. Estrogen receptor staining was positive in 52%, progesterone receptor in 46%, and HER2/neu in 14% of tumors represented on the array. Nuclear grade 3 (on a 1-3 scale) was noted in 28% of the specimens, and 59% of tumors were larger than 2 cm. The histologic subtypes included 72% invasive ductal carcinoma, 1% lobular carcinoma, and the remaining had mixed or other histology. The specimens were resected between 1962 and 1980, with a clinical follow-up ranging from 4 months to over 40 years, and a mean follow-up time of 12.6 years. Age at diagnosis ranged from 24 to 88 years (mean age, 58 years). Complete treatment history was not available for the entire cohort. Most patients were treated with local irradiation. None of the node-negative patients were given adjuvant systemic therapy. A minority of the node-positive patients (15%) received chemotherapy, and 27% received tamoxifen (after 1978). The time between tumor resection and tissue fixation was not available. A pathologist reviewed slides from all of the blocks to select representative areas of invasive tumor to be cored. The cores were placed on the tissue microarray using a Tissue Microarrayer (Beecher Instruments, Silver Spring, MD). The tissue microarrays were then cut to 0.5 mm sections and placed on glass slides using an adhesive tape-transfer system (Instrumedics, Inc., Hackensack, NJ) with UV cross-linking.

For immunohistochemistry, tissue array slides were deparaffinized by rinsing with xylene, followed by two washes with of 100% EtOH and two washes with water. The slides were then boiled in sodium citrate buffer (pH = 6.0) for antigen retrieval. To block endogenous peroxidase activity, the slides were incubated with 2.5% hydrogen peroxide in methanol for 30 minutes at RT. The slides were then washed with Tris-buffered saline (TBS), incubated in 0.3% BSA/1xTBS for 30 minutes at RT to reduce nonspecific background, and then stained with antibody against H2A.Z (Upstate, #07-594) at 1:5,000 dilution in BSA/TBS at 4°C overnight. The slides were rinsed three times in 1xTBS/0.05% Tween-20. Bound antibody was detected by applying anti-rabbit horseradish peroxidase-labeled polymer secondary antibody from the DAKO EnVision kit. The slides were washed with 1xTBS/0.05% Tween-20, and incubated for 10 min in 3,3'-diaminobenzidine in buffered substrate (Dako). Counterstaining was performed with hematoxylin, and slides were mounted with Immunomount (Shandon).

Scoring of Tissue Array Immunostaining: Intensity of H2A.Z nuclear staining was scored with a four-tiered system (0 to 3), with 0 representing negative staining, 1 representing weak staining, 2 representing intermediate staining, and 3 representing strong staining. Each tissue core was also semi-quantitatively scored by estimating the percentage of positive tumor cells in 5 categories as: 0%, < 10%, 10%-50%, 50%-80%, and > 80% (on a 0-4 scale). Combined scores were calculated by adding intensity score to percentage score, and then translated into a final score interpreted as negative (combined score: 0-2), weakly positive (score: 3), moderately positive (score: 4-5) and highly positive (score: 6-7) H2A.Z staining. For comparison of inter-observer variations, we performed correlation analysis and obtained concordance in the vast majority of cases with Pearson correlation $r = 0.8$, indicating reasonable accuracy and reproducibility of used scoring system.

Automated scoring was performed using the Automated Cellular Imaging System (ACIS) from

Clarient. ACIS software was programmed by M. Tretiakova to calculate the total number of pixels contained within all of the nuclei that are either brown (H2A.Z positive) or blue (negative), and calculate the ratio (%) of positively stained cells to all cells. The measurement of intensity of the nuclear staining was based on three related color parameters: the color defined by hue, the "darkness" defined as luminosity, and density of the color defined as the saturation. For each TMA core we obtained two major parameters: staining intensity and % positivity, however the later parameter was applied to all nuclei without their discrimination to tumor/non-tumor compartments. H2A.Z intensity measurements were then translated into the 4-tier system as negative, weak, moderate and strongly positive staining. These readings were then adjusted to exclude from final intensity score staining artifacts and cutting artifacts (from tape-transfer system). To reach consensus in scoring we compared manual semi-quantitative results (A.K, T.K) with automated scoring (M.T) in discrepant cases, and reviewed them at high magnification. H&E slide was used to analyze tissue histology in each core by T.K and M.T.

H2A.Z staining on TMA was further verified using our newly developed technology of automated quantitative analysis (AQUA) of TMAs on the second cohort of breast tumor TMA as described above. Details for automated image acquisition and algorithmic image analysis were previously described (Camp *et al*, 2002).

The JMP6 package (SAS Insitute Inc.) was used for the statistical analysis. Associations between H2A.Z protein level and other clinical and pathologic variables were examined using Chi-Square contingency test. Both univariate and multivariate analyses based on Cox proportional-hazards regression models, with overall survival as an end point, were preformed to test the prognostic significance of the variables. Kaplan-Meier survival curves were plotted (L or Low: negative; M or Moderate: weakly positive or moderately positive; H or High: highly positive in Fig. 5) and the Log-Rank test was performed with WinStat software (Kalmia Inc.).

II. References for Supplementary Information

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25:** 25-29.

Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4:** 2.

Camp RL, Chung GG, Rimm DL (2002) Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat Med* **8:** 1323-1327.

Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, Fox EA, Silver PA, Gingeras TR, Liu XS, Brown M (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38:** 1289-1297.

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, Wheeler R, Wong B, Drenkow J, Yamanaka M, Patel S, Brubaker S, Tammana H, Helt G, Struhl K, Gingeras TR (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116:** 499-509.

Cheng AS, Jin VX, Fan M, Smith LT, Liyanarachchi S, Yan PS, Leu YW, Chan MW, Plass C, Nephew KP, Davuluri RV, Huang TH (2006) Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters. *Mol Cell* **21:** 393-404.

Couronne O, Poliakov A, Bray N, Ishkhanov T, Ryaboy D, Rubin E, Pachter L, Dubchak I (2003) Strategies and tools for whole-genome alignments. *Genome Res* **13:** 73-80.

Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* **31:** 19-20.

