nature genetics

Genome-wide analysis of estrogen receptor binding sites

Jason S Carroll[1], Clifford A Meyer[2,3], Jun Song[2,3], Wei Li[2,3], Timothy R Geistlinger[1], Jérôme Eeckhoute[1], Alexander S Brodsky[4], Erika Krasnickas Keeton[1], Kirsten C Fertuck[1], Giles F Hall[5], Qianben Wang[1], Stefan Bekiranov[6,8], Victor Sementchenko[6], Edward A Fox[5], Pamela A Silver[5,7], Thomas R Gingeras[6], X Shirley Liu[2,3] & Myles Brown[1]

The estrogen receptor is the master transcriptional regulator of breast cancer phenotype and the archetype of a molecular therapeutic target. We mapped all estrogen receptor and RNA polymerase II binding sites on a genome-wide scale, identifying the authentic *cis* binding sites and target genes, in breast cancer cells. Combining this unique resource with gene expression data demonstrates distinct temporal mechanisms of estrogen-mediated gene regulation, particularly in the case of estrogen-suppressed genes. Furthermore, this resource has allowed the identification of *cis*-regulatory sites in previously unexplored regions of the genome and the cooperating transcription factors underlying estrogen signaling in breast cancer.

Recent work has focused on identifying gene expression signatures in breast cancer subtypes that predict response to specific treatment regimes and improved disease outcome[1–4]. Tumors with gene expression profiles that correlate with estrogen receptor α (hereafter referred to simply as 'estrogen receptor') expression have been termed luminal type[1,5] and are associated with response to endocrine therapy and improved survival, although the mechanisms by which estrogen receptor dictates tumor status are poorly understood.

Estrogen receptor–mediated transcription has been intensively studied on a small number of endogenous target promoters[6–8], and recent location analysis of estrogen receptor binding by chromatin immunoprecipitation combined with microarrays (ChIP-on-chip) also focused primarily on promoter regions[9,10]. We recently expanded on these analyses to map estrogen receptor binding sites in a less biased way that did not depend on preexisting concepts of classic promoter domains[11] and subsequently identified several new features of estrogen receptor transcription, including an involvement of distal *cis*-regulatory enhancer regions and a requirement for the Forkhead protein, FoxA1, in facilitating estrogen receptor binding to chromatin and subsequent gene transcription[11]. This analysis highlighted the importance of regions of chromatin distinct from the promoter-proximal regions and suggested an *in vivo* requirement for cooperating transcription factors. However, owing to technological limitations, this investigation was restricted to chromosomes 21 and 22, comprising <3% of the genome and containing few estrogen receptor–regulated genes[11]. Recent chromosome-wide transcript analyses have demonstrated the existence of multiple layers of transcription that are independent of known coding gene regions[12], implying that transcription factor activity cannot be described by a limited set of paradigms that are restricted to well-studied regions of the genome. To overcome these limitations, we conducted a genome-wide analysis of estrogen receptor and RNA polymerase II (PolII) binding by mapping estrogen-induced estrogen receptor and RNA PolII binding sites on all 1,500 Mb of nonrepetitive sequence in a breast cancer cell line at 35-bp resolution. The combination of this unique resource with gene expression data serves to elucidate the mechanisms underlying estrogen-regulated gene expression in breast cancer.

RESULTS

The MCF-7 breast cancer cell line has been extensively used as a model of hormone-dependent breast cancer. We deprived MCF-7 cells of hormones for 3 d and then synchronously induced transcription by the addition of estrogen for a brief period of time (45 min) known to result in maximal estrogen receptor–chromatin binding[6,11]. We used estrogen receptor–specific and RNA PolII–specific antibodies for ChIP and prepared precipitated chromatin as previously described[11]. We hybridized ChIP chromatin and input DNA to the Affymetrix Human tiling 1.0 microarrays representing the entire nonrepetitive human genome sequence (NCBI build 35) tiled at 35-bp resolution. We performed three biological replicates and identified enriched binding sites (**Supplementary Note** online) by the intersection of two independent methods: namely, a nonparametric generalized

[1]Department of Medical Oncology and [2]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney St., Boston, Massachusetts 02115, USA. [3]Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA. [4]Brown University, Laboratories for Molecular Medicine, Center for Genomics and Proteomics, Providence, Rhode Island 02903, USA. [5]Department of Cancer Biology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney St., Boston, Massachusetts 02115, USA. [6]Affymetrix, 3380 Central Expressway, Santa Clara, California 95051, USA. [7]Department of Systems Biology, Harvard Medical School, 200 Longwood Ave., Boston, Massachusetts 02115, USA. [8]Current address: Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA. Correspondence should be addressed to M.B. (myles_brown@dfci.harvard.edu) or X.S.L. (xsliu@jimmy.harvard.edu).

Received 19 July; accepted 8 September; published online 1 October 2006; corrected online 20 October 2006 (details online); doi:10.1038/ng1901

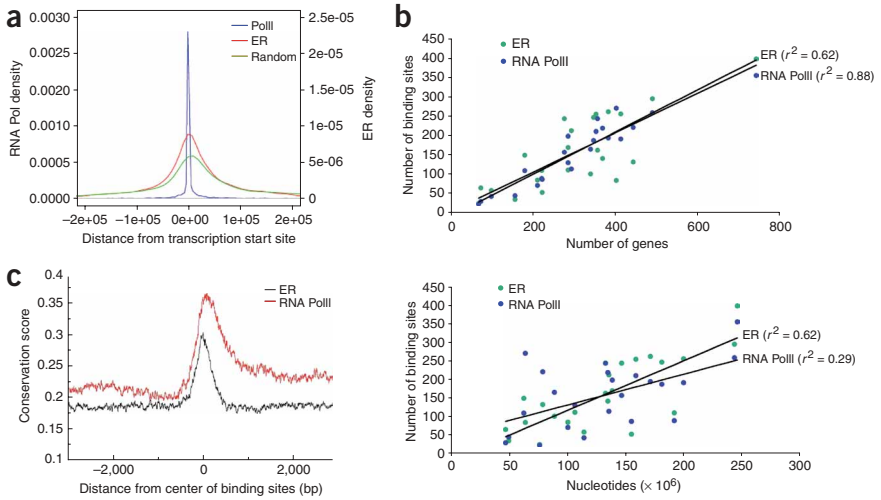

Figure 1 Summary of estrogen receptor and RNA PolII binding sites and correlation with nucleotide and gene number. (**a**) Location of estrogen receptor (ER) and RNA PolII sites relative to transcription start sites (TSS) of RefSeq genes. The scale on the left represents RNA PolII distribution, and the scale on the right represents estrogen receptor and random distribution. (**b**) Correlation of estrogen receptor and RNA PolII binding sites with each chromosome, ranked according to total gene number and total nucleotide number. (**c**) Conservation of all estrogen receptor binding sites (black line) and RNA PolII binding sites (red line) between human, mouse, rat, chicken and *Fugu rubripes* sequence. RNA PolII binding sites are shown in a 5′-to-3′ manner.

