Published in final edited form as:

Nature. 2009 November 5; 462(7269): 58–64. doi:10.1038/nature08497.

An Oestrogen Receptor α-bound Human Chromatin Interactome

Melissa J. Fullwood[1], **Mei Hui Liu**[1], **You Fu Pan**[1], **Jun Liu**[1], **Xu Han**[1], **Yusoff Bin Mohamed**[1], **Yuriy L. Orlov**[1], **Stoyan Velkov**[1], **Andrea Ho**[1], **Poh Huay Mei**[1], **Elaine G. Y. Chew**[1], **Phillips Yao Hui Huang**[1], **Willem-Jan Welboren**[2], **Yuyuan Han**[1], **Hong-Sain Ooi**[1], **Pramila N. Ariyaratne**[1], **Vinsensius B. Vega**[1], **Yanquan Luo**[1], **Peck Yean Tan**[1], **Pei Ye Choy**[1], **K. D. Senali Abayratna Wansa**[1], **Bing Zhao**[1], **Kar Sian Lim**[1], **Shi Chi Leow**[1], **Jit Sin Yow**[1], **Roy Joseph**[1], **Haixia Li**[1], **Kartiki V. Desai**[1], **Jane S. Thomsen**[1], **Yew Kok Lee**[1], **R. Krishna Murthy Karuturi**[1], **Thoreau Herve**[1], **Guillaume Bourque**[1], **Hendrik G. Stunnenberg**[2], **Xiaoan Ruan**[1], **Valere Cacheux-Rataboul**[1], **Wing-Kin Sung**[1,3], **Edison T. Liu**[1], **Chia-Lin Wei**[1], **Edwin Cheung**[1,4,5,*], and **Yijun Ruan**[1,4,*]

[1]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore [2]Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University, The Netherlands [3]Department of Computer Science, School of Computing, National University of Singapore, Singapore [4]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore [5]School of Biological Sciences, Nanyang Technological University, Singapore

Abstract

Genomes are organized into high-level 3-dimensional structures, and DNA elements separated by long genomic distances could functionally interact. Many transcription factors bind to regulatory DNA elements distant from gene promoters. While distal binding sites have been shown to regulate transcription by long-range chromatin interactions at a few loci, chromatin interactions and their impact on transcription regulation have not been investigated in a genome-wide manner. Therefore, we developed Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) for *de novo* detection of global chromatin interactions, and comprehensively mapped the

Correspondence and requests for materials should be addressed to Y.R. (ruanyj@gis.a-star.edu.sg) or E.C. (cheungcwe@gis.a-star.edu.sg).. * Corresponding authors: Yijun Ruan, PhD Genome Institute of Singapore, 60 Biopolis, Singapore 138672 T. 65 6478 8073; F. 65 6478 9059; ruanyj@gis.a-star.edu.sg Edwin Cheung, PhD Genome Institute of Singapore, 60 Biopolis, Singapore 138672 T. 65 6478 8184; F. 65 6478 9003; cheungcwe@gis.a-star.edu.sg.

Author contributions M.J.F. and Y.R. conceptualized the ChIA-PET strategy. M.J.F., E.C. and Y.R. designed the experiments. M.J.F., M.H.L., Y.F.P., J.L., A.H., P.H.M., E.G.Y.C., P.Y.Y.H., W-J.W., Y.H., Y.L., P.Y.T., P.Y.C., K.D.S.A.W., B.Z., K.S.L., S.C.L., J.S.Y., R.J., K.V.D., J.S.T., Y.K.L., T.H., H.G.S., X.R., and V.C-R. performed experiments. M.J.F., X.H., Y.B.M., Y.L.O, S.V., H-S.O., P.N.A., V.B.V., Y.K.L., R.K.M.K., G.B., H.G.S., X.R., V.C-R., W-K.S., C-L.W., E.C., and Y.R. analyzed experimental data. E.T.L., E.C., and C-L.W. commented on the manuscript drafts; M.J.F. and Y.R. wrote the paper.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A figure summarizing the main results of this paper is also included as SI (Figure 1). A ChIA-PET visualization browser is provided at http://cms1.gis.a-star.edu.sg (username is “guest”, password is “gisimsgtb”) to view the ERα ChIA-PET map.