Hertz GZ, Stormo GD (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15:** 563-577.

Hosack DA, Dennis G, Jr., Sherman BT, Lane HC, Lempicki RA (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol* **4:** R70.

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **34:** D354-357.

Keightley PD, Lercher MJ, Eyre-Walker A (2005) Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol* **3:** e42.

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34:** D108-110.

Peri S *et al* (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* **13:** 2363-2371.

Rifkin SA, Kim J, White KP (2003) Evolution of gene expression in the Drosophila melanogaster subgroup. *Nat Genet* **33:** 138-144.

Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32:** D91-94.

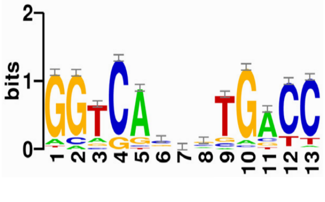
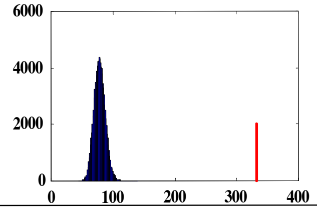
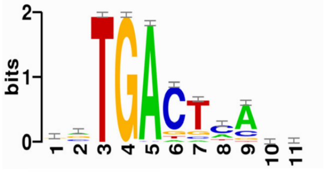
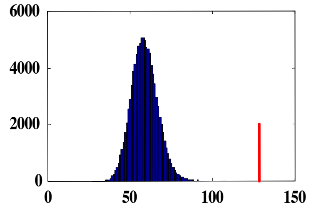
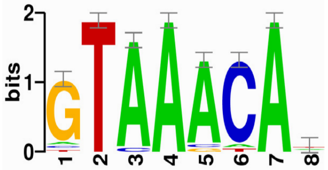
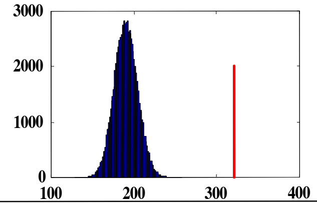
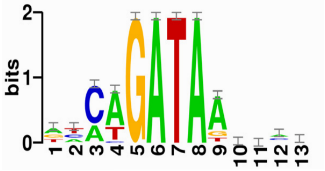
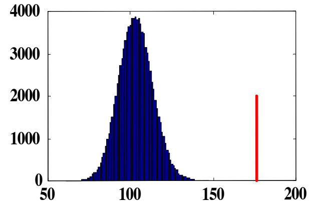
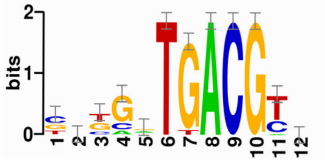
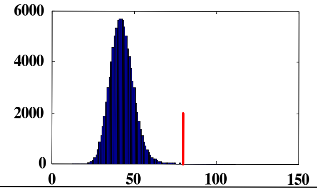
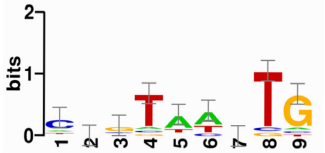
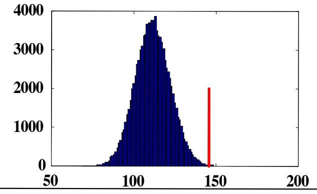
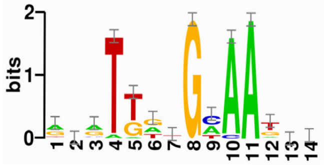
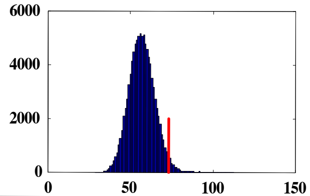
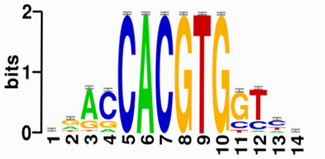
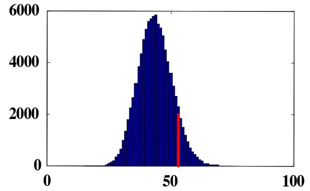
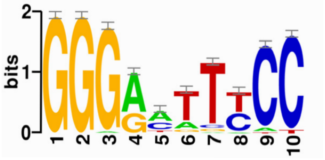
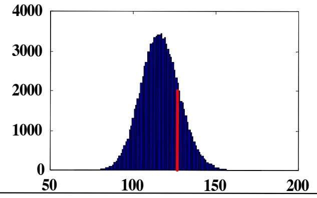
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498-2504.

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34:** D535-539.

