AD_____

Award Number: DAMD17-01-1-0222

TITLE: Dynamics of Estrogen Receptor Transcription Complex Assembly in Breast Cancer

PRINCIPAL INVESTIGATOR: Eli Hestermann, Ph.D.
Myles A. Brown, M.D.

CONTRACTING ORGANIZATION: Dana-Farber Cancer Institute
Boston, Massachusetts 02115

REPORT DATE: July 2004

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2004	Annual Summary (1 Jul 2001 - 30 Jun 2004)

4. TITLE AND SUBTITLE
Dynamics of Estrogen Receptor Transcription Complex Assembly in Breast Cancer

5. FUNDING NUMBERS
DAMD17-01-1-0222

6. AUTHOR(S)
Eli Hestermann, Ph.D.
Myles A. Brown, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dana-Farber Cancer Institute
Boston, Massachusetts 02115

E-Mail: Eli_hestermann@dfci.harvard.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

Estrogen plays a critical role in the development and progression of breast cancer. The biological activities of estrogen are mediated by estrogen receptors (ER). In addition, a large number of proteins termed cofactors are involved in ER signaling. Therapeutic agents, such as tamoxifen, also bind ER, but block proliferation in breast cells. However, tamoxifen increases the risk of endometrial cancer. We have used chromatin immunoprecipitation (ChIP) to investigate cofactor involvement in ER signaling *in vivo* and to understand the mechanisms underlying the different actions of tamoxifen in breast and endometrial cells. We have found that differences in cofactor expression underlie tissue-specific effects of tamoxifen. ChIP, in combination with tiled arrays of individual chromosomes, has been used to identify distant ER-binding sequences that regulate gene expression. Gene expression profiling has been used to identify differential regulation of ER targets in breast and endometrial cells, and these targets have been evaluated for their ability to regulate cell proliferation. The detailed understanding of tissue- and ligand-dependent changes in gene expression gained through these studies will lead to more effective therapies for ER-dependent breast cancer.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Estrogen receptor, coactivators, gene cloning, gene transcription	35
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Cover.....1

SF 298.....2

Introduction.....4

Body.....5

Key Research Accomplishments.....7

Reportable Outcomes.....7

Conclusions.....8

References.....8

Figures.....9

Appendix.....11

INTRODUCTION

Estrogen plays a critical role in the development and progression of breast cancer. The biological activities of estrogen are mediated by estrogen receptors (ER). In addition, a large number of proteins termed cofactors are involved in ER signaling. Therapeutic agents, such as tamoxifen, also bind ER, but block proliferation in breast cells. However, tamoxifen increases the risk of endometrial cancer. We have used chromatin immunoprecipitation (ChIP) to investigate cofactor involvement in ER signaling *in vivo* and to understand the mechanisms underlying the different actions of tamoxifen in breast and endometrial cells. We have found that differences in cofactor expression underlie tissue-specific effects of tamoxifen. ChIP, in combination with tiled arrays of individual chromosomes, has been used to identify distant ER-binding sequences that regulate gene expression. Gene expression profiling has been used to identify differential regulation of ER targets in breast and endometrial cells, and these targets have been evaluated for their ability to regulate cell proliferation. The detailed understanding of tissue- and ligand-dependent changes in gene expression gained through these studies will lead to more effective therapies for ER-dependent breast cancer.

BODY

(Note: In August 2003 the PI, Dr. Eli Hestermann, departed Dana Farber Cancer Institute. The award was not subsequently transferred to a new PI. The new data in this report represents the work of Dr. Hestermann in the period July – August 2003, as well as subsequent efforts by others to complete these projects. Dr. Hestermann has prepared the report.)

Task #1 To identify the coactivators that are involved in the estrogen-induced transcription complex and to determine the sequence of events and the dynamics involved in the assembly and disassembly of the transcription complex

Status: Complete (see previous report and Shang *et al* 2000 *Cell* 103:843-852)

Task #2 To compare the protein components of the tamoxifen-induced ER complex that occupies estrogen-responsive gene promoters in breast cancer cells and in endometrial cancer cells

a) Identify coactivators/corepressors that participate in tamoxifen-induced ER complex formation at target gene promoters in breast cancer cells
b) Identify coactivators/corepressors that participate in tamoxifen-induced ER complex formation at target gene promoters in endometrial cancer cells
c) Define the sequence of events that are involved in the assembly and disassembly of tamoxifen-induced ER complexes in both breast and endometrial cancer cells

Status: Objectives (a) and (b) accomplished for a several genes (see previous report and Shang and Brown 2002 *Science* 295:2465-2468). The new ChIP^2 technique (see task 3) makes this possible on a genome-wide scale.

The mechanisms of complex assembly and disassembly (objective c) have not been determined.

Task #3 To identify new estrogen- and tamoxifen-responsive genes in breast cancer cells and endometrial cancer cells

a) Isolate DNA fragments by ChIP and identify ER-regulated genes using microarrays.
b) Confirm the targets by quantitative RT-PCR of RNA from estrogen- and tamoxifen-treated breast and endometrial cells
c) Apply ChIP using antibodies for coactivators and corepressors to identify the components of ER complexes at each gene
d) Identify common and unique targets of ER in breast and endometrium

Status: The ChIP^2 technique (objective a) has been developed and has produced data on genome-wide promoter binding by ER and associated cofactors (objective c) – see Appendix. Expression microarray data has been mined to determine common and unique targets of ER in breast and endometrium (objective d), and targets of interest have been confirmed by quantitative RT-PCR (objective b). The ability of these targets to affect cell proliferation has also been determined.

During this final year of funding, the CHIP^2 technique has provided exciting new data on gene regulation by ER in breast cancer cells. Due to continuing technical difficulties, the printed DNA arrays used in previous years were abandoned in favor of tiled oligomer arrays that cover the sequence of entire chromosomes. Besides yielding more consistent results, these arrays will show **all** DNA binding sites, not just those in the specific promoters spotted on the previous arrays.

The details of this research are in the attached manuscript that has been submitted for publication (Appendix). In brief, ER-bound DNA was isolated from estrogen-treated cells, amplified, and hybridized to arrays covering chromosomes 21 and 22. Not only were known ER-binding sites identified by this technique, but sites over 100 kb distant from any known gene also showed specific ER binding. The *bona fide* activity of these sites is supported by recruitment of the cofactor AIB1 and RNA polymerase II to these sites, but not nearby control sites, as well as by their estrogen-dependent enhancement of reporter gene expression. Tamoxifen also caused ER recruitment to these sites, although not to the same extent as estrogen. Although the sites identified contained canonical ERE sequences, only a small fraction of all such sequences were occupied by ER. This same approach is now being used to probe the ER binding sites in the remaining chromosomes, as well as to expand the repertoire of cofactor binding sites.

The other advance made in the final year was in examining the functional significance of oncogenes that are regulated by estradiol in both breast and endometrial cells, but by tamoxifen only in endometrial cells. The genes c-myc, c-myb, and cyclin D1 had been shown to satisfy these criteria in previous experiments.We designed RNAi oligonucleotides to knock down expression (Figure 1). Consistent with earlier results, knockdown of c-myc also reduced cyclin D1 expression in both cell types and reduced c-myb expression in MCF-7 cells. Reducing expression of c-myb also affects levels of the other two genes in ECC-1 cells. This is unsurprising, since c-myc and c-myb are transfection factors that regulate expression of G1→S cyclins.

Following knock-down of expression by RNAi, the ability of MCF-7 breast cancer and ECC-1 endometrial cancer cells to proliferate in response to estradiol, tamoxifen, or serum was determined (Figure 2). Reduction of cyclin D1 expression reduced cell cycle progression in both cell types, nearly abrogating it in ECC-1 cells. Knockdown of c-myb in MCF-7 cells reduced cell cycle progression, but surprisingly knockdown of c-myc had little effect. As expected, MCF-7 cells did not proliferate following tamoxifen treatment. Reduction of either c-myc or c-myb levels caused a ~60% decrease in cell proliferation in ECC-1 cells. This may be a secondary effect, since both treatments also led to a reduction in cyclin D1 expression. The ECC-1 cells did not proliferate in response to tamoxifen, and this experiment will be repeated in a tamoxifen-responsive endometrial cell line. In both cell types, responses to complete serum were similar to those with estradiol alone, suggesting that estrogens are the critical mitogens in serum.