Mann-Whitney U-test[13] using a threshold of $P < 10^{-5}$ and a new model-based analysis tiling array algorithm, MAT[14]. This stringent approach ensures high confidence predictions to facilitate subsequent motif analysis, though it may introduce some false negatives with lower confidence (see the **Supplementary Note** for estrogen receptor and RNA PolII binding data at both the stringent and a lower threshold). The stringent threshold represents a false discovery rate (FDR) of ~1%, and the lower threshold represents an FDR of ~5%. After BLAT analysis[15] to eliminate redundant sequences, we identified a final set of 3,665 unique estrogen receptor binding sites and 3,629 unique RNA PolII binding sites using the stringent threshold, resulting in an estrogen receptor and RNA PolII binding site on average every 839 kb and 847 kb in the genome, respectively.

Correlation of binding with transcription start sites

We mapped the relative location of estrogen receptor and RNA PolII binding sites to transcription start sites (TSS) of known genes from RefSeq (**Fig. 1a**). Approximately 67% of RNA PolII sites map to promoter-proximal (–800 bp to +200 bp) regions of known genes, consistent with findings reported for transcription factor IID (TFIID)[16]. Identification of essential elements for estrogen receptor–mediated transcription of target genes have focused primarily on promoter-proximal regions, and recent estrogen receptor location analyses analyzed only promoter regions[9,10]. However, in our complete genome-wide approach, we find that only 4% of estrogen receptor binding sites mapped to 1-kb promoter-proximal regions at either the high or low threshold (**Fig. 1a**), and as such, almost all *in vivo* estrogen receptor binding events occur in regions previously unannotated as *cis*-regulatory elements within the genome. The low frequency of promoter-proximal binding sites found for estrogen receptor is unlikely to be due to a bias in the method, as we were able to find the vast majority of RNA PolII binding sites at promoters using this method as expected. However, within the list of estrogen receptor binding sites near promoter-proximal regions, we found a number of previously identified estrogen receptor targets, including *TFF1*, *EBAG9*, *TRIM25* (also known as *Efp*), *ESR1* and prothymosin α (*PTMA*), found using the stringent threshold, and cathepsin D (*CTSD*), *PGR* (also known as *PR*), keratin 19 (*KRT19*), *RARA* (also known as *RARα*) and *HSPB1* (also known as *Hsp27*), found using the more relaxed threshold (reviewed in refs. 17,18). Even when a very relaxed cutoff was analyzed corresponding to an FDR of >50%, only three additional promoter-proximal regions previously suggested to be estrogen receptor targets were identified (**Supplementary Table 1** online). The promoters identified using the lower thresholds may

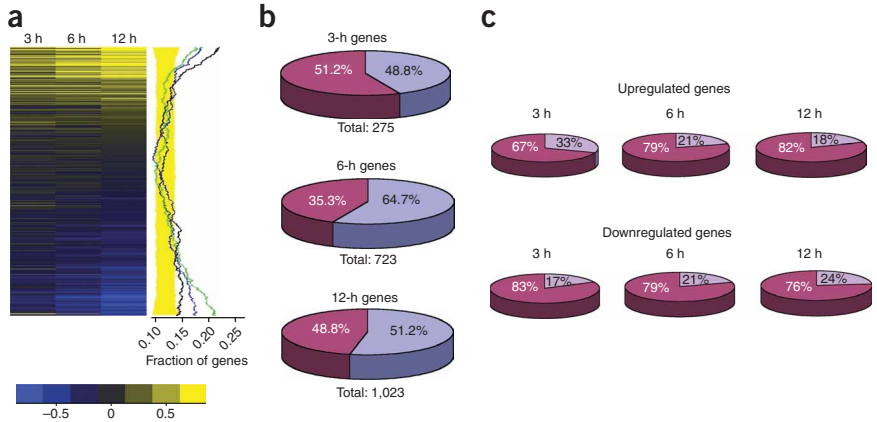

Figure 2 Estrogen-mediated transcript changes and correlation with estrogen receptor binding. (**a**) Expression changes of all genes as ranked by Welch *t* statistic at 3, 6 and 12 h relative to 0 h. Induction of gene expression relative to 0 h is represented as yellow and repression as blue. The graph represents the fraction of genes with an estrogen receptor binding site within 50 kb of the transcription start site. Genes were ranked by Welch *t* statistic between 3, 6 and 12 h and 0 h (control). The black (3 h), blue (6 h) and green (12 h) lines represent 2,000 gene moving averages of the fraction of genes that have one or more estrogen receptor binding sites within 50 kb of the transcription start site. The yellow band is a 99% confidence interval for the binding site moving average of genes in the 25%–50% 12-h *t* statistic range. (**b**) Summary of estrogen-mediated expression changes over a time course (0, 3, 6 and 12 h). Shown are the number of differentially expressed genes after estrogen treatment, relative to the vehicle-treated control (0 h). Blue segments represent upregulated genes, and red segments represent downregulated genes. (**c**) Percentage of genes upregulated or downregulated at each time point (relative to time 0 h) that contain estrogen receptor binding sites within 50 kb (purple sector).

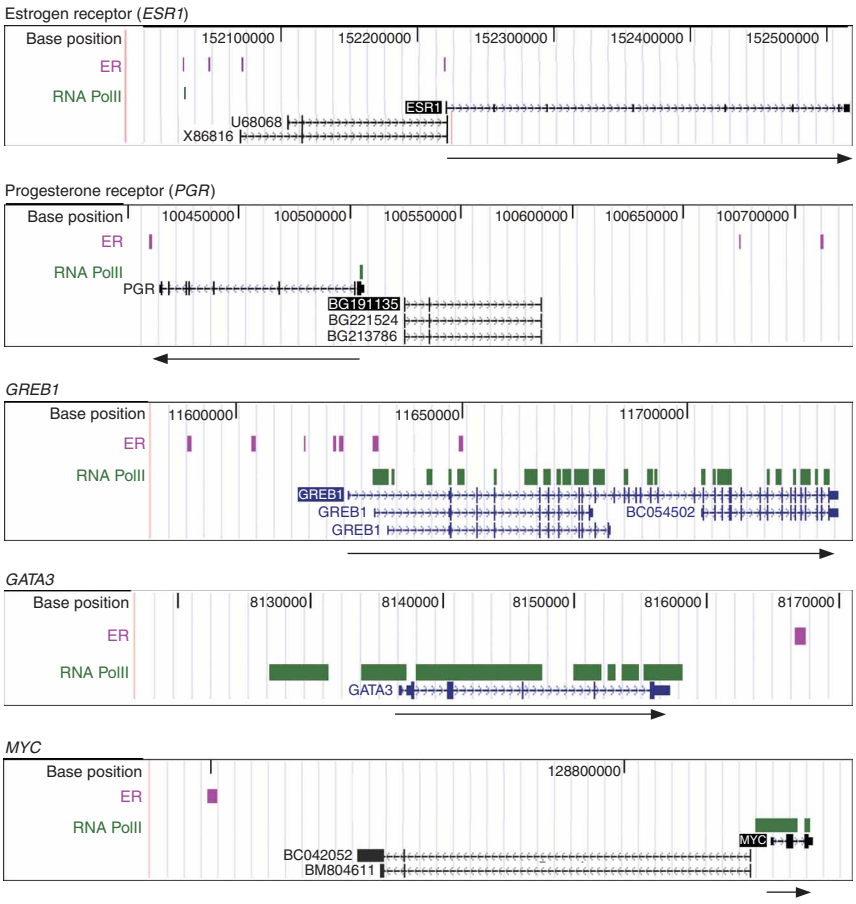

Figure 3 Estrogen receptor and RNA PolII binding relative to specific gene targets. The purple blocks represent estrogen receptor (ER) binding sites, and green blocks represent RNA PolII sites. *ESR1*, *GREB1*, *MYC* and *GATA3* are shown in their genuine 5′-3′ orientation, and *PGR* is shown in its genuine 3′-5′ orientation. The black arrows indicate the direction of the gene. Included are predicted transcripts that exist between the estrogen receptor binding sites and the target genes.

represent indirect or secondary binding sites, as assessed by the low enrichment (1.2- to 1.8-fold over background) by directed quantitative ChIP (**Supplementary Fig. 1** online and data not shown), compared with 15- to 160-fold for adjacent estrogen receptor binding sites distal from promoter regions.