Author information The raw sequences of the ChIA-PET libraries have been deposited with GEO and the NCBI Short Reads Archive.

The authors declare no competing financial interests.

chromatin interaction network bound by oestrogen receptor α (ERα) in the human genome. We found that most high-confidence remote ERα binding sites are anchored at gene promoters through long-range chromatin interactions, suggesting that ERα functions by extensive chromatin looping to bring genes together for coordinated transcriptional regulation. We propose that chromatin interactions constitute a primary mechanism for regulating transcription in mammalian genomes.

While genomic information is usually presented as a linear series of bases, genomes are known to be organized into three-dimensional structures *in vivo* through interactions with protein factors for nuclear process such as transcription1. The precise and coordinate regulation of transcription requires the binding of transcription factors to specific regulatory DNA sequences in the genome. ChIP-microarray2 (ChIP-Chip) and ChIP-sequencing3,4 (ChIP-PET and ChIP-Seq) have identified global transcription factor binding sites (TFBS) and revealed that many TFBS are far from gene promoters5. For example, the majority of TFBS bound by ERα in the human genome are distal to transcription start sites (TSS) of target genes6-10 . A major question arising from this observation is which distal TFBS are non-functional fortuitous binding sites, and which are involved in transcriptional activity through a remote control mechanism. Long-range chromatin interactions between DNA elements engaged in transcriptional regulation11,12 have been observed using Chromosome Conformation Capture (3C)13,14, and variants including ChIP-3C15,16, 4C17,18,19,20, 5C21, and 6C22, as well as RNA-TRAP23 and FISH24. However, these methods are limited to one-point or partial genome oriented detection, and are incapable of *de novo* detection of genome-wide chromatin interactions25.

To address whether and how DNA elements bound by protein factors interact through long-range chromatin looping in a genome-wide and unbiased manner, we conceived a novel strategy for Chromatin Interaction Analysis using Paired End Tag sequencing, called ChIA-PET. We applied ChIA-PET to characterize ERα-bound chromatin interactions in oestrogen-treated human breast adenocarcinoma cells (MCF-7), and generated the first human chromatin interactome map. Furthermore, using active promoter and transcriptional marks such as histone H3 lysine 4 trimethylation (H3K4me3) and RNA polymerase II (RNAPII) from ChIP-sequencing as well as gene expression microarray data, we show that ERα-bound chromatin interactions are functionally involved in regulating specific genes.

The ChIA-PET method

In ChIA-PET, long-range chromatin interactions are captured by formaldehyde cross-linking. Sonicated DNA-protein complexes are enriched by chromatin immunoprecipitation (ChIP). Tethered DNA fragments in each of the chromatin complexes are connected with DNA linkers via proximity ligation, and Paired-End Tags (PETs) are extracted for sequencing. The resulting ChIA-PET sequences are mapped to reference genomes to reveal relationships between remote chromosomal regions brought together in close spatial proximity by protein factors (Fig. 1a; Supplementary Fig. 1).

ChIA-PET proximity ligation generates two types of ligation products: self-ligation of the same DNA fragments and inter-ligation between different DNA fragments. PET sequences

derived from self-ligation products are mapped in the reference genome within a 3 Kb span, demarcating ChIP DNA fragments, similar to the standard ChIP-sequencing method3,8. Tethered DNA fragments in individual chromatin complexes can also ligate with each other, and the mapping results of such inter-ligation PET sequences would reveal if they are intrachromosomal (both tags of each PET are from the same chromosome) or interchromosomal (the tags are from different chromosomes). Singleton PETs are presumed experimental noise, and overlapping PET clusters are considered enriched putative binding sites or interaction events (Supplementary Fig. 2).