KEY RESEARCH ACCOMPLISHMENTS

- $CHIP^2$ has been used to identify ER binding sites on a chromosome-wide scale
- Novel ER binding sites distant from genes have been shown to recruit cofactors and stimulate transcription
- Only a small fraction (~1.5%) of the available estrogen response element sequences in the genome are bound by ER in response to estradiol or tamoxifen treatment
- Reducing expression of cyclin D1 or c-myb reduces estrogen-induced proliferation in both breast and endometrial cells, while reduction of c-myc reduces proliferation specifically in endometrial cells

REPORTABLE OUTCOMES

(The first four outcomes are new; the others represent the entire period of the award)

A manuscript showing the ER binding sites across two entire chromosomes has been submitted for publication.

A manuscript detailing the identification of tissue-specific ER targets involved in proliferation has been revised to include the proliferation data and will be submitted for publication.

Dr. Hestermann successfully applied for NIH BRIN funding in his new position, based in part on data collected under this award.

Dr. Hestermann has applied for an NIH R-15 award based in part on data collected under this award.

Two publications: 1. Shang *et al* 2000 *Cell* 103:843-852
2. Shang and Brown 2002 *Science* 295:2465-2468

Drs. Shang, Hestermann, and Brown have presented this work at several conferences, including the American Association for Cancer Research and the Nuclear Receptor and Enzymology of Chromatin Keystone meetings.

Based in large part on the success of this work, Dr. Shang was offered faculty positions at institutions both in the United States and abroad, and accepted a position at Beijing University.

Based in part on the success of this work, Dr. Hestermann was offered multiple faculty positions and is now a professor at Furman University.

CONCLUSIONS

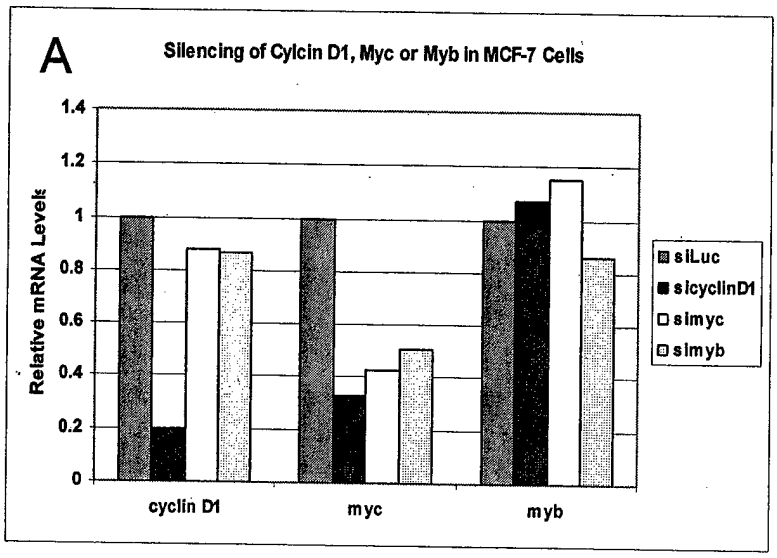
This project has produced a wealth of new information about regulation of gene expression by estrogen receptor in breast and endometrial cells. Identification and quantification of the binding of estrogen receptor and associated cofactors to promoters *in vivo* has provided novel insights into the mechanisms of gene regulation by ER. A key finding was that differential expression of the cofactor SRC-1 between breast and endometrium accounts for the action of tamoxifen in blocking proliferation in the former cells, while promoting proliferation in the latter. This finding should provide means for combating resistance to tamoxifen therapy in breast cancer treatment, as well as leading to mechanistic testing of novel therapeutic agents which are superior to tamoxifen for breast cancer prevention.

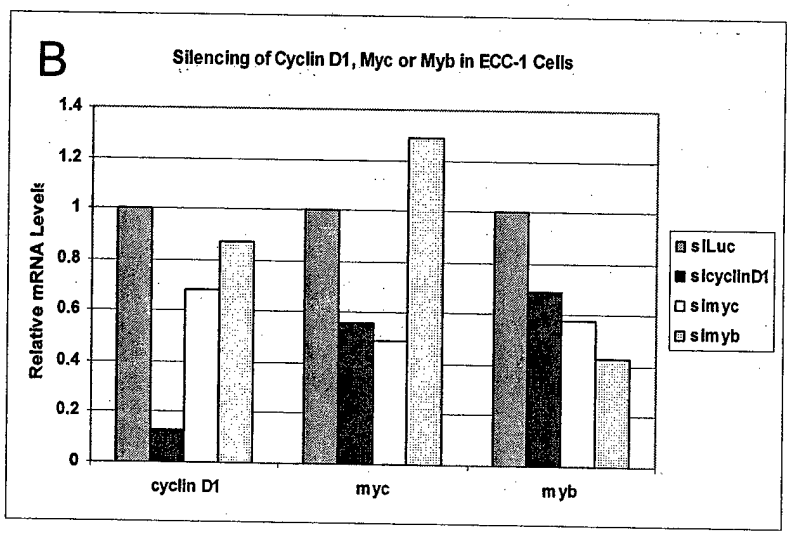
A combination of $CHIP^2$ and expression profiling data has identified novel targets of estrogen action and has also established the key targets of estrogen action that influence cell proliferation. $CHIP^2$ is also generating data on genome-wide promoter binding by ER and its associated cofactors. Using this technique and more traditional approaches, ER binding sites distant from genes have been shown to activate transcription. With specific targets of estrogen action identified and the mechanisms of their activation more clearly defined, targeted treatments of estrogen-dependent breast and endometrial cancer can be developed.

REFERENCES

Shang, Y., Xiao, H. DiRenzo, J., Lazar, M. A., and M. A. Brown. (2000) Cofactor dynamics and sufficiency in estrogen receptor-mediated transcription. *Cell* 103:843-852.

Shang, Y. and M. A. Brown. (2002) Molecular determinants for the tissue specificity of SERMs. *Science* 295:2465-2468.





C **Silencing of Myb and Cyclin D1 in MCF-7 and ECC-1 Cells**

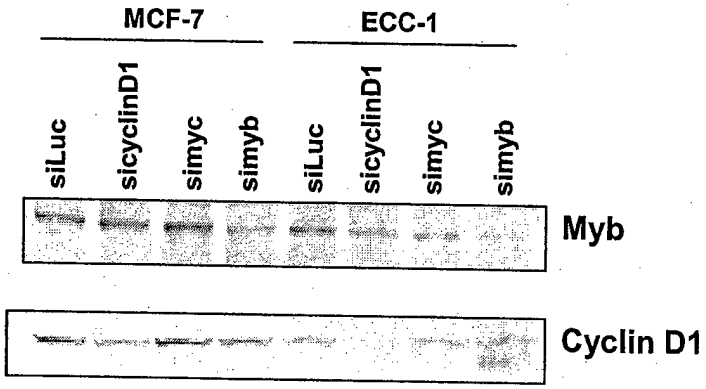


Figure 1. Expression of ER target genes in MCF-7 (A) or ECC-1 (B) cells following RNAi treatment. Cells were transfected for two days with oligos to knock down expression of luciferase (negative control) or the indicated ER targets. Expression of RNA was measured by qRT-PCR (A and B), and expression of protein was measured by immunoblot (C).