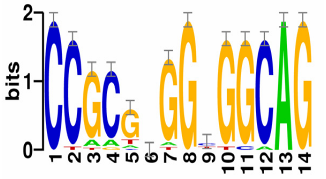
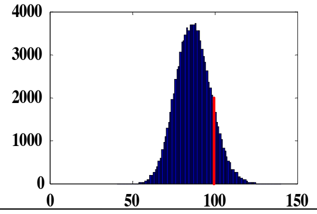
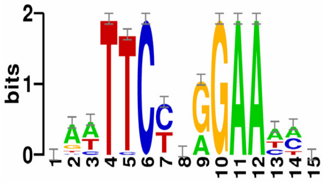
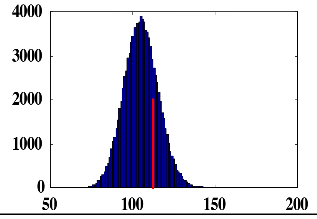
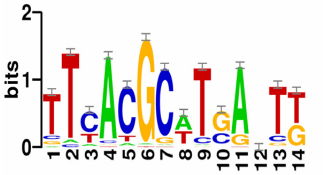
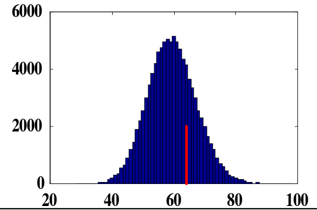
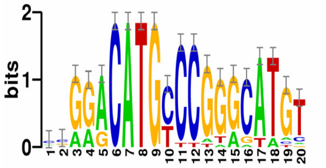
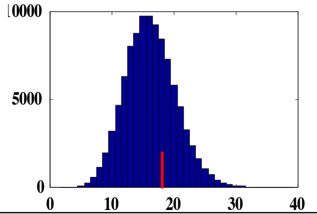
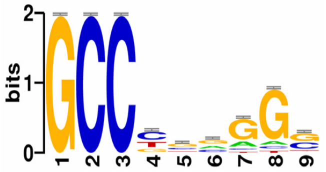
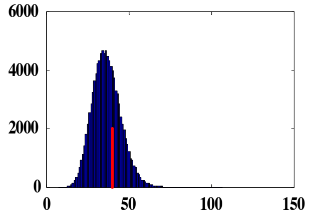
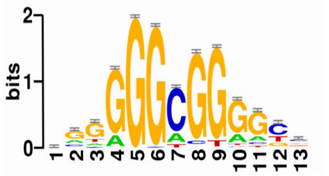
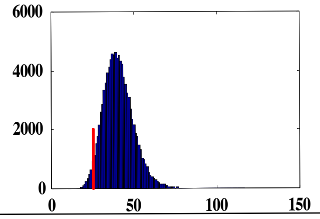
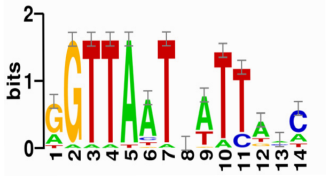
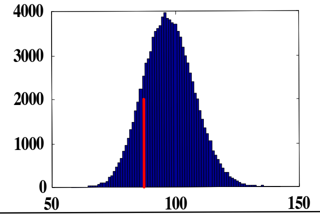
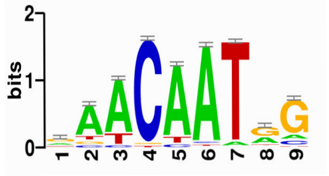
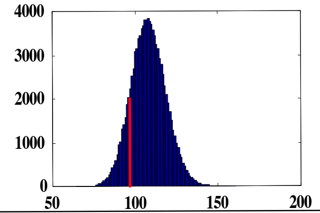
Townsend JP, Hartl DL (2002) Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. *Genome Biol* **3:** RESEARCH0071.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347:** 1999-2009.

Supplementary Figure 1

Transcription Factor	Class	Binding Motif	Random simulation (100,000 runs)	# of ERα ChIP sites containing motif(s)	Average # of random sites containing motif(s)	Fold enrichment	P-value
ERα	Cys4 zinc finger of nuclear receptor type			332	79	4.20	2.2e-163
AP-1	Leucine zipper factors (bZIP)			128	59	2.17	8.0e-18
FoxA1	Fork head/winged helix			321	190	1.69	1.5e-20
GATA	Diverse Cys4 Zinc fingers			176	103	1.71	8.2e-13
CREB1	Basic-leucine zipper (bZIP)			80	43	1.86	2.6e-7
Msx1	Homeo domain			146	111	1.32	6.6e-4
C/EBPβ	Leucine zipper factors (bZIP)			73	57	1.28	2.1e-2
MYC	Helix-loop-helix/leucine zipper (bHLH-ZIP)			53	44	1.20	9.3e-2
NFϰB	Rel homolog region (RHR)			127	116	1.09	1.8e-1

Supplementary Figure S1 (continued)

CTCF	Zinc fingers (multiple)			99	87	1.14	1.3e-1
STAT5α	STAT			112	105	1.07	2.6e-1
Pax6	Paired-box			64	59	1.08	2.8e-1
P53	P53			18	16	1.12	3.3e-1
AP-2γ	AP2/ERE BP-related factors			40	36	1.11	3.4e-1
Sp1	Cys2His2 zinc finger			25	41	0.61	6.2e-1
TCF1	High mobility group (HMG)			87	97	0.90	8.4e-1
Sox9	High mobility group (HMG)			97	108	0.90	8.5e-1

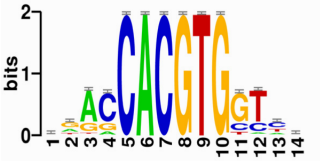
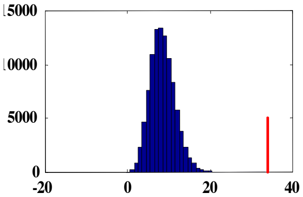
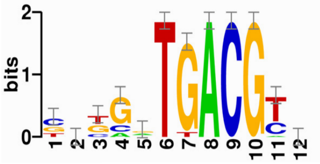
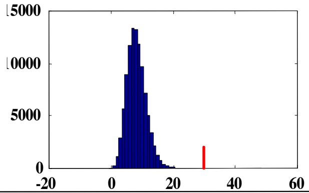
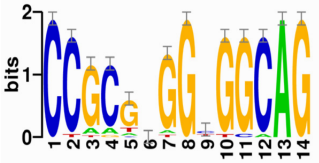
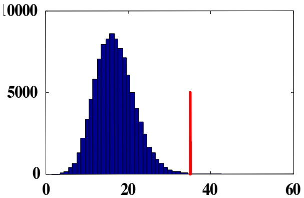
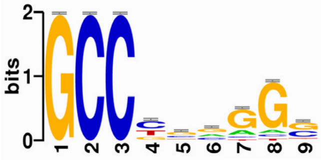
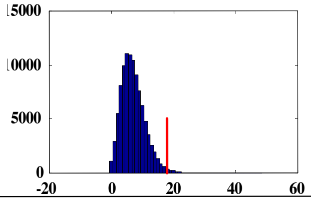
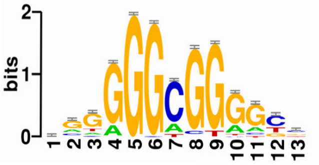
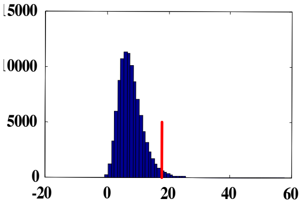
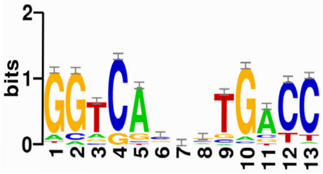
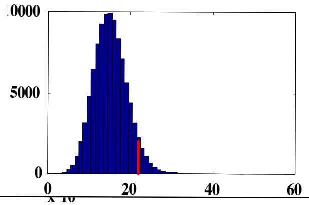
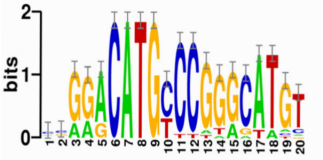
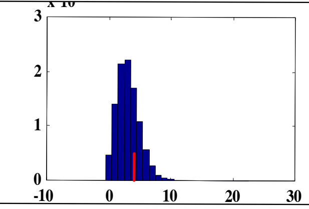
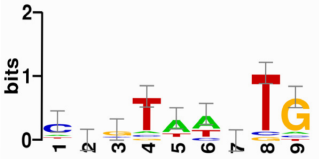
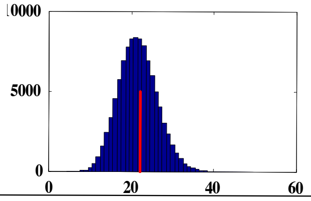
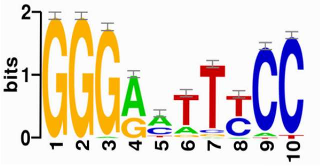
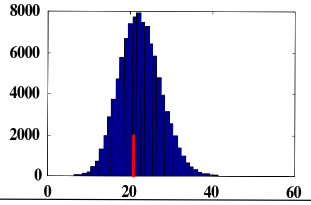
(**Supplementary Figure 1**, continued)