To test the ChIA-PET strategy, we constructed two ChIA-PET libraries from independent ERα ChIP-enriched oestrogen-treated MCF-7 chromatin preparations, and generated two replicate pilot datasets (IHM001H and IHM001N) using Roche/454 pyrosequencing. Our analysis showed that both ChIA-PET libraries produced comparable putative binding sites and interactions. To assess levels of false positive chromatin interactions, we created a negative control ChIP-PET library (IHM043) from the same ChIP sample, wherein the DNA was reverse cross-linked before proximity ligation. We also analyzed a previously reported cloning-based ChIP-PET library (SHC007)8. Both libraries generated abundant binding sites but no interactions. As an additional control, we used IgG, which binds to chromatin nonspecifically, to perform a mock ChIA-PET analysis (IHM062), and only a few binding sites and interactions were identified (Table 1; Supplementary Figs. 2-3; Supplementary Text I).

In proximity ligation-based analyses including 3C, the level of non-specific chimeric DNA ligations between different chromatin complexes can be high and thus may confound data analysis. To address this, we designed linker nucleotide barcodes in the ChIA-PET method to specifically identify such chimeric ligation PETs in another ERα ChIA-PET replicate. Linker barcoding analysis suggests that chimeric ligations are random and do not overlap with each other to form false positive interactions (Table 1, Supplementary Fig. 4, and Supplementary Text II). A possible complication is that ChIP-enriched loci with more DNA fragments would result in proportionally higher chances of inter-ligations, leading to false positive interactions comprising randomly overlapping inter-ligation PETs among highly-enriched ChIP DNA fragments. Hence, we devised a statistical scheme to calculate such probabilities and neutralize the potential ChIP-enrichment bias (Supplementary Materials and Methods; validations in Supplementary Fig. 5).

Together, these libraries indicate that the prevalent chromatin interactions (Supplementary Figs. 2d-g) identified by ERα ChIA-PET data depend on proximity ligations of chromatin complexes, and not technical artifacts of ligations between random DNA fragments, nor mapping errors.

ERα-bound chromatin interactome map

Next, we generated a large ERα ChIA-PET dataset (IHM001F) with 32 million PET sequences using Illumina GAII paired-end sequencing (Table 1; Supplementary Materials and Methods) for comprehensive analyses of ERα binding and chromatin interactions in oestrogen treated MCF-7 cells. Of 4.6 million uniquely mapped PET sequences, 1.2 million

(27%) were self-ligation PETs. Among the self-ligation PETs, 16.7% clustered to form overlapping PET groups, representing 14,468 putative ERαBS (FDR< 0.01, PET count per ERαBS ≥5, Supplementary Table 1). Of the inter-ligation PETs, 0.23 million (5.1% of uniquely aligned PETs) were intrachromosomal and 3.2 million (68%) were interchromosomal (Table 1). After statistical analyses wherein we discarded singleton inter-ligation PETs as either very weak interactions or background noise, clustered overlapping inter-ligation PETs, corrected for ChIP enrichment biases, and filtered out obviously false interactions due to structural variations in the MCF-7 genome (Supplementary Materials and Methods), we identified a large set of 1,451 intrachromosomal and a small set of 15 interchromosomal overlapping clusters consisting of 3 or more inter-ligation PETs per cluster (FDR < 0.05). These represent paired inter-ligating ChIP DNA fragments which indicate potential distant chromatin interactions bound by ERα (Supplementary Table 2).

Each chromatin interaction detected by an inter-ligation PET cluster features two anchor regions (interacting loci) and a loop (the intermediate genomic span between the two anchors), and is therefore called a “duplex interaction” (Supplementary Table 2). Most anchors (1,893/2,008= 94%) involve self-ligation PET-defined ERαBS (FDR< 0.01). Interestingly, many nearby duplex interactions are inter-connected, linking three or more anchors into “daisy-chain” aggregated complex interactions (Figs. 1b-d; Supplementary Fig. 6). For example, multiple duplex interactions with 3 ERαBSs in the *SIAH2* region interconnect to form a complex interaction. Hence, we further assembled 1,036 duplex interactions into 274 complex interactions based on overlapping of interaction anchors (Supplementary Materials and Methods). The remaining interactions (415) are stand-alone duplex interactions. Collectively, we identified 689 ERα-bound chromatin interaction regions (Supplementary Table 3).