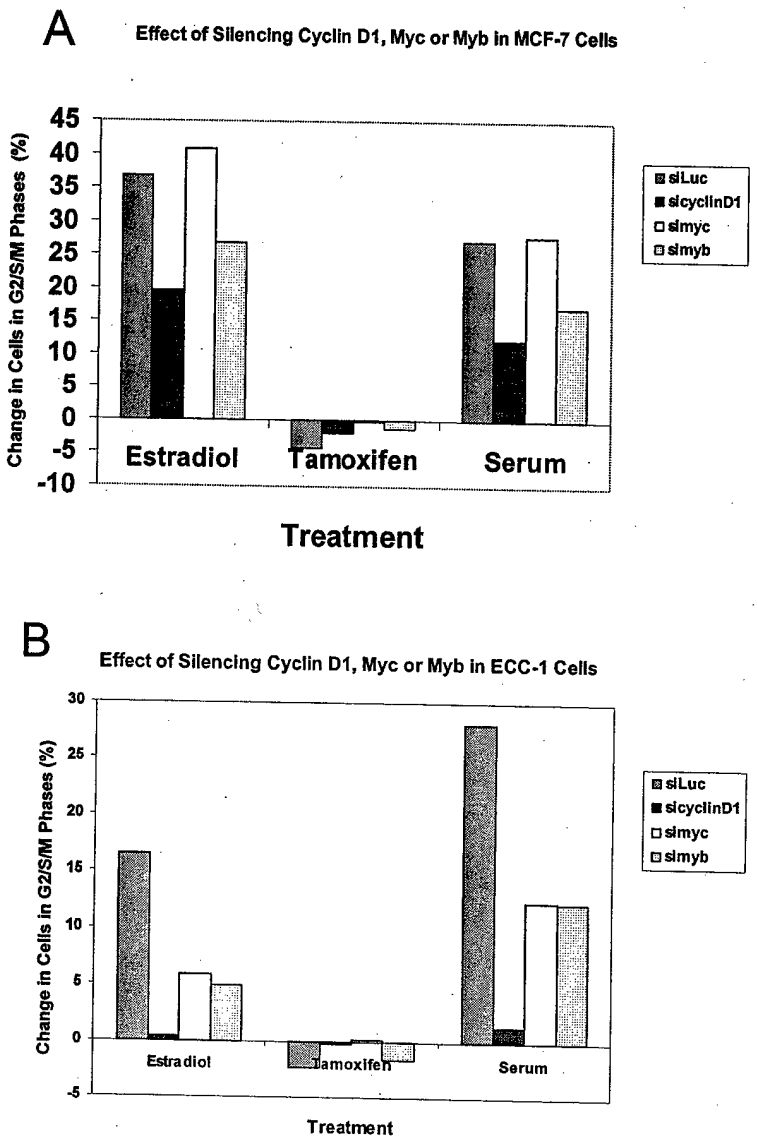


Figure 2. Cell cycle progression in MCF-7 (A) or ECC-1 (B) cells following RNAi treatment. Cells in hormone-depleted medium were transfected for two days with oligos to knock down expression of luciferase (negative control) or the indicated ER targets. Cells were then treated with 100 nM estradiol, 1 μM 4-hydroxytamoxifen, or 10% complete serum. After 24 hours, cells were fixed, stained with propidium iodide, and analyzed for cell cycle stage by flow cytometry. Change in fraction of cells in S or G2/M is relative to ethanol-treated controls.

Long-distance transcriptional regulation by the Estrogen Receptor

Jason. S. Carroll[1], Xiaole. S. Liu[2,4], Alex. S. Brodsky[3], Wenlin. Shao[1], Eli. V. Hestermann[5], Timothy. R. Geistlinger[1], Edward. A. Fox[3], Pamela. A. Silver[3] and Myles Brown[1*].

[1]Department of Medical Oncology, [2]Department of Biostatistics and [3]Department of Cancer Biology, Dana-Farber Cancer Institute, Harvard Medical School. 44 Binney St, Boston, MA, 02115. [4]Harvard School of Public Health. [5]Department of Biology, Furman University, 3300 Ponsett Hwy, Greenville, SC, 29613

*To whom correspondence should be addressed. E-mail: myles_brown@dfci.harvard.edu

Knowledge of the transcriptional regulation of estrogen target genes has been limited to promoter motif mapping and *in vitro* transcriptional data [1-3]. Here we have explored the *in vivo* association of the Estrogen Receptor with all of the non-repetitive sequence of chromosomes 21 and 22 by combining chromatin immunoprecipitations (ChIP) with tiled microarrays. 32 clusters of ER binding sites were observed across the 35 million base pairs represented including sites more than 100 kb distal to known estrogen regulated genes. A number of binding sites were further validated by ChIP and PCR. Additionally, RNA polymerase II and the oncogenic cofactor AIB-1, were shown to be recruited to all these sites in an estrogen dependent manner, thus supporting their role as *bone fide* transcription regulatory domains. Furthermore, many of the assessed binding sites contained intrinsic transcriptional activity in a luciferase reporter assay suggesting they are sites of localised transcription. A canonical ERE occurred in most binding sites, but only 1.5% of predicted EREs recruited ER, confirming a higher level of complexity with regards to *in vivo* ER-chromatin interactions and transcriptional regulation of target genes.

Recent clarification of estrogen action has highlighted the increasing complexity of estrogen-mediated transcriptional regulation of target genes. A cyclic association of the Estrogen Receptor (ER), p160 cofactors (such as AIB-1), Histone Acetyl Transferases (HAT) and chromatin modifying molecules, such as p300/CBP and p/CAF, with the promoter of the Cathepsin D and pS2/TFF-1 genes are seen in a temporal and ordered fashion[4,5]. Estrogen Response Elements (ERE) have been identified within the 1 kb 5'

proximal region of the estrogen-regulated genes, TFF-1, EBAG9 and Cathepsin D genes[6] and putative Sp-1 and AP-1 regulatory regions within the promoters of the c-Myc and IGF-I genes[2,3]. A wealth of studies on β-globin gene regulation has contributed to our understanding of general mechanisms of transcriptional regulation and has shown that Locus Control Regions (LCR) up to 25 kb from the gene, are capable of enhancing gene transcription (recently reviewed in[7]). However, the role and identification of long distance regulatory regions in estrogen-mediated transcription has not yet been addressed. Such a study is fundamental to our understanding of estrogen action, an important issue considering the list of estrogen target genes known to regulate cell proliferation[8,9] and the importance of ER status in breast cancer development and progression[10]. This investigation combined chromatin immunoprecipitation (ChIP) analyses of *in vivo* protein-DNA complexes with novel 35 bp tiled microarrays that cover all the non-repetitive sequence within chromosomes 21 and 22[11,12]. The 780 characterized or predicted genes on chromosomes 21 and 22[11] represent about 2% of the total 30,000-35,000 genes in the genome and provide a representative model for identifying paradigms of gene regulation by the estrogen pathway. Such a combination allows for an unbiased chromosome-wide analysis of *in vivo* association of ER with the genes and intergenic sequence represented by the million probes without limiting the investigation to gene proximal regions.

Estrogen-dependent MCF-7 cells were deprived of hormones and stimulated with estrogen or vehicle for 45 minutes, a time period known to result in maximal recruitment of ER to the promoters of several known gene targets, including Cathepsin D and TFF-1,

which is located on chromosome 21[4]. ER associated DNA was amplified under non-biased conditions, labelled and hybridized to the microarrays. Positive probe signals that were enriched by ER ChIP relative to input (total genomic DNA) were identified by subtractive modeling. Identification of clusters of enriched DNA sequences after replicate ER ChIP revealed a total of 32 discrete clusters of binding sites (Fig 1a, 1b, table 1 and supplemental data 1). Localisation of ER to 400 bp within the proximal promoter of the estrogen-regulated gene, TFF-1, was observed, as was association with a region 10.5 kb upstream from the transcription initiation site (Fig 1a). The ERE had been previously mapped to a region 393 to 405 bp upstream from the transcription start site of TFF-1[1] and functioned as a positive control for the ChIP-microarray assay. Interestingly, a putative DNase hypersensitivity site had been mapped 10.5 kb upstream from the TFF-1 gene[13], precisely where we observed ER association, and supports the *in vivo* association of ER with the two discrete sites upstream from TFF-1.