Supplementary Figure 1 Enrichment of Transcription Factor Binding Motifs in ERα-Bound Regions.

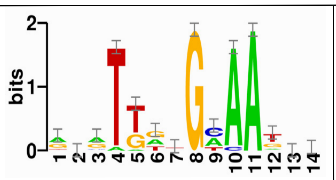
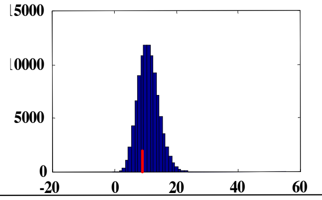
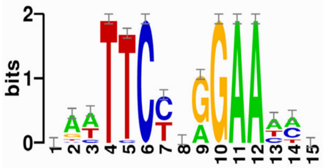
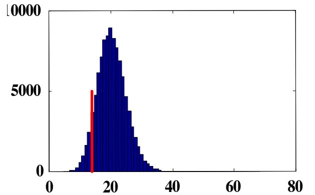
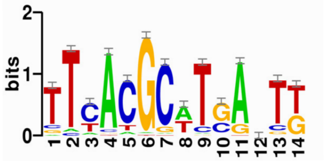
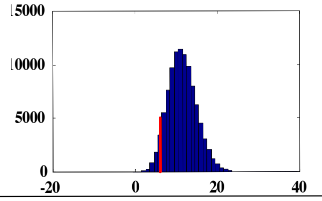
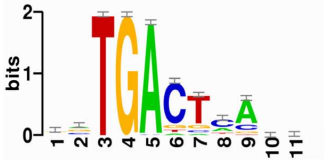
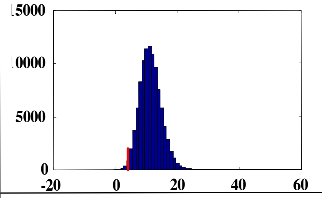
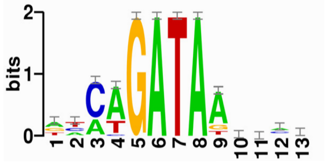
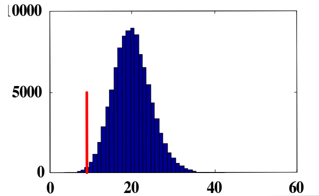
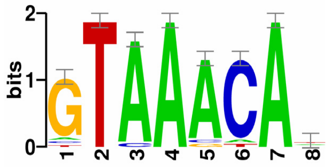
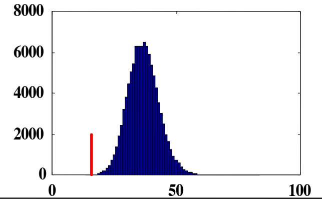
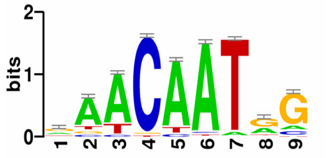
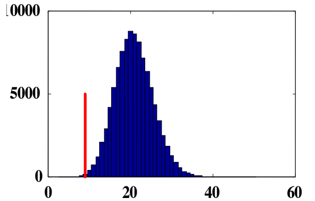
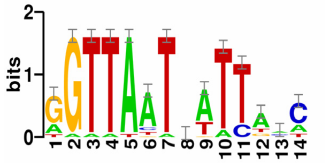
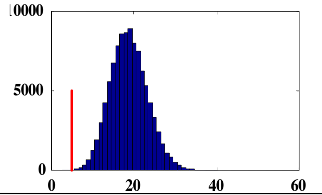
Computational analysis of the high-confidence ERα-binding sites detected by ChIP-chip (see Supplementary Table 1) for enrichment of consensus estrogen response elements (EREs). Similar analysis was performed to identify significant enrichment of putative recognition sites for the transcription factors AP-1, FOXA1, GATA, CREB1, and MSX1 (all of which were significantly enriched at ERα bound regions, $p < 0.01$).

To test for the enrichment of predicted transcription factor binding motifs in ChIP-identified ERα- or MYC-bound regions, we obtained position weight matrices for transcription factor binding motifs from the TRANSFAC database (release 6.1)(Matys *et al*, 2006), the JASPAR database (Sandelin *et al*, 2004), and from manually collected literature on ERα binding. These transcription factors principally include ones related to estrogen signaling, normal mammary development and breast carcinogenesis. PATSER (Hertz and Stormo, 1999) was applied to scan repeat masked human genomic sequence for matches to the weight matrices. Fold enrichment and significance were estimated by comparing the number of putative binding motifs within ERα- or MYC-bound regions (unified length = 500 bp) with the number of predicted motifs within the same number of randomly selected 500 bp genomic regions (100,000 randomized sampling runs were performed).