To verify the ChIA-PET results, we validated a number of new ERαBS identified in this study by ChIP-qPCR (Supplementary Fig. 7), as well as putative intrachromosomal interaction sites (20 genomic loci) by 3C, ChIP-3C, 4C, and FISH experiments (three examples are shown in Fig. 1, others are in Supplementary Figs. 8-11; Supplementary Tables 4 and 5). Moreover, the 3C and FISH experiments showed higher levels of chromatin interactions in oestrogen-treated compared with untreated conditions, indicating that the interactions are oestrogen-dependent. We also examined 3 putative interchromosomal interactions by FISH; however, none of them were positive (Supplementary Table 4, Supplementary Text III), suggesting most ERα-bound intrachromosomal interactions are *bona fide*, while the putative interchromosomal interactions are false positives, or too weak to be validated.

Collectively, the ERαBS and chromatin interactions identified by ChIA-PET data constitute a whole genome chromatin interaction map bound by ERα. The genomic spans of most duplex interactions (86%) are in the range of <100 Kb, about 13% are from 100 Kb to 1 Mb, and < 1% are over 1 Mb. Complex interactions extend genomic span by connecting multiple duplex interactions. Many complex interactions (47%) have genomic spans in the range of 100 Kb to 1 Mb, with a few that are over 1 Mb (Supplementary Fig. 12; Supplementary Table 3).

To determine the reproducibility of this chromatin interactome map, we generated an additional ERα ChIA-PET library using a different antibody against ERα10. For this biological replicate (IHH015F), we obtained 20 million PET sequences (Table 1; Supplementary Materials and Methods). Overall, the two ERα ChIA-PET libraries are very similar with many overlapping ERαBS and intrachromosomal interactions but few interchromosomal interactions (Table 1; Supplementary Tables 1 and 2). The ERαBS identified in these two libraries showed high reproducibility, especially for highly enriched binding peaks. The 2,513 ERαBS with ≥50 PET counts per cluster (high-enrichment) overlapped with over 70% of the ERαBS in the replicate ChIA-PET library (Supplementary Table 6). Furthermore, these high-enrichment ERαBS intersect well with previously reported ERα binding maps9,10 (Fig.2a; Supplementary Fig. 13). Therefore, high-enrichment ERαBS are more reliable than low-enrichment sites. Many intrachromosomal interaction regions are detected in both replicate libraries. Highly abundant chromatin interactions are mostly reproducible. 86 of the top 100 most abundant chromatin interactions in IHM001F could be found in IHH015F (more analyses in Supplementary Table 7). Furthermore, all interactions previously identified and validated in this study are found in both replicate libraries (Supplementary Table 5). Conversely, none of the putative interchromosomal interactions were reproducible.

Together, our results demonstrate that the ChIA-PET method is highly reliable. Furthermore, our data suggests that ERα functions primarily via intrachromosomal mechanism. Thus, our subsequent analyses focused on intrachromosomal interactions. Downstream analyses for both ChIA-PET replicate libraries showed similar results; for simplicity, we discuss our results here using IHM001F, but results for IHH015F can be found in Supplementary Text IV.

We asked how many ERαBS are involved in complex- and duplex-interactions, or no-interactions (Fig. 2b-d). Our analysis showed that high-enrichment ERαBS are much more frequently involved in interactions (53%) compared to low-enrichment ERαBS (only 9%) (Fig. 2e; Supplementary Fig. 13), suggesting high-confidence and strong ERαBS are more likely to be involved in chromatin interactions than weaker ERαBS. To further understand ERαBS with respect to ERα target genes, we analyzed how many ERαBS are proximal or distal to gene promoters, based on a cut-off of 5 Kb from Transcription Start Sites (TSS) from the UCSC Gene database. Of 2,342 ERαBS involved in chromatin interactions, 387 (17%) are proximal and 1,955 (83%) are distal to TSS (Supplementary Fig. 14). We also observed the same ratio for no-interaction ERαBS: 2,043 (17%) are proximal and 10,175 (83%) are distal. Therefore, most ERαBS are distal to gene TSS, which is in agreement with previous studies7,8,10.