Within the small list of 32 ER-binding clusters, we observed interaction with regions upstream from the gene encoding the XBP-1 transcription factor and within the DSCAM-1 gene. In direct support of the ER-binding data, XBP-1 has been implicated as an estrogen-regulated gene and DSCAM-1 has been upregulated in microarray analyses after estrogen stimulation[14,15]. Interestingly, neither gene recruited ER to the 'classic' 5' promoter region, but possessed divergent patterns of association. The XBP-1 gene, recruited estrogen-ER to three distinct and discrete regions 13.2 kb to 22.9 kb from the transcription start site (Fig 1b). DSCAM-1 showed a different profile of ER-association, namely a clustering of four to five ER-association sites between exons 6 and 9, more than

half a megabase from the transcription initiation site. All of the ER-binding sites adjacent to TFF-1, XBP-1 and DSCAM-1 were validated by ER ChIP and standard PCR (Fig 2a-c). Also, quantitative PCR was performed on each of the sites after ER ChIP (Fig 2f) confirming these *in vivo* binding sites as genuine ER-association domains. PCR of the XBP-1 promoter and an intergenic region between the TFF-1 promoter and novel putative enhancer were used as negative controls (primer sequences can be found in supplemental material 2). RNA from vehicle or estrogen-stimulated MCF-7 cells was converted into double stranded cDNA and hybridised to the chromosome 21 and 22 tiled microarrays to identify the repertoire of estrogen-regulated genes and their locations relative to the small pool of ER-binding sites on these two chromosomes. A number of genes were transcriptionally active at the 3 hr time point, but only 3% of all the genes on these two chromosomes were upregulated by estrogen stimulation, including TFF-1 and XBP-1 (Supplementary data 1). This provides evidence that XBP-1 is a direct transcriptional target of the ER pathway. DSCAM-1 was not upregulated by estrogen stimulation at the 3 hr time point assessed, but was transcribed 7-fold after 24 hr of estrogen stimulation (data not shown). The immediate association with ER to the DSCAM-1 gene, but the long delay in transcriptional activation may be a consequence of a requirement for subsequent modification of the receptor complex or delayed pre-initiation complex (PIC) assembly. Despite the mechanism for the delay, it now appears that early and delayed estrogen-regulated genes recruit the receptor with the same kinetics and events after ER-binding are responsible for initiating transcription.

SOD-1, a gene observed to be estrogen-regulated (supplemental data 1) was greater than 130 kb from a single ER-binding site. This ER-binding site is further than previously conceived for regulatory regions. SOD-1 is a superoxide dismutase gene involved in scavenging oxygen free radicals[16,17] and implicated in tamoxifen-resistant progression in MCF-7 xenografts[18], but not previously implicated as an estrogen target gene. NRIP-1, another gene seen to be estrogen-regulated (supplemental data 1) is a transcriptional cofactor implicated in influencing ER activity[19,20]. No proposed method of NRIP-1 transcription by estrogen has been hypothesised. Five binding sites were identified 200-240 kb from the start of NRIP-1, in a region of chromosome 21 well known for its scarcity of genes[21]. Standard ChIP of three of the NRIP-1 sites and the SOD-1 site confirmed that ER associates with all of these distant domains in an estrogen-dependent manner (Fig 2d and 2e). 5'RACE was conducted on the NRIP-1 gene to identify if previously missed exons existed closer to the ER-binding sites than initially suggested. Surprisingly, sequencing of the 5' terminus of the NRIP-1 transcript after estrogen stimulation revealed the presence of two previously missed exons for NRIP-1, 74.96 kb and 97.39 kb from the previously annotated gene start site (data not shown). Therefore, the ER-binding sites are between 107 kb and 144 kb from the *bone fide* first exon of NRIP-1.

Functional investigation of the identified binding sites as essential transcriptional regulators is limited due to the large amount of sequence between the putative regulatory region and the gene. However, we explored the possibility that these ER-binding sites may recruit components indicative of transcriptional activation. RNA PolII ChIP

followed by real-time PCR was performed on all of the regulatory regions adjacent to TFF-1, XBP-1, DSCAM-1, NRIP-1 and SOD-1 genes. Interestingly, RNA PolII association was seen with all of these sites in an estrogen-dependent manner (Fig 2f). Furthermore ChIP of AIB-1, an oncogenic nuclear receptor cofactor[22] confirmed that AIB-1 is also present on all of these 'regulatory' sites following estrogen exposure (Fig 2f). This data provide evidence for these discrete ER-binding sites as functionally relevant estrogen regulatory domains, despite the fact that many exist further from the estrogen-regulated gene than was previously thought for enhancers.

Little is known of the chromatin binding sites utilised by tamoxifen to inhibit transcription of the ER pathway. Using the specific and discrete pool of *in vivo* estrogen-ER-binding sites identified near TFF-1, NRIP-1, XBP-1, SOD-1 and DSCAM-1 genes, we performed ER ChIP following tamoxifen treatment of MCF-7 breast cancer cells. A majority of the association sites employed by ER in estrogen-mediated transcription were also sites of ER association during transcriptional abrogation by tamoxifen (Fig 2b) despite the dichotomy in protein recruitment[4,23]. However, not all of the ER-binding domains are common between estrogen and tamoxifen, suggesting that tamoxifen-specific binding sites might also occur.

To distinguish if the distal enhancer regions and transcription initiation machinery are in physical contact, we performed ChIP of the TFIID (TBP) transcriptional initiation factor, a component known to promote PIC assembly prior to elongation [24,25]. Quantitative PCR confirmed that TFIID is recruited to the promoters of both TFF-1 and XBP-1 following

estrogen treatment, but not on the enhancers of these genes (Fig 3a and 3b). Deletion of enhancer regions from the mouse β–globin gene can dramatically decrease transcription of the globin gene, but the promoter can still recruit the transcription machinery[7] supporting the hypothesis that the PIC does not require enhancer regions for assembly on the promoter. Furthermore, TBP assembly on an Ig promoter was found to be independent of RNA PolII recruitment to enhancer regions[26], directly substantiating the conclusions here, that proteins associated with distal enhancers and promoter-associated proteins can assemble independently. Therefore enhancer-promoter interactions appear to function with similar characteristics in nuclear receptor mediated transcription.

Cawley *et. al.* proposed the theory that transcription factor binding sites correlate with regions of non-coding RNAs[12]. To explore the possibility that the regulatory regions we have identified may function as more than simple binding sites and may in fact be sites of discrete localised transcription, we cloned each of the binding sites (between 400 and 500 bp from each enhancer) adjacent to TFF-1, XBP-1, NRIP-1 and DSCAM-1 genes into a reporter vector containing the Renilla luciferase gene that lacked any eukaryotic promoters or enhancer elements and transfected these into MCF-7 cells which where subsequently treated with estrogen or vehicle control. The introduction of non-specific sequence from regions between the promoter and enhancer of TFF-1 and 40 kb upstream from the XBP-1 gene were used as negative controls. Unexpectedly, all enhancers, with the exception of NRIP-1 enhancer 2 and DSCAM-1 enhancer 3 were capable of driving luciferase expression (Fig 3d). 5'RACE of the Luciferase transcript driven by NRIP-1 enhancer 1 and DSCAM-1 enhancer 1 vectors confirmed that the transcription start site

of the luciferase transcript originated from within the enhancer itself (data not shown). Therefore, a number of the assessed regulatory regions are not simply ER-binding sites potentiating transcription of downstream genes, but can be sites of transcriptional initiation themselves, whereas other appear to function solely as receptor docking regions.

An algorithm designed to identify common motifs[27] within all the ER binding sites on chromosomes 21 and 22 revealed that only one domain was consistently represented. A canonical ERE[6] was present in 50% of all the ER binding sites on chromosomes 21 and 22 (Fig 3c). The likelihood of an ERE occurring in one of the 32 clusters of ER-binding sites was dramatically increased when compared to total chromosomal DNA ($p < 5.6 \times 10^{-9}$). A complete map of predicted EREs on chromosomes 21 and 22 (Fig 1a) emphasises that the presence of a canonical ERE does not represent a guaranteed ER binding site. Furthermore, ChIP of ER followed by PCR of regions flanking a number of perfect EREs on chromosome 21 and 22 did not reveal any ER association (data not shown). This, for the first time provides evidence that receptor binding motifs do not dictate receptor-DNA association and suggest that other levels of regulation occur, perhaps at a more complex, chromatin level. It also raises questions about the validity of motif screens as methods for identifying receptor interaction domains.

These ChIP-microarray experiments demonstrate the ability to identify genuine *in vivo* ER protein binding sites in previously unexplored intergenic sequence with complete confidence. ER binding was not a random occurrence, with only a limited number of

sites, many adjacent to genes that were shown to be estrogen-induced. Interaction of the ER with classic promoters was rarely seen and instead binding sites more than 100 kb from regulated genes was observed. Many of these binding domains were shown to potentiate transcription in a promoter-less context, but others appeared to function solely as sites of ER recruitment. Furthermore, a screen of EREs revealed limited usage of potential ER binding sites and the occurrence of other non-ERE sites as *in vivo* interaction regions. These data illuminate the complexity of ER-mediated transcriptional activation in an *in vivo* setting and provide a novel and informative understanding of chromatin-regulated transcription in general.