Supplementary Figure 2

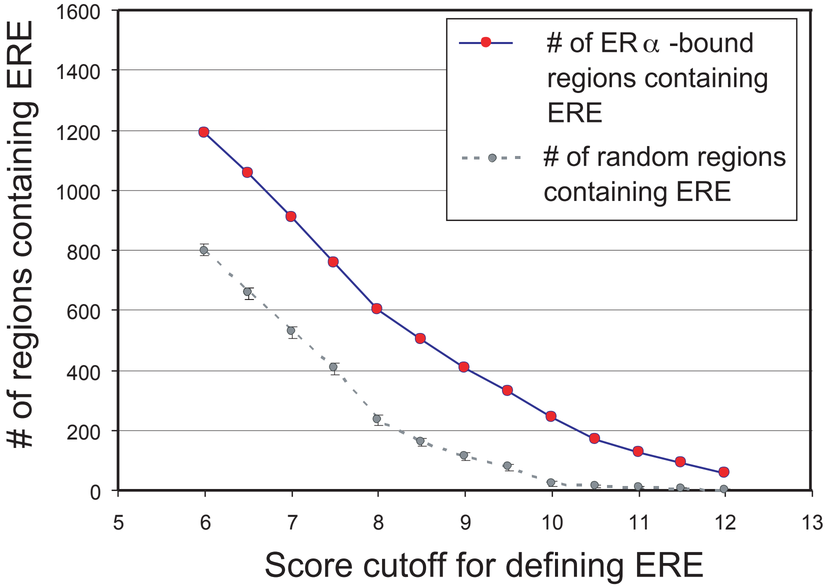
Transcription Factor	Class	Binding Motif	Random simulation (100,000 runs)	# of Myc ChIP sites containing motif(s)	# of random sites containing motif(s)	Fold enrichment	P-value
MYC	Helix-loop-helix/leucine zipper (bHLH-ZIP)			34	8	4.25	2.1e-17
CREB1	Basic-leucine zipper (bZIP)			30	8	3.75	1.8e-11
CTCF	Zinc fingers (multiple)			35	16	2.19	5.6e-5
AP-2γ	AP2/EREBP-related factors			18	7	2.57	2.4e-3
Sp1	Cys2His2 zinc finger			18	8	2.25	5.8e-3
ERα	Cys4 zinc finger of nuclear receptor type			22	15	1.47	4.8e-2
P53	P53			4	3	1.33	3.1e-1
Msx1	Homeo domain			22	22	1.00	4.6e-1
NFϰB	Rel homolog region (RHR)			21	22	0.95	6.2e-1

Supplementary Figure 2 (continued)

C/EBPβ	Leucine zipper factors (bZIP)			9	11	0.82	7.2e-1
STAT5α	STAT			14	20	0.70	9.1e-1
Pax6	Paired-box			6	11	0.54	9.4e-1
AP-1	Leucine zipper factors (bZIP)			4	11	0.36	9.8e-1
GATA	Diverse Cys4 Zinc fingers			9	20	0.45	9.9e-1
FoxA1	Fork head/winged helix			16	37	0.43	9.9e-1
Sox9	High mobility group (HMG)			9	21	0.43	9.9e-1
TCF1	High mobility group (HMG)			5	19	0.26	9.9e-1

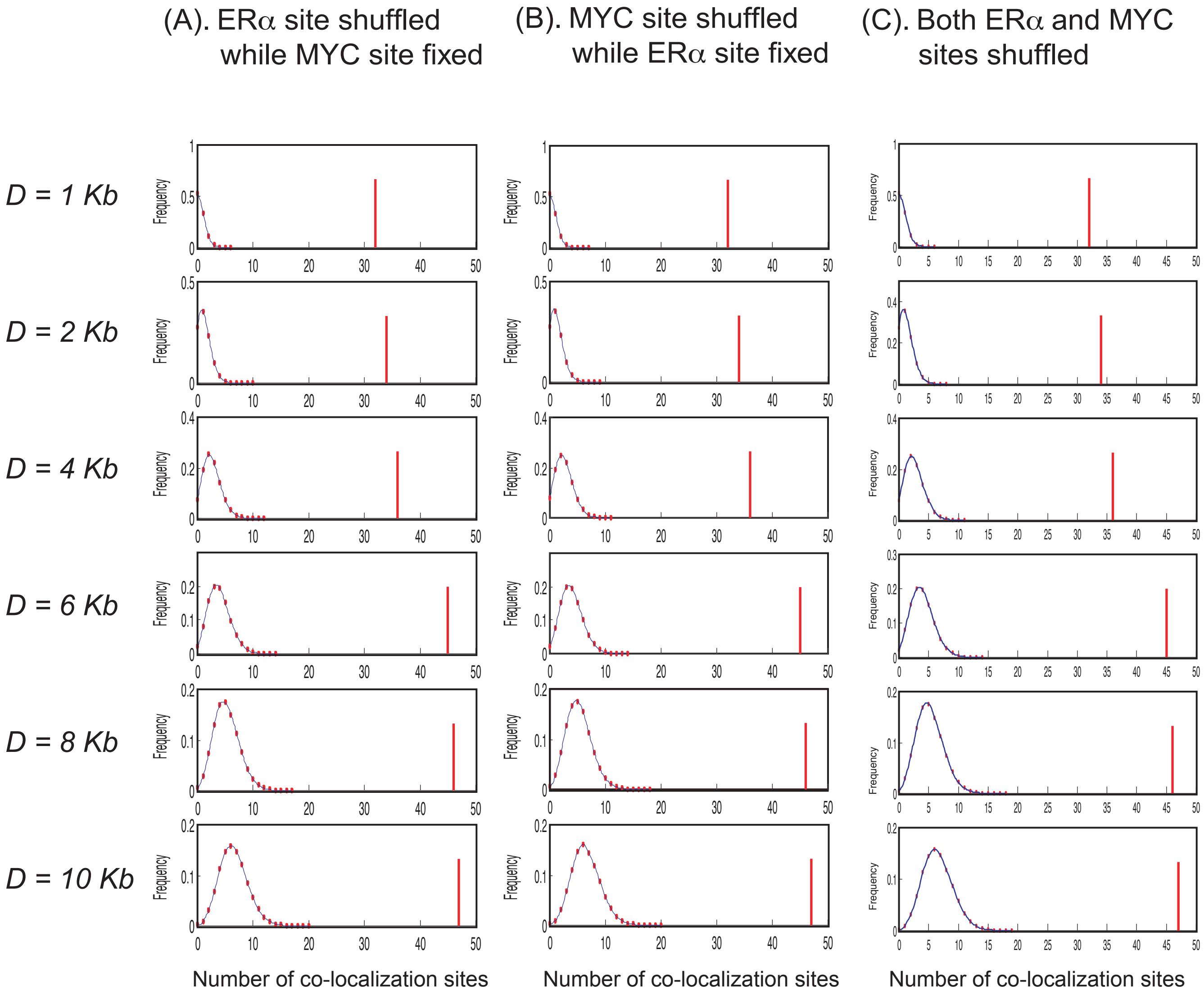
Supplementary Figure 2 Enrichment of Transcription Factor Binding Motifs in MYC-Bound Regions.