Conserved *cis* elements define estrogen receptor binding

RNA PolII binding correlated well ($r^2 = 0.88$) with gene number, not chromosome length ($r^2 = 0.29$), as its binding was predominately promoter proximal (**Fig. 1b**). Compared with RNA PolII, estrogen receptor binding was less well correlated with gene number ($r^2 = 0.62$) and equally correlated with chromosome size, as estrogen receptor binding is distributed within and between genes rather than being restricted to promoters (**Fig. 1b**).

Sequence comparison of all the estrogen receptor binding sites between the genomes of multiple vertebrate species showed high conservation within the binding sites, but not in immediate surrounding regions (**Fig. 1c**); conservation was almost to the same level as for coding sequences. Conservation analysis of RNA PolII binding sites showed a similar degree of sequence preservation, although in contrast to estrogen receptor, this was also maintained in the surrounding coding sequence (**Fig. 1c**). Therefore, the evolutionary maintenance of the estrogen receptor binding sites supports their putative role as functional *cis*-regulatory domains distinct from promoters.

Gene expression correlates with binding

To correlate estrogen receptor and RNA PolII binding data with the estrogen transcriptional response, we performed gene expression profiling by microarray analyses, which were performed in triplicate over an estrogen stimulation time course (0, 3, 6 and 12 h), with 3 h representing immediate transcriptional targets[7] and both 6 and 12 h representing delayed targets (complete data sets are available; see **Supplementary Note**). Relative to time 0 h, 134 genes were upregulated after 3 h of estrogen treatment (**Fig. 2a**,**b**), which is a small fraction of the RNA PolII binding sites present in MCF-7 cells under these conditions. However, RNA PolII binding sites identified by ChIP-on-chip represent not only the genes differentially regulated by estrogen, but also estrogen-independent binding sites within actively transcribed genes[13].

Correlation of estrogen receptor binding sites with early (3 h) and late (6 h and 12 h) estrogen-induced genes showed a bias of binding sites within 50 kb of TSS of both early and delayed estrogen-induced genes ($P < 0.001$) (**Fig. 2a**,**c**). Although there is significantly greater estrogen receptor binding bias toward early upregulated genes, the bias observed near late-upregulated genes suggests that either these late transcripts are produced early and do not accumulate to detectable levels for more than 3 h, or more likely, their transcription requires estrogen induction of a secondary or cooperating transcription factor.

Estrogen-mediated gene repression

Most work investigating estrogen-regulated transcription focuses on upregulated genes, although downregulated genes constitute a significant fraction of all estrogen-dependent expression changes in cell lines[19] and tumor samples[20]. In our expression array analysis, 51.2% of early (3 h) gene changes are downregulated events (**Fig. 2b**). Of the different possible mechanisms for this early gene inhibition, one hypothesis is a sequestration of limiting factors away from downregulated genes[21], so-called physiologic squelching. In support of this hypothesis, correlation of estrogen receptor binding sites with downregulated genes did not show any statistical bias to the TSS of genes downregulated at 3 h (**Fig. 2a**). We took several different experimental approaches to assess if physiologic squelching was a primary mode of early downregulation. RNA PolII binding at the promoters of early-downregulated genes decreased after only 45 min of estrogen stimulation, coincident with RNA PolII binding at promoters of early-upregulated genes (data not shown). Furthermore, pretreatment of MCF-7 cells with the translational inhibitor cycloheximide for 1 h before estrogen stimulation did not influence the early decreases in a number of assessed transcripts (**Supplementary Fig. 2** online), suggesting that these genes are primary, yet indirect, targets of estrogen receptor action.

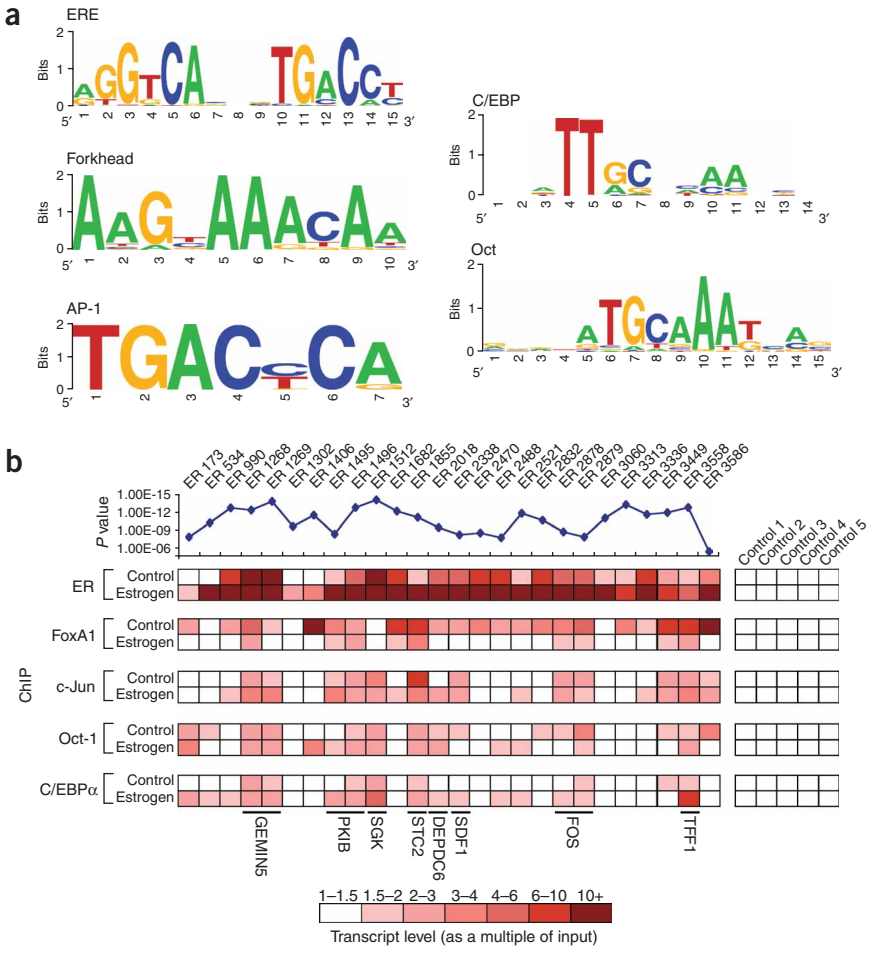

Figure 4 Identification of enriched motifs within the estrogen receptor binding sites and validation of transcription factor binding. (**a**) A computational screen for enriched motifs within all estrogen receptor binding regions demonstrates the presence of ERE, Forkhead, AP-1, Oct and C/EBP sites, with nucleotide bias shown using Weblogo (http://weblogo.berkeley.edu/). A complete list of enriched motifs can be found in **Supplementary Table 2**. (**b**) Directed ChIP of transcription factors that bind to these enriched motifs was performed on 26 estrogen receptor (ER) binding sites and five control regions. The binding sites were chosen to cover a range of enrichment values but also included sites near a select number of estrogen-regulated genes. The relative *P* value for each of the binding sites assessed is provided. Estrogen receptor binding sites adjacent to estrogen-regulated genes are shown by the gene name. The real-time PCR data are shown as a multiple of input DNA and are the average of independent replicates.