Materials and Methods

Chromatin Immunoprecipitation (ChIP)-microarray preparation

ChIPs were performed as previously described[4], with the following modifications. 2 μg of antibody was prebound for a minimum of 4 hr to protein A and protein G Dynal magnetic beads (Dynal Biotech, Norway) and washed three times with ice-cold PBS plus 5% BSA, and then added to the diluted chromatin and immunoprecipitated overnight. The magnetic bead-chromatin complexes were precipitated and washed 6 times in RIPA buffer (50 mM HEPES pH 7.6, 1 mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5 M LiCl). Elution of the DNA from the beads was as previously described[4]. Antibodies used were: ERα (Ab-10) from Neomarkers (Lab Vision, UK), ERα (HC-20), RNA PolII (H-224), AIB-1/RAC3 (C-20), TFIID (SI1), mouse IgG (sc-2025) and rabbit IgG (sc-2027) from Santa Cruz (Santa Cruz Biotechnologies, CA). Ligation-Mediated PCR was performed as previously described[28]. Following DNA amplification, DNA was subjected

to DNase I treatment (3 U/μ□ in NEBuffer 4) at 37°C for 30min, followed by denaturation at 95°C for 15min. 5 μl was run on an agarose gel to assess the efficiency of fragmentation. To the remaining 50 μl, 13 μl TdT buffer (Promega), 0.6 U TdT, and 1 μl Biotin-N6-ddATP (Perkin Elmer, MA) was added and the following program was run: 37°C 16hr, 95°C 10min, 4°C hold.

Data Analysis

1,054,325 probe pairs were mapped to chromosome 21 and 22 according to the NCBIv33 GTRANS Libraries provided by Affymetrix. (PM-MM) value was recorded for each probe pair, and a probe pair was removed if either PM or MM was flagged as outlier by the Affymetrix GCOS software. The five samples (three ER+ ChIP and three genomic inputs) were normalized by quantile normalization (Bolstad et al. 2003) based on a combined 76 ChIP experiments obtained from public domain and Dana-Farber Cancer Institute. The behavior of every probe pair i, assumed to be $N(\mu_i, \sigma_i^2)$, was estimated from the 76 normalized experiments. A two-state (ChIP-enriched state and non-enriched state) Hidden Markov Model with the following parameters was applied to each sample to estimate the probability of ChIP-enrichment at each probe pair location:

Transition probabilities: $\frac{300}{1{,}054{,}325}$ for transition to a different state

$1-\frac{300}{1{,}054{,}325}$ for staying in the same state

Emission probabilities: $N(\mu_i, \sigma_i^2)$ for non-enriched hidden state

$$N(\mu_i + 2\sigma_i, (1.5\sigma_i)^2) \text{ for enriched hidden state}$$

To combine the results from the six samples, an enrichment score was calculated as the average enrichment probability in the three ER+ ChIP samples subtracted by the average enrichment probability in the two genomic input samples. Since the tiling array has one 25-mer probe in every 35 bp of non-repeat regions, the coverage of every probe was extended by 10-bp on both ends. An enriched regions is defined as run of probes with enrichment score > 50% and covering at least 125 bp. Each enriched region can tolerate up to two neighboring probes with enrichment score between [10%, 50%]. If two neighboring probes are more than 210 bp apart, the enriched region is broken into two separate blocks. A summary enrichment score was obtain for each enriched region, which is the enrichment score summation for all the probes in the region divided by the square root of the number of probes in the region. This summary enrichment score represents the relative confidence of a predicted enriched region.

The genomic DNA of every ChIP-enriched region was retrieved from UCSC genome browser, and ranked by the summary enrichment score. MDscan algorithm (Liu et al. 2002) was applied to the sequences to find enriched sequence pattern that is the putative estrogen receptor binding motif. To find a motif of width w, MDscan first enumerates each w-mer in the highest ranking sequences, and collects other w-mers similar to it in these sequences to construct a candidate motif as a probability matrix. A semi-Bayes scoring function was used to remove low scoring candidate motifs, and refine the rest by checking all w-mers in all the ChIP-enriched sequences. A high scoring motif (with

similar consensus) consistently reported multiple times at different motif widths indicates a strong prediction. From all the ChIP-enriched sequences, MDscan discovered a strong 15-mer motif (Figure).

Real time PCR

Primers were selected using Primer Express (Applied Biosystems). 5 μl of precipitated and purified DNA was subjected to PCR using the Applied Biosystems SYBR Green Mastermix. Relative DNA quantities were measured using the PicoGreen system (Molecular Probes, OR). All primer sequences and locations are listed in S2.

Double stranded cDNA synthesis

Total RNA was converted to double stranded cDNA according to the Invitrogen Superscript Double stranded cDNA synthesis manufacturers instructions. The RNA was primed with 250 ng Oligo(dT) (Invitrogen) and 25 ng random hexamers (Gibco). cDNA was fragmented and labelled as described above.

Renilla luciferase transcriptional activity

Regulatory domains (approximately 400-500 bp) were amplified by PCR and cloned into the pRL-Null vector (Promega). Hormone depleted MCF-7 cells were transfected with each of the regulatory domains with Lipofectamine 2000 (Invitrogen) and total protein lysate was harvested after estrogen or ethanol addition for 24 hr. Renilla luciferase activity was assessed using the Renilla Luciferase Assay kit (Promega).

5'RACE

5' RACE was performed according to the manufacturers instructions (Invitrogen). The primers sequences used were: NRIP-1 RT primer (5'-TGCCTGATGCATTAGTAATCC-3'), NRIP-1 nested primer 1 (5'-GAGCCAAGCTCTTCTCCATGAGTCATGTTC-3') and NRIP-1 nested primer 2 (5'-ACCTTCCATCGCAATCAGAGAGAGACGTACTG-3'). The PCR product was cloned and sequenced by standard methods. The primers used to identify the 5' end of the Renilla luciferase transcripts were Renilla RT primer (5'-GCCTGATTTGCCCATACCAAT-3'), Renilla nested primer 1 (5'-CATGTCGCCATAAATAAGAAGAGGCCGCGT-3') and Renilla nested primer 2 (5'-CATCTGGCCCACCACTGCGGACCAGTTATCAT-3').

Acknowledgements

The authors would like to thank Dale Porter for helpful suggestions and Danielle Lynch and Mitchell Lazar for reviewing the manuscript. We thank Drs Cliff A. Meyer, Wei Li, Molin Wang, David Harrington, Cheng Li, and Jun S. Liu for their insights and advices on the tiling array data analysis. This work was supported by NIH grant xxxxxxxxx, by the DF/HCC Breast Cancer Spore Grant, and the U.S Army Medical Research and Materiel Command under an Army Breast Cancer Fellowship (DAMD17-01-1-0222) to E.H.

Figure legends

Fig. 1

Map of ER binding sites on chromosomes 21 and 22 after estrogen stimulation. The visual representation of positive probe signal clusters in the ER-ChIP for chromosomes 21 (a) and 22 (b) are shown. Genes (represented as blue bars) in the 5' to 3' orientation are shown above the chromosome and 3' to 5' below. Gene locations are based on the 2003 genome freeze in the UCSC browser using Genbank and RefSeq positions. Binding site probe clusters were identified when signal of PM>MM probes were adjacent to at least two neighbouring positive probes, no further than three probes from another positive signal (minimum run of 60 bp and maximum gap of 80 bp) and constituents of a larger cluster, and are represented as red bars. The positive probe signals shown were positive in at least two of three separate biological replicates. Predicted EREs are shown as black bars. An expanded view of the TFF-1 gene and adjacent binding sites is shown as signal difference between ER ChIP and Input DNA for both the estrogen and vehicle treated cells. The TFF-1 gene is shown in its genuine 3'-5' orientation **(b)** Expanded view of ER binding site clusters adjacent to the XBP-1 gene on chromosome 22. Positive probe signal from ER ChIP after estrogen or ethanol vehicle treatment are shown as a difference above Input DNA (total genomic DNA) binding signals. The gene is shown in its genuine 3'-5' orientation.