Computational analysis of the high-confidence MYC-binding sites detected by ChIP-on-chip (see Supplementary Table 2) for enrichment of the MYC recognition site (a.k.a. E-box). Similar analysis was performed to identify significant enrichment of putative recognition sites for the transcription factors CREB1, CTCF, AP-2g and Sp1 (all of which were significantly enriched at MYC bound regions, p < 0.01).



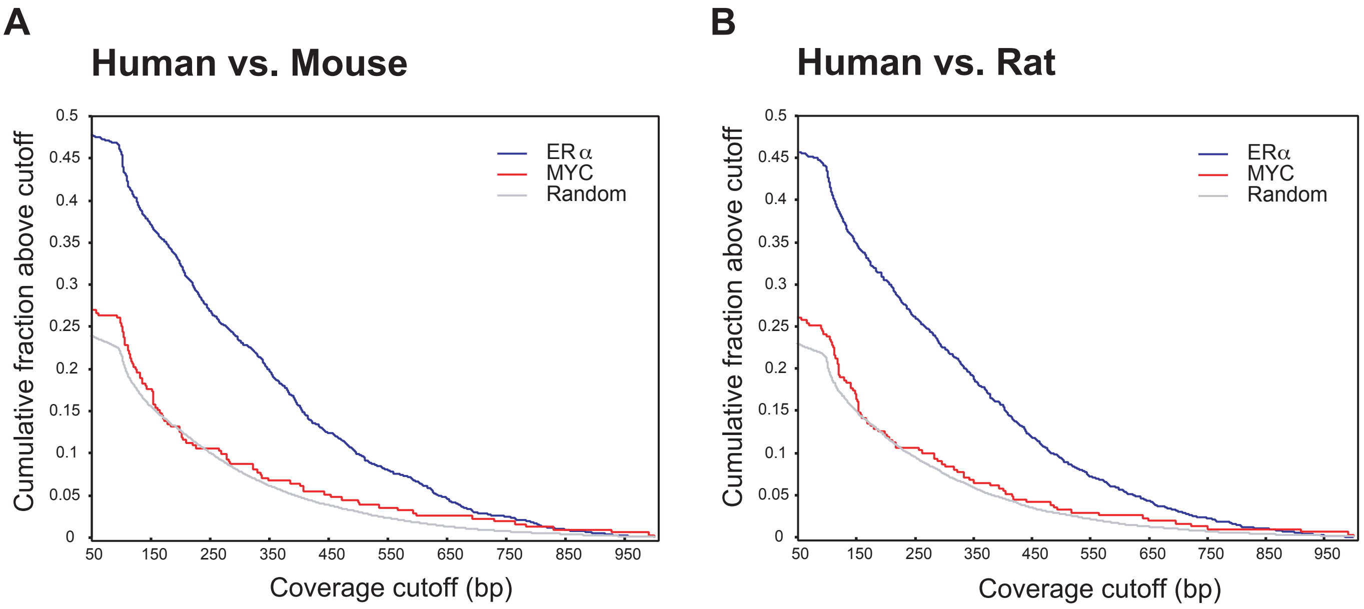
Supplementary Figure 3 Effect of Position Weight Matrix Stringency on ERE Sequence Detection from ERα-Bound Loci.

To test for the enrichment of predicted ERE binding motifs in ChIP-identified ERα-bound regions, we obtained position weight matrices for ERE motifs from the TRANSFAC database (release 6.1)(Matys et al, 2006) and the JASPAR database (Sandelin et al, 2004). PATSER (Hertz et al, 1999) was applied to scan repeat masked human genomic sequence for matches to the weight matrices according to the threshold score cutoffs noted (Y axis). For all matrix thresholds tested there was enrichment of putative binding motifs within ERα-bound regions (unified length = 500 bp) compared with the number of predicted motifs within the same number of randomly selected 500 bp genomic regions (100,000 randomized sampling runs were performed).



Supplementary Figure 4 Co-occurrence of ERα- and MYC-Bound Regions Detected by ChIP-chip in MCF7 Cells.