In contrast to the early-downregulated genes, when we mapped the relationship between estrogen receptor binding and the TSS of genes downregulated at the later 6- and 12-h time points, we observed a significant enrichment of estrogen receptor binding sites within 50 kb of promoter regions (**Fig. 2a**). This bias of estrogen receptor binding adjacent to late-downregulated genes suggests that in contrast to the majority of early-downregulated genes, which are likely to result from a preponderance of indirect mechanisms such as physiologic squelching, most downregulation late requires estrogen receptor binding. The lag suggests the necessity for the transcription of an estrogen-induced repressor or corepressor capable of associating with chromatin-bound estrogen receptor to facilitate subsequent transcriptional inhibition of adjacent genes. In support of this hypothesis, pretreatment of MCF-7 cells with cycloheximide before estrogen stimulation abrogated the late downregulation of a number of assessed transcripts (**Supplementary Fig. 2**), confirming the requirement for translation of a secondary factor.

Diversity of estrogen receptor regulatory mechanisms

The ChIP-on-chip data suggest that a diversity of binding profiles exist. As examples, autoregulation of the *ESR1* gene involved estrogen receptor binding at the promoter as previously implicated[22] but also may involve three estrogen receptor binding sites 150 kb to 192 kb upstream of the gene (**Fig. 3**). The gene encoding the progesterone receptor, a steroid receptor that is critical in female reproduction and lactation[23] and pathological in breast cancer, contained a RNA PolII binding site at the promoter and two estrogen receptor binding sites 168 kb and 206 kb upstream of the gene. In contrast, approximately half of early, direct estrogen-upregulated genes have estrogen receptor binding sites within 100 kb. As examples, *GREB1*, an estrogen-regulated gene[24] with no previously identified mechanism of estrogen regulation, contained RNA PolII and an estrogen receptor binding site at the promoter of a specific isoform, as well as a cluster of five other estrogen receptor sites upstream of the gene. *GATA3*, a transcription factor that correlates with estrogen receptor status in breast cancer cells[25], contained one estrogen receptor binding site close to the 3′ end of the gene. Previous work delineating mechanisms of estrogen induction of *MYC* have implicated non-estrogen-responsive elements (EREs) within the promoter[26] as the estrogen receptor binding site[7], but we observed a single estrogen receptor binding site approximately 67 kb upstream from *MYC*. We validated estrogen receptor binding to most of this subset of newly identified binding sites using directed estrogen receptor ChIP and real-time PCR (**Supplementary Fig. 1**). In support of the ChIP-on-chip data, estrogen receptor binding was only marginally enriched at the *MYC* promoter by ChIP and quantitative PCR (1.5-fold over input DNA) compared with the newly identified upstream enhancer (15-fold over input DNA), substantiating the assertion that the *MYC* promoter is not the primary estrogen receptor binding site. It should be noted that in the cases of *ESR1*, *PGR* and *MYC*, predicted transcripts exist in the region between the binding sites and the hypothesized target, as shown in **Figure 3**, although there is no evidence for their expression in MCF-7 cells. Future studies will be needed in order to prove the particular functional significance of any of these estrogen receptor binding sites; however, in the absence of this unique resource, the existence of these sites would be unknown. These examples typify the gene-specific complexity of estrogen receptor transcriptional regulation and reinforce the concept that the historical bias towards promoter-proximal regions does not fully identify the primary sites of estrogen regulation in most cases.

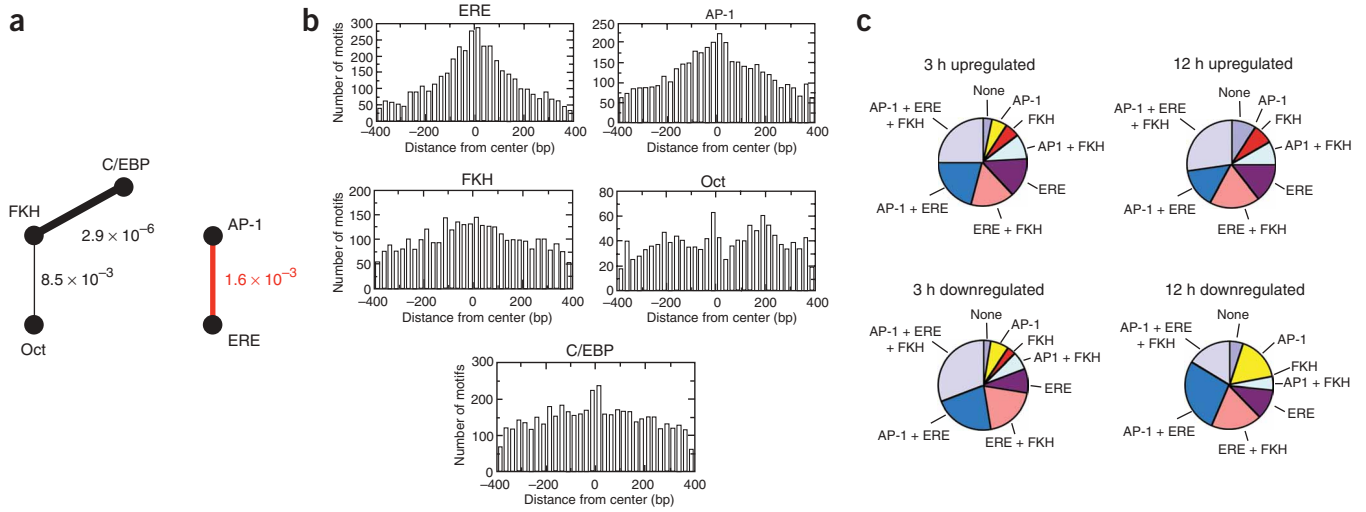

Figure 5 Involvement of cooperating transcription factors at estrogen receptor binding sites. (**a**) Pairwise analysis between ERE, Forkhead (FKH), AP-1, Oct and C/EBP motifs. A positive correlation is shown as a black line, and a negative correlation is shown as a red line. Statistical significance is shown numerically and also indicated by line thickness. (**b**) Distribution of ERE, Forkhead (FKH), AP-1, Oct and C/EBP motifs within estrogen receptor binding sites relative to the center of the binding sites (represented as 0). (**c**) Fraction of specific binding sites containing ERE, AP-1 and Forkhead (FKH) motifs adjacent to genes up- or downregulated early (3 h) or late (12 h). The top 200 differentially expressed genes at each time point (based on the Welch *t* test) were included in the analysis. For each gene, only motifs in the nearest ChIP region within 50 kb were considered.