Fig. 2

Validation of the *in vivo* binding of the transcription complex to regulatory regions. ChIP of ER and standard PCR of the promoter, intergenic region and enhancer of TFF-1 **(a)**,

the promoter and three binding sites adjacent to XBP-1 **(b)**, four binding sites within DSCAM-1 **(c)**, the promoter and three binding sites upstream from NRIP-1 **(d)** and a single binding site distal from SOD-1 **(e)**. The lanes are vehicle (C), estrogen (E) and Input (I). The primers against the TFF-1 intergenic region, the XBP-1 promoter and the NRIP-1 promoter were used as ER ChIP negative controls. Primer sequences are included in S2. **(f)** ChIP of ER, RNA PolII, AIB-1 or IgG control and real time PCR of the enhancer regions. The data are estrogen-mediated fold enrichment of protein-DNA association compared to vehicle (ethanol) control and are the average of three separate replicates. The colour intensity reflects the fold change as described in the legend. ChIP of ER after tamoxifen treatment is also included and is shown as fold enrichment over vehicle control. TFF-1 non-specific and XBP-1 non-specific primers were included as negative controls.

Fig. 3

Identification of estrogen-TFIID binding, transcriptional activity of the regulatory domains and presence of common binding domains. ChIP recruitment of TFIID (TBP) to the regulatory regions and promoters of **(a)** TFF-1 and **(b)** XBP-1. **(c)** A motif screen of all the ER binding sites on chromosomes 21 and 22 revealed the presence of one common motif. **(d)** Transcriptional activity of all of the regulatory regions in a Renilla luciferase vector lacking eukaryotic promoter or enhancer sequences. The data are the difference in luciferase units between estrogen and vehicle stimulated cells and are the average of three separate experiments.

Table 1

Location of ER binding sites and the closest gene in both the 5'-3' and 3'-5' orientations. The distance between the beginning of the gene and the binding site is listed as is any information on the transcriptional status of the gene after estrogen stimulation for 3hr. The number of binding sites within each cluster is also given. The distance from the binding sites is from the furthest binding site within a cluster.

References and Notes

1. Berry, M., Nunez, A. M. & Chambon, P. Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. *Proc Natl Acad Sci U S A* **86**, 1218-22 (1989).

2. Dubik, D. & Shiu, R. P. Mechanism of estrogen activation of c-myc oncogene expression. *Oncogene* **7**, 1587-94 (1992).

3. Umayahara, Y. et al. Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer. *J Biol Chem* **269**, 16433-42 (1994).

4. Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843-52 (2000).

5. Metivier, R. et al. Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751-63 (2003).

6. Klinge, C. M. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res* **29**, 2905-19 (2001).

7. Bulger, M., Sawado, T., Schubeler, D. & Groudine, M. ChIPs of the beta-globin locus: unraveling gene regulation within an active domain. *Curr Opin Genet Dev* **12**, 170-7 (2002).

8. 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**, 4562-74 (2003).

9. Kian Tee, M. et al. Estradiol and selective estrogen receptor modulators differentially regulate target genes with estrogen receptors alpha and beta. *Mol Biol Cell* **15**, 1262-72 (2004).

10. Henderson, B. E., Ross, R. & Bernstein, L. Estrogens as a cause of human cancer: the Richard and Hinda Rosenthal Foundation award lecture. *Cancer Res* **48**, 246-53 (1988).

11. Kapranov, P. et al. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916-9 (2002).

12. 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).

13. Giamarchi, C. et al. Chromatin structure of the regulatory regions of pS2 and cathepsin D genes in hormone-dependent and -independent breast cancer cell lines. *Oncogene* **18**, 533-41 (1999).

14. Pedram, A., Razandi, M., Aitkenhead, M., Hughes, C. C. & Levin, E. R. Integration of the non-genomic and genomic actions of estrogen. Membrane-initiated signaling by steroid to transcription and cell biology. *J Biol Chem* **277**, 50768-75 (2002).

15. Wang, D. Y., Fulthorpe, R., Liss, S. N. & Edwards, E. A. Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. *Mol Endocrinol* **18**, 402-11 (2004).

16. Beckman, J. S., Carson, M., Smith, C. D. & Koppenol, W. H. ALS, SOD and peroxynitrite. *Nature* **364**, 584 (1993).

17. Singh, R. J. et al. Reexamination of the mechanism of hydroxyl radical adducts formed from the reaction between familial amyotrophic lateral sclerosis-associated Cu,Zn superoxide dismutase mutants and H2O2. *Proc Natl Acad Sci U S A* **95**, 6675-80 (1998).

18. Schiff, R. et al. Oxidative stress and AP-1 activity in tamoxifen-resistant breast tumors in vivo. *J Natl Cancer Inst* **92**, 1926-34 (2000).

19. Cavailles, V. et al. Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. *Embo J* **14**, 3741-51 (1995).

20. Soulez, M. & Parker, M. G. Identification of novel oestrogen receptor target genes in human ZR75-1 breast cancer cells by expression profiling. *J Mol Endocrinol* **27**, 259-74 (2001).

21. Katsanis, N., Ives, J. H., Groet, J., Nizetic, D. & Fisher, E. M. Localisation of receptor interacting protein 140 (RIP140) within 100 kb of D21S13 on 21q11, a gene-poor region of the human genome. *Hum Genet* **102**, 221-3 (1998).

22. Kuang, S. Q. et al. AIB1/SRC-3 deficiency affects insulin-like growth factor I signaling pathway and suppresses v-Ha-ras-induced breast cancer initiation and progression in mice. *Cancer Res* **64**, 1875-85 (2004).

23. Lavinsky, R. M. et al. Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. *Proc Natl Acad Sci U S A* **95**, 2920-5 (1998).

24. Maldonado, E., Ha, I., Cortes, P., Weis, L. & Reinberg, D. Factors involved in specific transcription by mammalian RNA polymerase II: role of transcription factors IIA, IID, and IIB during formation of a transcription-competent complex. *Mol Cell Biol* **10**, 6335-47 (1990).

25. Takada, R. et al. Identification of human TFIID components and direct interaction between a 250-kDa polypeptide and the TATA box-binding protein (TFIID tau). *Proc Natl Acad Sci U S A* **89**, 11809-13 (1992).

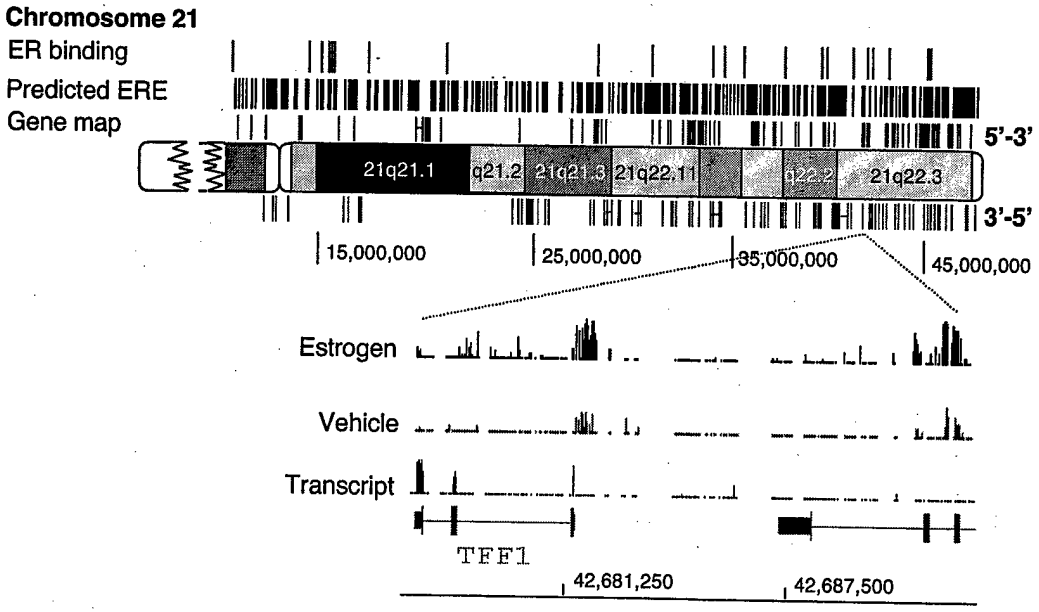
26. Bertolino, E. & Singh, H. POU/TBP cooperativity: a mechanism for enhancer action from a distance. *Mol Cell* **10**, 397-407 (2002).