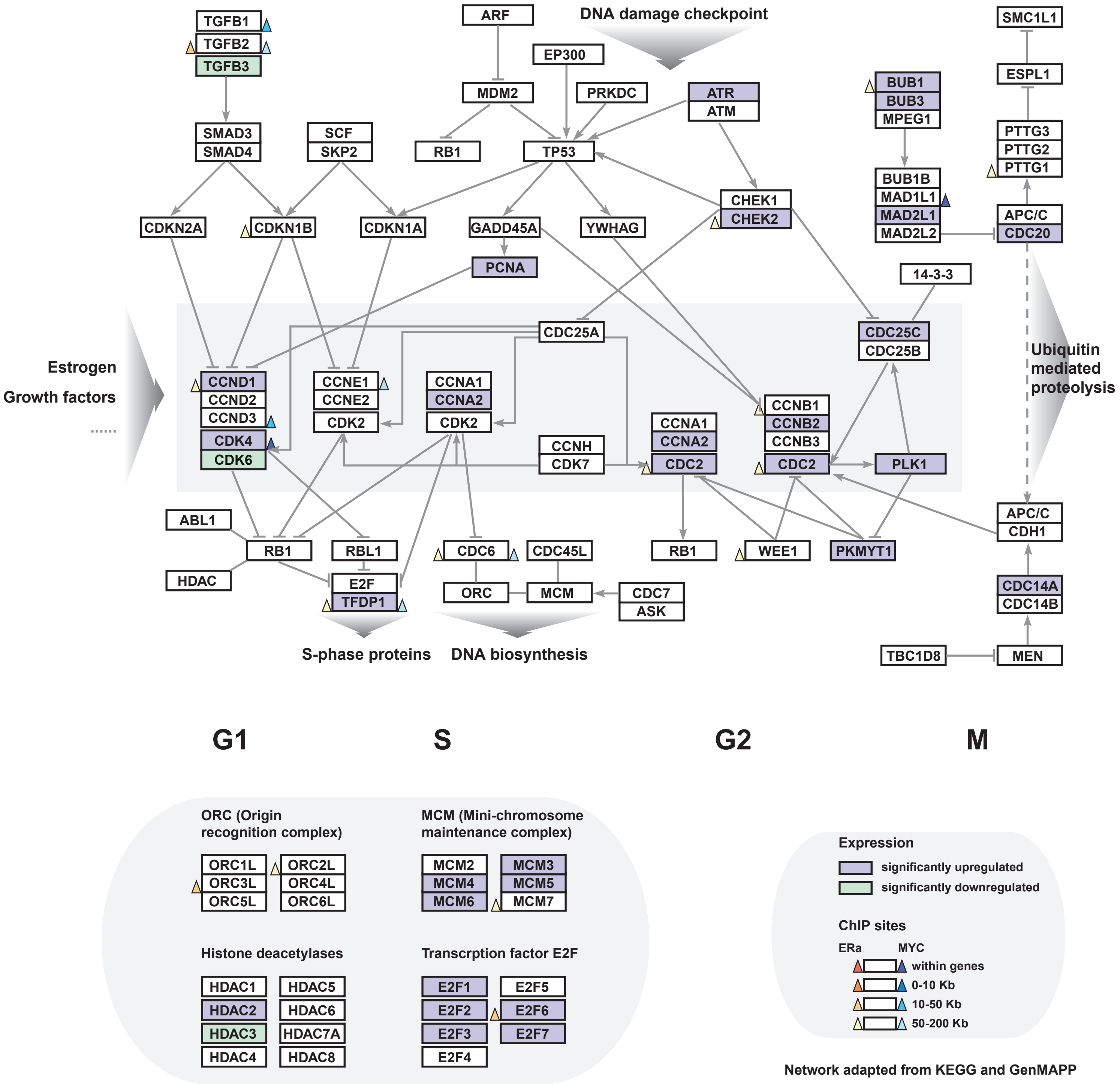
This figure demonstrates statistically significant co-occurrence of ERα- and MYC-bound regions detected by ChIP-chip. We tested the statistical significance of this co-occurrence with three different random distribution models: Using a model that fixes experimentally-derived MYC-bound loci and randomly shuffles ERα-bound sites (A), one that fixes experimentally-derived ERα-bound loci and randomly shuffles MYC-bound sites (B), and one that randomly shuffles both ERα- and MYC-bound regions (C). We obtained similar statistical significance when applying each of the three models, above ($p < 1e-16$ for colocalization of MYC sites within 1Kb of an ERα binding site using model C, above). The red line in each panel indicates the number of MYC-bound regions co-localized within a given genomic distance (D) of an ERα-bound region. Positions of MYC and/or ERα binding sites were randomly shuffled and 100,000 random simulation runs were performed to estimate the distribution of numbers of MYC sites with adjacent ERα binding sites within a given genomic distance (the blue curve in each panel). These data support a model of cooperative functions between ERα and MYC at many genomic targets in breast cancer cells(Cheng et al, 2006).



Supplementary Figure 5 Evolutionary Conservation of Regions Bound by ERα and MYC.

Conservation of ERα or MYC binding regions (unified length = 1 Kb) was estimated at the nucleotide level by the extent of coverage of conserved genomic regions based upon whole genome alignments between human-mouse (hg16-mm4) and human-rat (hg16-rn3) (criteria of ≥ 100 bp and ≥ 70 percent identity) generated by the Dubchak Lab at UC Berkeley (Couronne et al, 2003). Cumulative distributions of coverage values (i.e., the fraction of transcription factor bound sites above the cutoff criteria) for ERα-bound regions (blue lines) and MYC-bound regions (red lines) compared to random genomic regions (length = 1 Kb) (grey lines) based on the human-mouse genomic alignment and the human-rat alignment are shown in (A) and (B), respectively.

ERα-binding regions sustained relatively high sequence conservation between human and mouse or human and rat (about 75 million year divergence times) as compared to alignments of genomic background, while MYC-binding regions did not show sequence conservation above background levels. Interestingly, in primate genomes, regions closer to gene start sites, in which many MYC sites predominate, have been found to be undergoing rapid evolution (Keightley et al, 2005). Our results indicate that, for MYC-bound regions, mutation accumulation has reached saturating levels between humans and rodents while evolutionary comparisons are still informative for ERα-bound regions.