Involvement of cooperating factors

To systematically identify the network of transcription factors that modulate estrogen receptor function, we searched all estrogen receptor binding sites for enriched DNA binding elements by both *de novo* and candidate scanning approaches. This screen identified EREs and Forkhead motifs, as previously implicated[11], as well as a number of other putative binding motifs (a complete list of enriched motifs can be found in **Supplementary Table 2** online), including AP-1, Oct and C/EBP motifs (**Fig. 4a**), supporting the suggestion that these sites serve as enhancers. Using ChIP followed by real-time PCR of 15 randomly selected estrogen receptor binding sites with different enrichment values, 11 sites adjacent to estrogen regulated genes and five negative controls (regions containing EREs or ERE half sites, but not identified as estrogen receptor binding sites) (**Supplementary Table 3** online), we confirmed estrogen receptor recruitment to all of the tested ChIP-on-chip–identified sites but not to any of the negative controls (**Fig. 4b**). FoxA1 binding occurred at most of these sites (but not at any of the controls), and the signal was generally diminished after estrogen addition, as we previously found for sites on chromosomes 21 and 22 (ref. 11) (**Fig. 4b**).

To validate specific transcription factor association with the enriched AP-1, Oct and C/EBP motifs, we focused initially on members of each transcription factor family that were abundant in MCF-7 cells. As an example, Oct-1 was expressed in MCF-7 cells, and Oct-1 protein was shown by ChIP to be recruited to a number (73%) of the assessed sites (**Fig. 4b**), supporting the data showing Oct-1 as a nuclear receptor–interacting transcription factor[27] and a putative regulator of estrogen target genes[28]. Similarly, c-Jun and C/EBPα were shown to bind to a subset of estrogen receptor binding sites, but not to the negative controls. C/EBPα has been shown to interact with estrogen receptor in GST pull-down experiments[29], and c-Jun has an extensively characterized role modulating estrogen target genes[30,31], although general roles for these transcription factors in estrogen receptor–mediated transcription have not been previously shown. Importantly, these motifs were not statistically enriched in the promoter-proximal regions of estrogen-regulated genes (data not shown).

We performed pairwise analysis to identify combinatorial interactions between ERE, Forkhead, Oct, AP-1 and C/EBP motifs within all estrogen receptor binding sites and found a strong negative correlation between ERE and AP-1 elements (**Fig. 5a**), suggesting that ERE and AP-1 motifs occur exclusively. The pairwise analysis also showed a positive correlation between C/EBP, Oct and Forkhead motifs (**Fig. 5a**), implying that these motifs tend to cluster together within the same estrogen receptor binding sites. The C/EBP, Oct and Forkhead motif cluster had equal likelihood of occurring with ERE or AP-1 motifs.

The relative positional distribution of the enriched motifs within the estrogen receptor binding sites show that both ERE and AP-1 motifs typically occur at the center of the estrogen receptor binding sites (**Fig. 5b**), whereas Forkhead, C/EBP and Oct motifs were less biased toward the center of the binding sites, possessed a more even distribution across the estrogen receptor binding sites and, in the case of Oct motifs, seemed to be multimodal, with clusters occurring approximately 200 bp on both sides of the center of the binding sites. This suggests that the primary interaction of estrogen receptor with chromatin can occur either through direct interaction with an ERE or through a tethering mechanism involving AP-1 factors, as previously suggested[18,31], with C/EBP, Oct and Forkhead[32,33] motifs functioning as adjacent binding sites for cooperating factors.

NRIP1-mediated gene repression

We next investigated whether there were functional differences between estrogen receptor binding sites centered on an ERE versus an AP-1 motif in binding sites adjacent to the highest differentially regulated genes. In contrast to the early-regulated genes, there was a clear bias of AP-1–centered estrogen receptor binding sites adjacent to late (12 h)-downregulated versus late-upregulated genes ($P < 0.01$; **Fig. 5c**). As this bias in AP-1 motifs was not observed early, it suggested that the late direct estrogen receptor binding–mediated

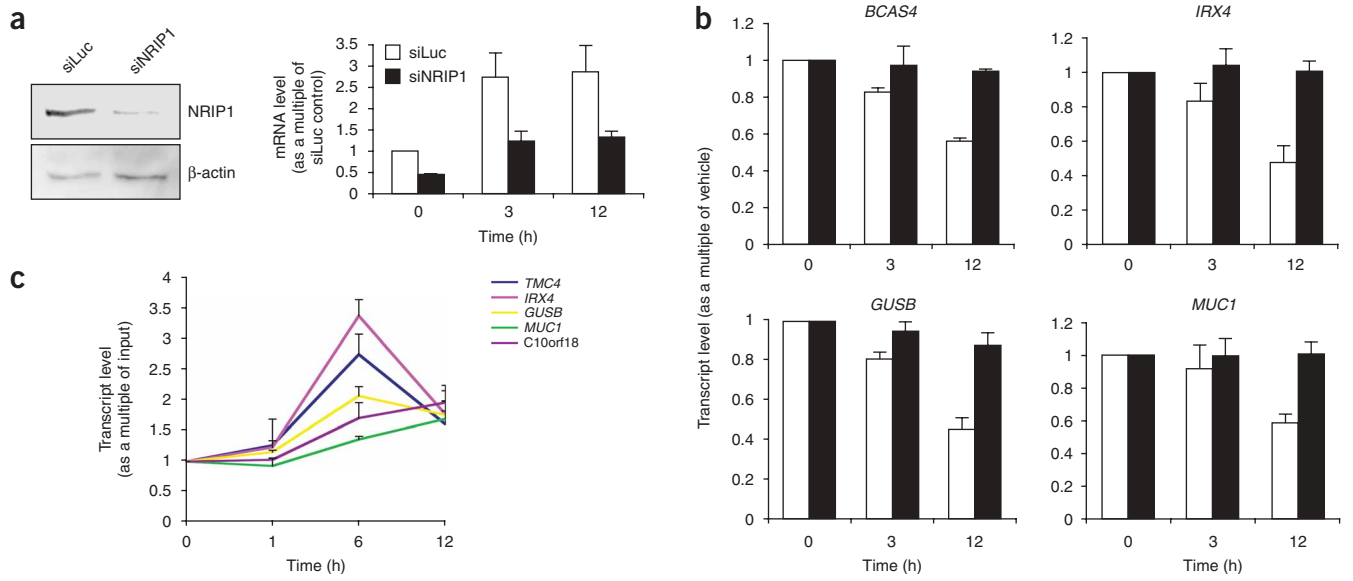

Figure 6 The role of NRIP1 in mediating gene repression. (**a**) siRNA to control (siLuc) or NRIP1 was transfected into hormone-depleted MCF-7 cells, and NRIP1 protein levels were assessed (left). β-actin functioned as a loading control. *NRIP1* mRNA levels were assessed after estrogen stimulation in the presence of control (siLuc) or siNRIP1. (**b**) Transcript levels of candidate late-downregulated genes with estrogen receptor binding sites containing AP-1 elements (*BCAS4*, *IRX4*, *GUSB* and *MUC1*) were assessed after siLuc control or siNRIP1 transfection and subsequent estrogen stimulation. The data are normalized to vehicle-treated conditions. (**c**) We assessed NRIP1 recruitment to estrogen receptor binding sites containing AP-1 elements adjacent to late-downregulated genes by NRIP1 ChIP after estrogen treatment for increasing time periods. Real-time PCR was performed on the estrogen receptor binding sites and data were normalized to vehicle-treated conditions. The data are the mean of independent replicates ± s.d.

transcriptional inhibition might be mediated via an estrogen-induced factor capable of interaction with estrogen receptor tethered to AP-1 motifs.