27. 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**, 835-9 (2002).

28. Ren, B. et al. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev* **16**, 245-56 (2002).

Figure 1

A


Chromosome 21
ER binding
Predicted ERE
Gene map
5'-3'
21q21.1
q21.2
21q21.3
21q22.11
21q22.3
3'-5'
15,000,000
25,000,000
35,000,000
45,000,000
Estrogen
Vehicle
Transcript
TFF1
42,681,250
42,687,500

B

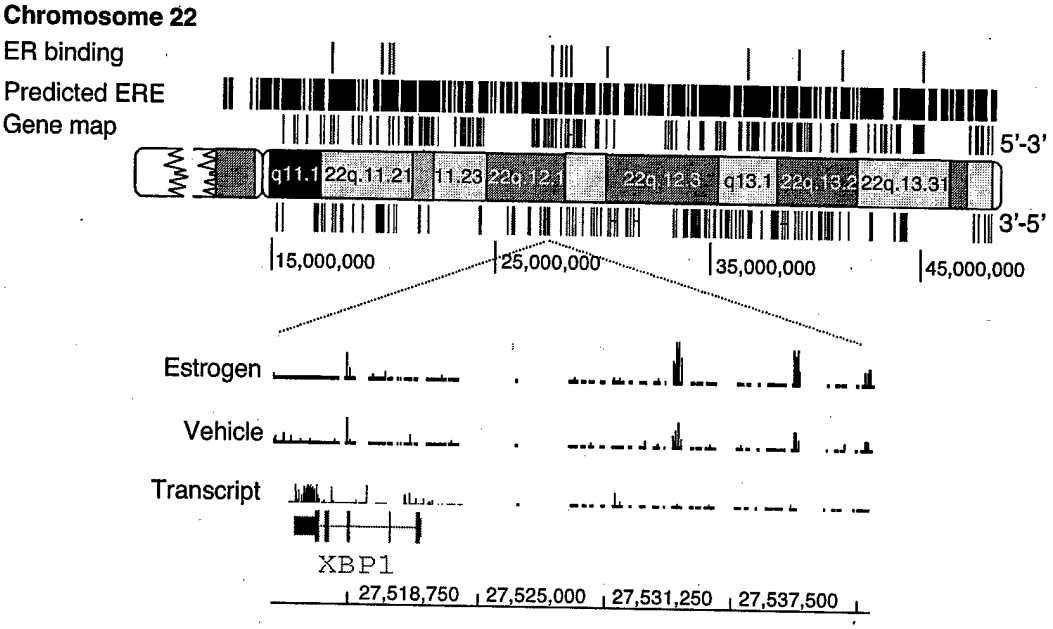
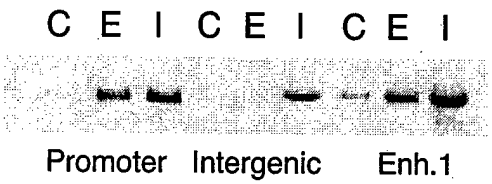
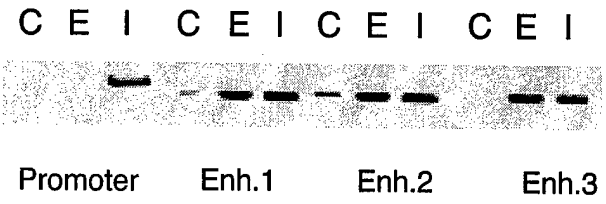

Chromosome 22
ER binding
Predicted ERE
Gene map
5'-3'
q11.1
22q.11.21
11.23
22q.12.1
22q.12.3
q13.1
22q.13.2
22q.13.31
3'-5'
15,000,000
25,000,000
35,000,000
45,000,000
Estrogen
Vehicle
Transcript
XBP1
27,518,750
27,525,000
27,531,250
27,537,500