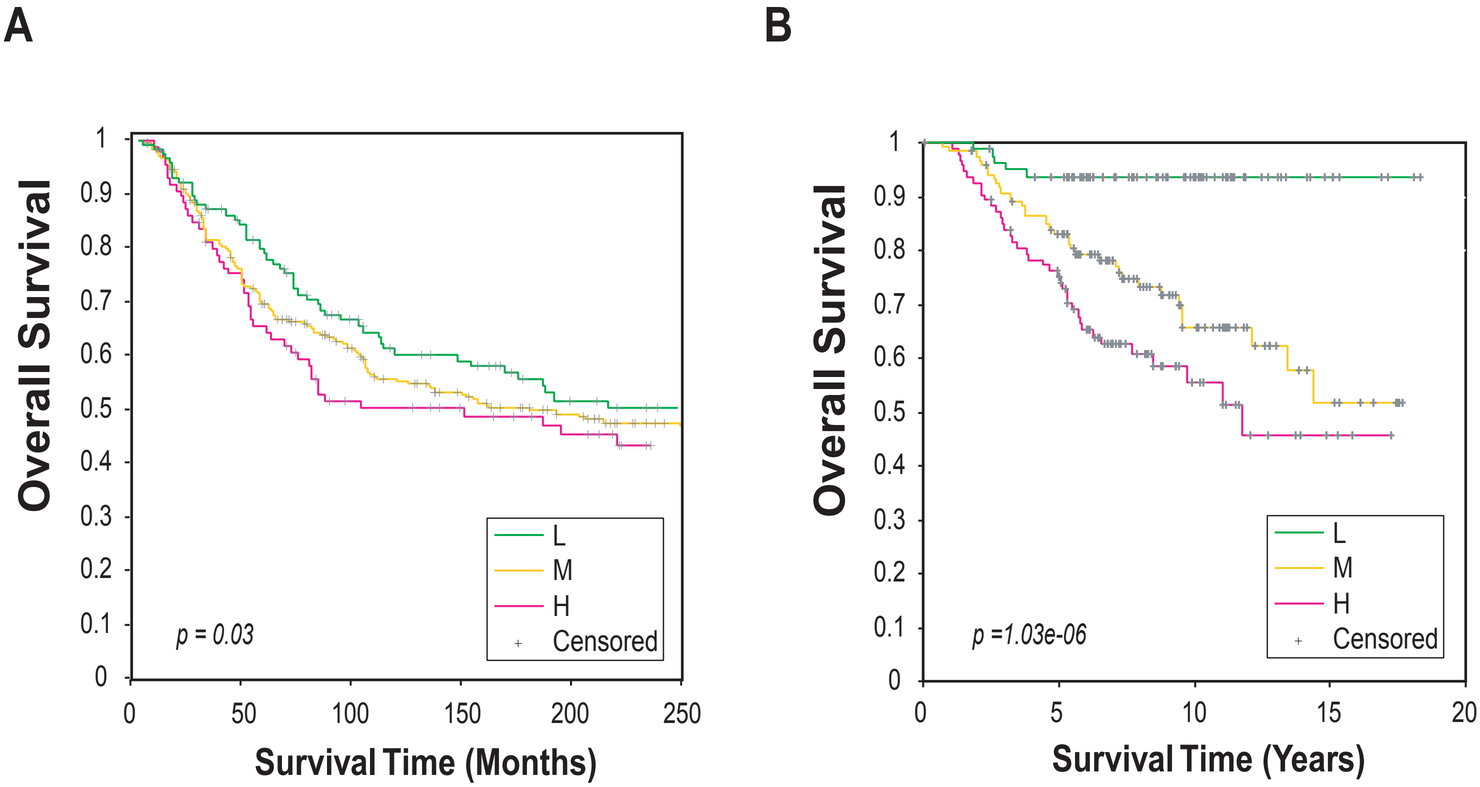
Network adapted from KEGG and GenMAPP

Supplementary Figure 6 Schematic of Mediators and Regulators of the Cell Cycle with Estrogen-Regulated Genes Highlighted and ERα and/or MYC Binding Sites Indicated.

Estrogen acts as a potent mitogen in MCF7 cells and affects many downstream targets which participate in cell cycle regulation. Estrogen-induced and estrogen-repressed genes (all $p < 0.01$) are highlighted with purple and green colors, respectively. The genes containing ERα or MYC binding sites are labeled with color-coded triangles which also indicate binding-site location relative to the target genes (see color legend). The network structure is adapted from KEGG (http://www.genome.jp/kegg/pathway/hsa/hsa04110.html) and GenMAPP (http://www.genmapp.org/).

(**Supplementary Figure 6**, continued)

Expression Analysis Systematic Explorer (EASE v 2.0) (Hosack *et al*, 2003) was used to analyze over-represented functional categories or pathways within the E2 signaling target list identified by ChIP-Chip and expression profiling analyses. The function or pathway annotation systems included Gene Ontology (GO) (Ashburner *et al*, 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al*, 2006), GenMAPP (Dahlquist *et al*, 2002) and BioCarta (http://www.biocarta.com/genes/). One-tailed Fisher exact probability and EASE scores, a conservative adjustment to the Fisher exact probability test, were calculated to identify significantly over-represented categories. Enriched functional categories or pathways were defined as those with a P-value < 0.05 and with the number of list hits of at least 3 from each category or pathway (see Supplementary Table 5).



Supplementary Figure 7 Kaplan-Meier Survival Curves for H2A.Z Protein Staining on a Second Cohort of Breast Tumor Samples and for H2A.Z mRNA Levels from an Independent Breast Tumor Set.

A newly established tissue microarray technology (automated quantitative analysis, or AQUA) was applied to detect H2A.Z protein levels on a second cohort of breast tumor samples (n=459). Two cut-points of continuous AQUA scores at 27.6 and 62.2 were used to divide all tumor samples into 3 intensity categories: Low/L, 113 samples; Moderate/M, 262 samples; High/H, 84 samples. Consistent with the primary tumor cohort (Figure 8), high H2A.Z protein levels were again associated with decreased breast cancer survival as demonstrated by Kaplan-Meier survival analysis and the Log-Rank test (p = 0.03) (A). An independent dataset included H2A.Z mRNA levels from 295 breast tumor samples (van de Vijver et al, 2002) was used to perform similar analysis (B). All samples were divided into 3 groups based on two cut-points of normalized H2A.Z mRNA expression ratio (log-transformed) (–0.40 and 0.27): Low/L, 80 samples; Moderate/M 122 samples; High/H, 93 samples. Statistically significant association between high H2A.Z mRNA level and short patient survival was observed (p = 1.03e-6) (B).