We therefore searched for genes that were estrogen induced at the early (3 h) time point that were known to interact with either estrogen receptor or AP-1 proteins. One such candidate was the coregulator NRIP1, which (i) is upregulated at 3 h of estrogen treatment, (ii) is a nuclear receptor corepressor[34] and (iii) has been shown *in vitro* to specifically antagonize estrogen receptor transcription via its interaction with AP-1 proteins[35].

To assess whether NRIP1 was a required factor mediating late, direct gene repression via estrogen receptor binding to AP-1–containing elements, we developed short interfering RNA (siRNA) to the *NRIP1* transcript and transfected this into hormone-depleted MCF-7 cells. NRIP1 protein levels were effectively reduced after specific siRNA transfection, and the early estrogen-induced accumulation of *NRIP1* transcript in control siRNA-treated cells was significantly inhibited by the presence of siNRIP1 (**Fig. 6a**).

We next measured transcript levels by quantitative RT-PCR of several late (12 h after estrogen treatment) downregulated genes that contained adjacent estrogen receptor binding sites centered on AP-1 elements, including *BCAS4*, *IRX4*, *GUSB* and *MUC1*. All of these target genes were substantially downregulated at 12 h by estrogen, but these effects were markedly reversed in the presence of siNRIP1 (**Fig. 6b**), demonstrating that NRIP1 is necessary for the downregulation of these genes. We found that a number of control target genes that are upregulated late by estrogen were unaffected by the presence of siRNA to *NRIP1* (data not shown). Furthermore, NRIP1 ChIP followed by real-time PCR of the estrogen receptor binding sites adjacent to these late-downregulated genes confirmed NRIP1 binding at either 6 or 12 h of estrogen treatment (**Fig. 6c**).

Function of binding sites in human breast cancers

In order to determine whether the estrogen receptor binding sites defined in MCF-7 cells is cell line specific, we assessed the function of a subset of estrogen receptor binding sites in another estrogen receptor–positive breast cancer cell line, T47D. All of the small subset of tested sites functioned as estrogen receptor binding sites in another breast epithelial cell line (**Fig. 7a**).

To test whether the estrogen receptor binding sites as defined in MCF-7 cells are relevant to the pattern of gene expression observed in authentic human breast cancers, we compared the estrogen receptor binding with the gene expression signatures from two independent studies, one involving 286 primary breast tumors[4] and the other 295 breast tumors[3]. When we compare the position of an estrogen receptor binding site with the genes correlated with estrogen receptor expression in each of the two studies we find a significant (Wang, $P < 3.0 \times 10^{-8}$, and van de Vijver, $P < 1.0 \times 10^{-6}$) enrichment of estrogen receptor binding adjacent to the positively correlated genes (**Fig. 7b**). The percentage of genes with estrogen receptor binding sites within 100 kb are 56% and 59% for the van de Vijver and Wang studies, respectively. This relationship is very similar to the one found for estrogen-regulated (3 h) genes in MCF-7 cells of ~50%. As a comparison, we examined estrogen receptor binding profiles adjacent to estrogen-regulated genes in MCF-7 cells (**Fig. 7c**). This result suggests that the estrogen receptor binding profile identified in MCF-7 cells both predicts the gene expression signature and identifies functional regions of the genome that control estrogen responses in primary human breast cancers.

DISCUSSION

The identification of the set of *cis*-acting targets of a *trans*-acting factor such as the estrogen receptor across the whole genome provides an important new resource for the study of gene regulation. The classic

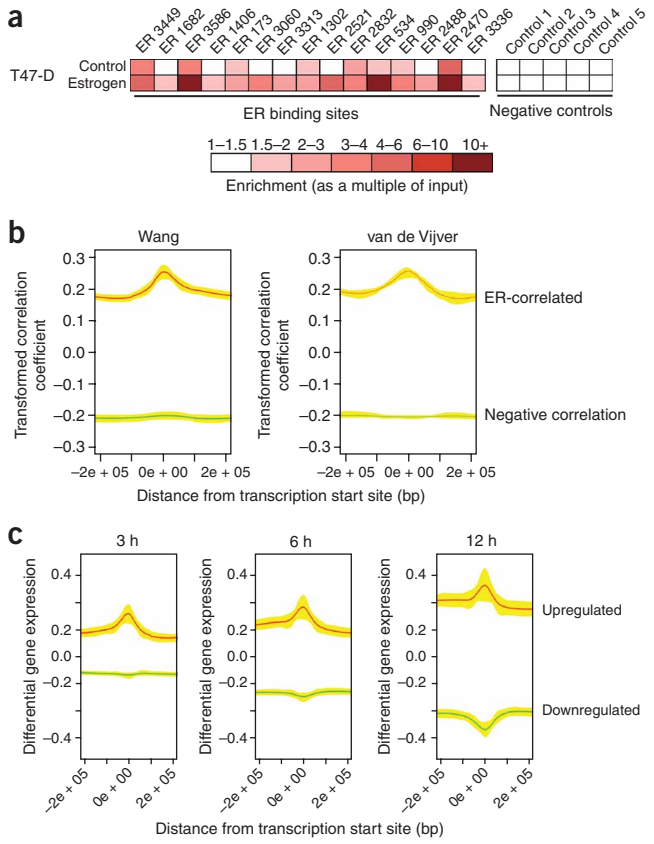

Figure 7 Assessment of estrogen receptor binding properties in different cell systems. (**a**) Estrogen receptor (ER) ChIP was performed after vehicle or estrogen stimulation in T-47D breast cancer cells. Real-time PCR of estrogen receptor binding sites previously identified in MCF-7 cells was performed, and the data are shown as a multiple of input. (**b**) Correlation of estrogen receptor binding sites relative to transcription start sites of the highest estrogen receptor–correlated genes from two independent primary breast cancer gene expression studies. (**c**) Correlation of estrogen receptor binding sites with transcription start sites of genes either estrogen-upregulated or estrogen-downregulated at 3, 6 or 12 h (relative to 0 h) in MCF-7 cells.

paradigm of estrogen receptor function involves binding to promoter-proximal regions and subsequent gene regulation. However, it now seems that the promoter-proximal regions, although important for some genes, do not constitute the majority of estrogen receptor target sites. Instead, it is apparent that a full definition of estrogen receptor binding to *cis*-regulatory regions distinct from promoters is required to fully understand the estrogen response. Similar analyses of c-Myc, p53 and Sp-1 binding to chromosomes 21 and 22 has also shown analogous enhancer binding profiles[36], suggesting that studies that focus on promoter regions[9,10] are insufficient. In contrast, TFIID[16] and RNA PolII ChIP-on-chip analyses (in this investigation) confirm that the basal transcription machinery is significantly biased to promoter-proximal regions. In general, it seems that communication is often mediated at great distances between the transcription factors that initiate gene expression events and the transcription machinery that execute it.

Almost one-third of early–estrogen upregulated genes have estrogen receptor binding sites within 50 kb of the TSS, confirming a clear statistical bias for regulation of genes in the vicinity of chromatin-interaction sites. Other estrogen-stimulated genes that do not have an estrogen receptor binding site within 50 kb may use sites that are greater than 50 kb from the gene[11], use enhancers on different chromosomes[37] or induce transcription independent of estrogen receptor binding events. It is of interest to note that there are many more estrogen receptor binding sites in the genome than differentially regulated genes, as has been previously suggested[38]. It is likely that a significant number of these binding sites are not functional in MCF-7 cells under the specific experimental conditions used and may be functional in other cell types or under different conditions.