Figure 2

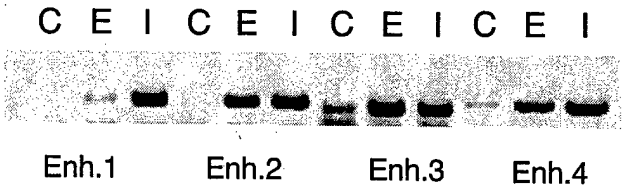
A



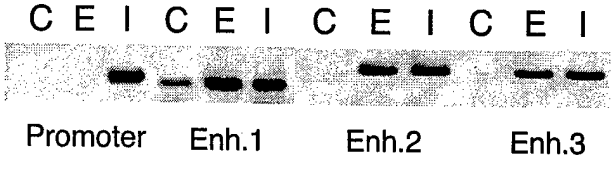
B



C



D



E

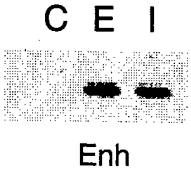


Figure 2f

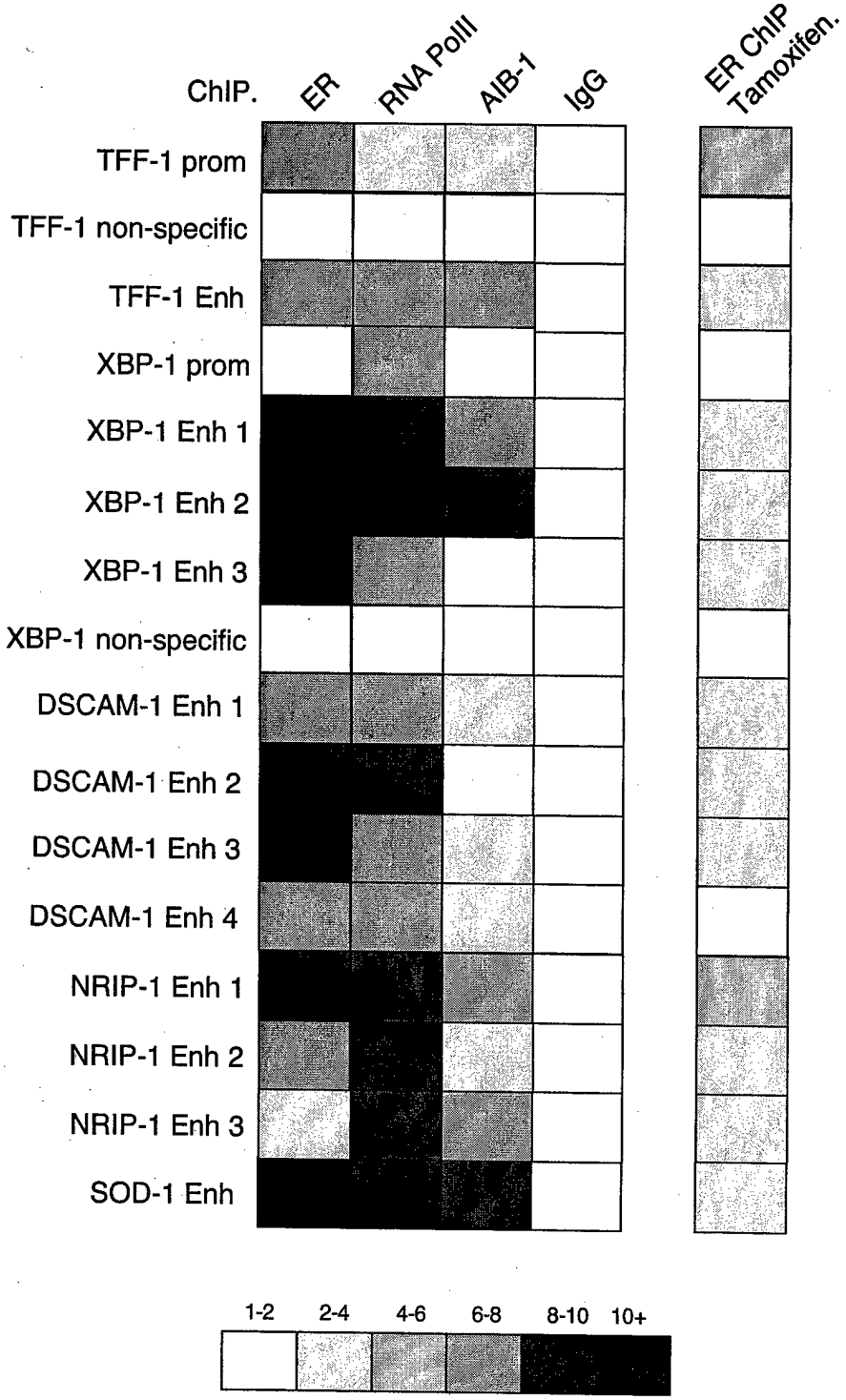

ChIP.
ER
RNA PolII
AIB-1
IgG
ER ChIP Tamoxifen.
TFF-1 prom
TFF-1 non-specific
TFF-1 Enh
XBP-1 prom
XBP-1 Enh 1
XBP-1 Enh 2
XBP-1 Enh 3
XBP-1 non-specific
DSCAM-1 Enh 1
DSCAM-1 Enh 2
DSCAM-1 Enh 3
DSCAM-1 Enh 4
NRIP-1 Enh 1
NRIP-1 Enh 2
NRIP-1 Enh 3
SOD-1 Enh
1-2
2-4
4-6
6-8
8-10
10+

Figure 3

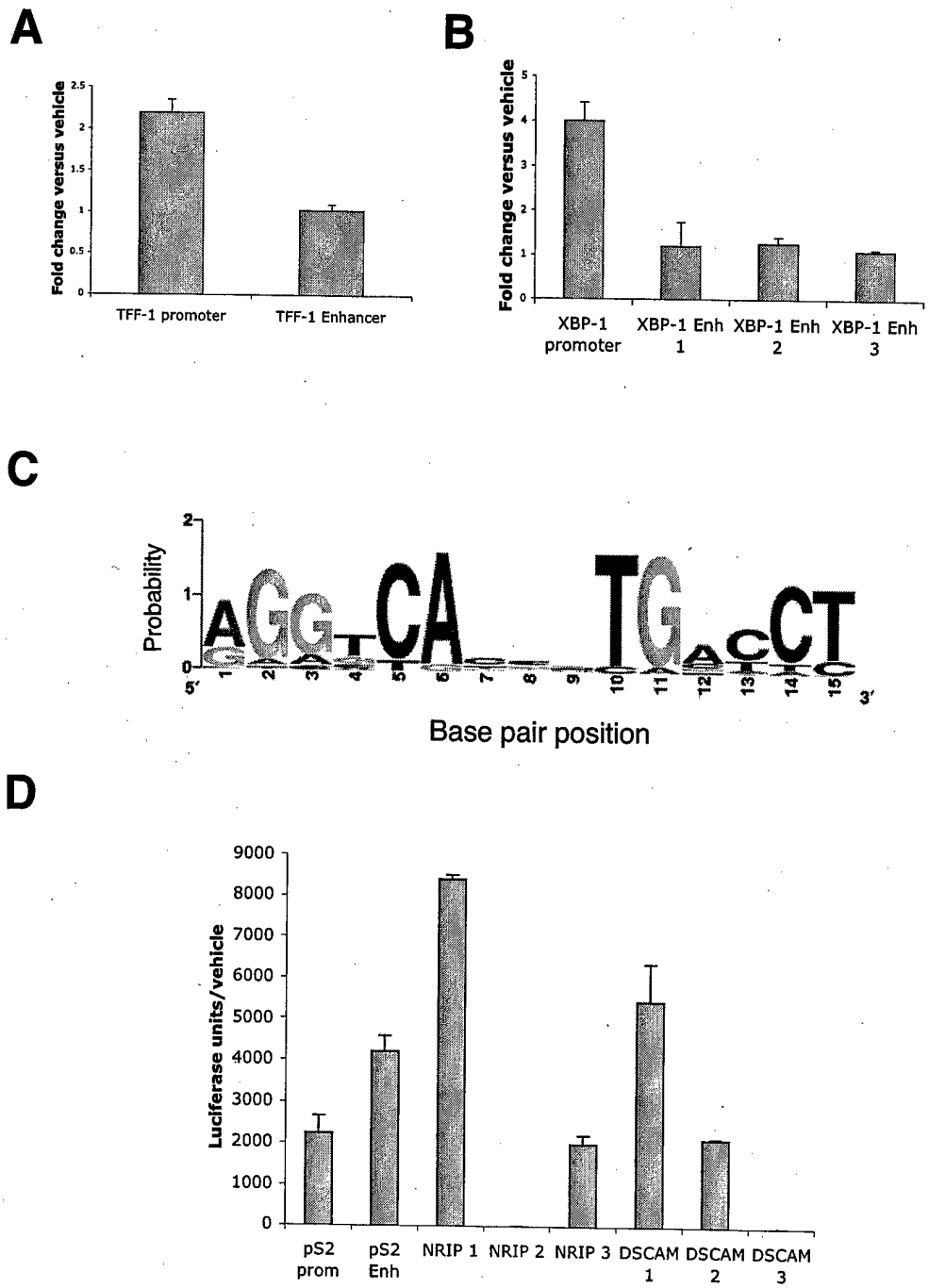

A
Fold change versus vehicle
2.5
2
1.5
1
0.5
0
TFF-1 promoter
TFF-1 Enhancer
B
Fold change versus vehicle
5
4
3
2
1
0
XBP-1 promoter
XBP-1 Enh 1
XBP-1 Enh 2
XBP-1 Enh 3
C
Probability
2
1
0
5'
3'
Base pair position
D
Luciferase units/vehicle
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
pS2 prom
pS2 Enh
NRIP 1
NRIP 2
NRIP 3
DSCAM 1
DSCAM 2
DSCAM 3

Table 1. Location of ER binding sites and the closest gene in both the 5'-3' and 3'-5' orientations. The distance between the beginning of the gene and the binding site is listed as is any information on the transcriptional status of the gene after estrogen stimulation for 3hr. The number of binding sites within each cluster is also given. The distance from the binding sites is from the farthest binding site within a cluster.

ER binding location	Orientation	Closest gene	Distance from binding site	Transciptional status	Number of binding sites
Chr. 21					
14600566	5'-3'	USP25	1.42Mb	No	1
	3'-5'	C21orf81	325.5Kb	No	1
15467330	5'-3'	USP25	500Kb	No	3
	3'-5'	NRIP-1	145Kb	Yes	3
21291444	5'-3'	NCAM2	1kb	Yes	1
	3'-5'	PRSS7	2.6Mb	No	1
31822242	5'-3'	SOD1	127kb	Yes	1
	3'-5'	TIAM-1	Intron	No	1
35027500	5'-3'	CBR-1	1.3Mb	No	1
	3'-5'	DSCR-1	120Kb	No	1
35510634	5'-3'	CBR-1	851Kb	No	1
	3'-5'	RUNX-1	170Kb	No	1
36481084	5'-3'	C21orf5	Intron	No	1
	3'-5'	C21orf18	128Kb	No	1
40607235	5'-3'	DSCAM-1	Intron	No	5
	3'-5'	C21orf88	770Kb	No	5
41911961	5'-3'	FLJ36335	520Kb	No	1
	3'-5'	TMPRSS2	112Kb	No	1
42681597	5'-3'	UBASH3A	37Kb	No	2
	3'-5'	TFF-1	10.5Kb	Yes	2
45663607	5'-3'	Col18A1	18Kb	No	1
	3'-5'	C21orf93	82Kb	No	1
Chr.22					
17192673	5'-3'	DGCR6	75Kb	No	1
	3'-5'	BID	561Kb	No	1
19824303	5'-3'	UBE2L3	422Kb	Yes	2
	3'-5'	SLC7A4	232Kb	No	2
27534672	5'-3'	KREMEN-1	259Kb	No	3
	3'-5'	XBP-1	24Kb	Yes	3
28238326	5'-3'	NF-2	87Kb	No	1
	3'-5'	C22orf19	Intron	No	1
41361730	5'-3'	BIK	376Kb	No	1
	3'-5'	A4GALT	14Kb	No	1
45100080	5'-3'	C22orf4	270Kb	No	1
	3'-5'	CELSR-1	Intron	No	1