Although previous work has shown numerous estrogen receptor–cooperating proteins at the promoters of estrogen-regulated genes[6,8], we find that transcriptional activity of estrogen receptor from the *cis*-regulatory elements also involves combinations of cooperating transcription factors. We previously found an enrichment of Forkhead motifs within estrogen receptor binding sites on chromosomes 21 and 22 and subsequently showed a requirement for FoxA1 in mediating estrogen receptor binding to chromatin[11] supporting the role of FoxA1 as a pioneer factor[32,33]. Using the statistical power of all 3,665 estrogen receptor binding sites in the entire human genome, we both confirmed the role of FoxA1 and identified several additional enriched motifs that were not identified in our previous investigation[11], including DNA-binding motifs for AP-1, C/EBP and Oct transcription factors. Previous work has shown an estrogen-dependent role for c-Jun, Oct-1 and C/EBP proteins in transcription of cyclin D1 (ref. 28), but the unbiased identification of these binding motifs within estrogen receptor binding sites suggests a more general role for these cooperating factors in estrogen receptor transcription.

AP-1 family members have an extensively characterized role in estrogen receptor–regulated transcription[31], and the estrogen receptor can bind to DNA via ERE or AP-1 elements[18,30] involving different protein complexes[39]. A positive role for AP-1 proteins in the estrogen-mediated induction of target genes is established, but we now show a role for AP-1 proteins in gene repression. Our data show that gene changes that occur late (at 6 and 12 h of estrogen stimulation) can be clearly divided into two categories: genes that are upregulated, which have adjacent estrogen receptor binding sites more likely to contain EREs, and genes that are downregulated, which generally contain AP-1 elements. We now show the mechanisms defining these two classes of estrogen receptor binding sites, with estrogen inducing the corepressor NRIP1, which subsequently interacts with estrogen receptor–AP-1 complexes[35] to effect direct repression of adjacent target genes. Our previous work identified the mechanism of estrogen receptor–mediated *NRIP1* induction: several distant enhancers (~150 kb from the TSS of *NRIP1*) function as primary estrogen receptor binding sites, and chromatin loops between these *NRIP1* enhancers and its promoter exist in the presence of estrogen[11].

The estrogen receptor is critical in determining the phenotype of human breast cancers and is the most important therapeutic target. The complete set of estrogen receptor binding sites across the genome defined in these studies establishes a new resource for understanding estrogen action in breast cancer. It correctly predicts the genes coexpressed with the estrogen receptor in primary breast tumors and thus identifies important and previously unexplored regions of the genome that are the critical regulators of the estrogen dependence of breast cancer.

METHODS

ChIP-on-chip analysis. ChIP and chromatin preparation were performed as previously described[11,40,41]. We used antibodies to ERα (Ab-10; Neomarkers,

Lab Vision); ERα (HC-20) and RNA PolII (H-224) (Santa Cruz) and RNA PolII (4H8; Abcam). All three replicates were performed on the Affymetrix Human tiling 1.0 microarrays (14-chip set). The only difference between replicates is that the Affymetrix imagine software GCOS rotated the CEL files 90° in the first two replicates but not in the third replicate. We applied the generalized Mann-Whitney *U* test[13] to identify regions at least 600 bp in length that were enriched in ChIP samples compared with the controls. A total of 5,712 regions were predicted at the *P* value cutoff of 1×10^{-5}. MAT[14] was applied to the same data to predict the highest-scoring 5,712 ChIP regions (equivalent to a MAT score cutoff of 10.27 and a *P* value of 7.1×10^{-6}). The two predictions had a high degree of concordance, and we reported the intersection between them. In addition, 17 regions predicted by MAT as the top 1,000 but missed by the generalized Mann-Whitney method were added to the final list of estrogen receptor binding sites. BLAT analysis[15] was performed to eliminate redundant sequences.

Expression microarrays. MCF-7 cells were deprived of hormones as previously described[11] and stimulated with 100 nM estrogen for 0, 3, 6 or 12 h, after which total RNA was collected using Trizol (Invitrogen). Expression microarrays were Affymetrix U133Plus2.0, and all experiments were performed in triplicate. Data were analyzed using the RMA algorithm[42] with the newest probe mapping[43], and the Welch *t* statistic was used to calculate the level of differential expression at each time point relative to 0 h.

Directed ChIP and real-time PCR. ChIP was performed as previously described[11]. We used antibodies to ERα (Ab-10; Neomarkers, Lab Vision); ERα (HC-20), HNF-3α/FoxA1 (H-120), c-Jun (N), Oct-1 (C-21), C/EBPα (14AA) and NRIP1 (H-300) (Santa Cruz); and NRIP1 (ab3425; Abcam). Quantitative real-time PCR was performed as previously described[11].

siRNA. siRNA experiments were performed as previously described[11]. *NRIP1* siRNA sequences (Dharmacon) were as follows: sense, 5′-GAAGCGUG CUAACGAUAAA-3′, and antisense, 5′-UUUAUCGUUAGCACGCUUC-3′. Antibodies used in the protein blot were NRIP1/RIP-140 R5027 (Sigma Aldrich) and β-actin A1978 (Sigma Aldrich).

Real-time RT-PCR. RNA was collected as described above. Real-time RT-PCR was performed as described above for real-time PCR, with the exception that 10 units of MultiScribe (Applied Biosystems) were added, and a reverse transcription step of 48 °C for 30 min was included before PCR cycling. Primer sequences can be found in **Supplementary Table 3**.

Sequence conservation analysis. The 3,665 estrogen receptor ChIP regions were aligned at their centers and uniformly expanded to 3,000 bp in each direction, and phastCons[44] scores were retrieved (http://genome.ucsc.edu) and averaged at each position.

Screen of estrogen receptor binding sites for enriched motifs. The ChIP regions and 3,800 promoters of non-differentially expressed RefSeq genes located within 200 kb of the ChIP regions were scanned for transcription factor motifs using 533 well-defined position-specific score matrices (PSSM) from TRANSFAC[45], JASPAR[46] and ref. 11. The background nucleotide frequencies were computed from the whole genome. For each matrix, we considered all PSSM matches with cutoff scores from 5.0 (90% of relative entropy) up to 12.0, in increments of 0.5. At each cutoff level, the resulting two sets of motifs were then tested for significance using the criteria of binomial $P < 1 \times 10^{-4}$ and minimum change (with respect to control) of 1.5-fold. We report the relevant statistics for only those PSSM score cutoffs with maximum changes with respect to control. In addition to the PSSM scan, we performed *de novo* motif scans using LeitMotif [47], a modified MDscan[48] with ninth-order Markov dependency for the genome background.

Conditional independence graphical models[49] were constructed to understand the association of transcription factors. The 3,665 estrogen receptor ChIP regions were uniformly resized to 400 bp in each direction from their centers. PSSM scans for ERE, Forkhead, AP-1 and Oct were performed with 90% of relative entropy (RE) cutoff and for C/EBP at a cutoff of 5.0 because of its very low RE. The PSSM scores were then normalized as (score – RE)/motif length, and when two motifs overlapped, only the motif with higher normalized score was kept. The resulting five-dimensional motif hit contingency table for the distribution of the motifs in estrogen receptor ChIP regions was then analyzed with MIM (http://www.hypergraph.dk) graphical modeling software. Using 100% relative entropy adds one more interaction edge between Oct-1 and C/EBP; the corresponding model is shown in **Supplementary Figure 3** online.

Correlation of estrogen receptor binding to gene expression profiles in tumor samples. We downloaded the gene expression index from 286 lymph node–negative individuals who had not received adjuvant systemic treatment[4] and 295 individuals with either lymph node–negative or lymph node–positive disease[3] from GEO (accession 2034) and http://www.rii.com/publications/2002/nejm.html, respectively. Pearson correlation coefficients of estrogen receptor expression relative to every other UCSC known gene were calculated within the Wang and van de Vijver data sets, respectively. Fisher's transformation of the correlation coefficient, $z = 0.5 \log((1 + c) / (1 - c))$, was fitted to the oriented distance to the nearest estrogen receptor ChIP region. A cubic spline with 11 knots between –1 Mb and +1 Mb with equal numbers of data points between knots was applied to smooth the graph (**Fig. 7b**).

URLs. Data to accompany the **Supplementary Note** can be downloaded from http://research.dfci.harvard.edu/brownlab/datasets/index.php?dir=ER_whole_human_genome/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors would like to thank D. Lynch, M. Lupien, M. Laza and O. Jänne for helpful discussions and review of the manuscript. This work was supported by grants from the US National Institute of Diabetes and Digestive and Kidney Diseases (DK074967) and the US National Cancer Institute–sponsored Dana-Farber/Harvard Cancer Center Breast Cancer SPORE Grant (CA089393); by the Claudia Adams Barr Program in Innovative Basic Cancer Research (M.B., P.A.S., X.S.L.); by US Department of Defense Breast Cancer Research Program Awards (M.B., J.S.C.) and by fellowships from the Fondation Recherche Medicale (J.E.), the Susan G. Komen Breast Cancer Foundation (E.K.K.) and the US National Institutes of Health (T.R.G. and C.A.M.).

AUTHOR CONTRIBUTIONS

This study was designed by J.S.C., C.A.M., A.S.B., P.A.S., X.S.L., T.R.G. and M.B. The experimental procedures were performed by J.S.C., with assistance in specific areas from T.R.G., J.E., E.K.K., K.C.F. and Q.W. Biostatistical support was provided by C.A.M., J.S., W.L. and X.S.L. Microarray support and analysis was provided by G.F.H., S.B., V.S., E.A.F. and T.R.G.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics

Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869–10874 (2001).
2. van 't Veer, L.J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
3. van de Vijver, M.J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
4. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671–679 (2005).
5. Perou, C.M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
6. Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843–852 (2000).
7. Shang, Y. & Brown, M. Molecular determinants for the tissue specificity of SERMs. *Science* **295**, 2465–2468 (2002).
8. Metivier, R. *et al.* Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763 (2003).
9. Laganiere, J. *et al.* Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc. Natl. Acad. Sci. USA* **102**, 11651–11656 (2005).
10. Cheng, A.S. *et al.* Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters. *Mol. Cell* **21**, 393–404 (2006).

11. Carroll, J.S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long -range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43 (2005).
12. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154 (2005).
13. Brodsky, A.S. *et al.* Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol.* **6**, R64 (2005).
14. Johnson, W.E. *et al.* MAT: Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci. USA* (in the press) (2006).
15. Kent, W.J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
16. Kim, T.H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
17. Klinge, C.M. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res.* **29**, 2905–2919 (2001).
18. O'Lone, R., Frith, M.C., Karlsson, E.K. & Hansen, U. Genomic targets of nuclear estrogen receptors. *Mol. Endocrinol.* **18**, 1859–1875 (2004).
19. Frasor, J. *et al.* 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**, 4562–4574 (2003).
20. Nishidate, T. *et al.* Genome-wide gene-expression profiles of breast-cancer cells purified with laser microbeam microdissection: identification of genes associated with progression and metastasis. *Int. J. Oncol.* **25**, 797–819 (2004).
21. Kalaitzidis, D. & Gilmore, T.D. Transcription factor cross-talk: the estrogen receptor and NF-kappaB. *Trends Endocrinol. Metab.* **16**, 46–52 (2005).
22. Lazennec, G., Huignard, H., Valotaire, Y. & Kern, L. Characterization of the transcription start point of the trout estrogen receptor-encoding gene: evidence for alternative splicing in the 5′ untranslated region. *Gene* **166**, 243–247 (1995).
23. Lydon, J.P. *et al.* Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes Dev.* **9**, 2266–2278 (1995).
24. Ghosh, M.G., Thompson, D.A. & Weigel, R.J. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. *Cancer Res.* **60**, 6367–6375 (2000).
25. Hoch, R.V., Thompson, D.A., Baker, R.J. & Weigel, R.J. GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int. J. Cancer* **84**, 122–128 (1999).
26. Dubik, D. & Shiu, R.P. Mechanism of estrogen activation of c-myc oncogene expression. *Oncogene* **7**, 1587–1594 (1992).
27. Prefontaine, G.G. *et al.* Selective binding of steroid hormone receptors to octamer transcription factors determines transcriptional synergism at the mouse mammary tumor virus promoter. *J. Biol. Chem.* **274**, 26713–26719 (1999).
28. Cicatiello, L. *et al.* Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor (progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter. *Mol. Cell. Biol.* **24**, 7260–7274 (2004).
29. Boruk, M., Savory, J.G. & Hache, R.J. AF-2-dependent potentiation of CCAAT enhancer binding protein beta-mediated transcriptional activation by glucocorticoid receptor. *Mol. Endocrinol.* **12**, 1749–1763 (1998).
30. Umayahara, Y. *et al.* Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer. *J. Biol. Chem.* **269**, 16433–16442 (1994).
31. Kushner, P.J. *et al.* Estrogen receptor pathways to AP-1. *J. Steroid Biochem. Mol. Biol.* **74**, 311–317 (2000).
32. Cirillo, L.A. *et al.* Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* **17**, 244–254 (1998).
33. Cirillo, L.A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
34. Lee, C.H., Chinpaisal, C. & Wei, L.N. Cloning and characterization of mouse RIP140, a corepressor for nuclear orphan receptor TR2. *Mol. Cell. Biol.* **18**, 6745–6755 (1998).
35. Teyssier, C., Belguise, K., Galtier, F., Cavailles, V. & Chalbos, D. Receptor-interacting protein 140 binds c-Jun and inhibits estradiol-induced activator protein-1 activity by reversing glucocorticoid receptor-interacting protein 1 effect. *Mol. Endocrinol.* **17**, 287–299 (2003).
36. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
37. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R. & Flavell, R.A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).
38. van Steensel, B. Mapping of genetic and epigenetic regulatory networks using microarrays. *Nat. Genet.* **37**(Suppl.), S18–S24 (2005).
39. Webb, P. *et al.* The estrogen receptor enhances AP-1 activity by two distinct mechanisms with different requirements for receptor transactivation functions. *Mol. Endocrinol.* **13**, 1672–1685 (1999).
40. Ren, B. *et al.* E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev.* **16**, 245–256 (2002).
41. Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919 (2002).
42. Irizarry, R.A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
43. Dai, M. *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **33**, e175 (2005).
44. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
45. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
46. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
47. Song, J.S. & Liu, X.S. LeitMotif: *de novo* motif discovery for ChIP-Chip in mammalian genomes. In Preparation.
48. Liu, X.S., Brutlag, D.L. & Liu, J.S. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat. Biotechnol.* **20**, 835–839 (2002).
49. Darroch, J., Lauritzen, S. & Speed, T. Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8**, 522–539 (1980).