Chromatin interaction and transcriptional regulation

To investigate the functions of ERαBS and ERα-bound chromatin interactions in transcription activation, we generated genome-wide maps of H3K4me3 and RNAPII ChIP-Seq data from MCF-7 cells under oestrogen induction (Supplementary Materials and Methods). H3K4me3 is a histone modification which specifically marks active promoters26, and the presence of RNAPII is strong evidence for genes that are actively transcribing27.

We also analyzed previously reported FoxA1 ChIP-Chip data9, because FoxA1 is an important cofactor of ERα6,9. Generally, H3K4me3, RNAPII, and FoxA1 marks showed enrichment around ERαBS in our analyses (Fig. 3a). When we compared interaction ERαBS with no-interaction ERαBS, we found a significant enrichment gradient of RNAPII and FoxA1 binding around ERαBS: most association was with complex-interaction ERαBS, followed by duplex-interactions, and lastly no-interactions (Fig. 3a; Supplementary Fig. 15a; significance tests in Supplementary Text V).

Next, we examined the H3K4me3, RNAPII, and FoxA1 marks with respect to ERαBS proximal or distal to gene promoters and their involvement in chromatin interactions. Proximal ERαBS whether involved in interactions or not were highly enriched in H3K4me3, but this was not the case with distal ERαBS, which was expected since H3K4me3 is a known mark for promoter regions (Fig. 3b; Supplementary Fig. 15b; significance tests in Supplementary Text V). Proximal ERαBS were also highly enriched with RNAPII marks, but the enrichment for both proximal and distal ERαBS involved in interactions was significantly higher than that of the proximal and distal ERαBS that are not involved in interactions. Intriguingly, although RNAPII showed less enrichment around interaction distal ERαBS compared to interaction proximal ERαBS, the enrichment was significantly higher than that with stand-alone no-interaction distal ERαBS. Conversely, FoxA1 binding was more enriched around distal ERαBS than proximal ERαBS, and most enriched around distal interaction ERαBS (Supplementary Fig. 15c), and differences were statistically significant (significance tests in Supplementary Text V). This indicates that RNAPII and FoxA1, but not H3K4me3, predict interactions at distal ERαBS, and suggests that RNAPII and FoxA1 participate in tethering chromatin interactions. While RNAPII is strongly associated with ERαBS for transcription activation, FoxA1 is more directly correlated with ERα's regulatory function at distal ERαBS. At least 6 interacting ERαBS bracket the *FoxA1* gene, signifying ERα-mediated chromatin interactions may regulate *FoxA1* (Fig. 2b), further supporting the hypothesis that FoxA1 and ERα may regulate each other28.

Subsequently, we examined the 689 ERα-bound chromatin interaction regions with respect to looping structure and gene transcription. We envisage that multiple ERαBS may function as "anchor" regions forming chromatin looping structures in 3-dimensional space (Fig. 4a). Genes close to interaction anchors are considered as "anchor genes", and genes in the interaction loop regions and faraway from anchors are "loop genes". We annotated the interaction regions in relation to UCSC Gene database transcripts29 (a gene may have multiple transcripts; here we report transcript numbers, but gene numbers can be found in Supplementary Text VI). A gene was considered associated with a chromatin interaction region if the TSS of a gene is within 20 Kb of the interaction boundaries (Supplementary Fig. 14), a parameter that includes many known and validated ERα target genes. Most interaction regions (393/689=57%) were associated with "anchor genes" (TSS to interaction anchor within 20 Kb). Altogether, 1,575 "anchor genes" and 3,767 "loop genes" (TSS >20 Kb away from interaction anchors) were assigned to interaction regions (Supplementary Tables 3 and 8). Using the same distance parameter (20 Kb), we assigned 11,790 genes to 12,126 stand-alone ERαBS not involved in interactions (Supplementary Text VI).

Within interaction regions with at least one anchor gene, there are 1,073 distal ERαBS and 387 proximal ERαBS (< 5 Kb to TSS); and all distal ERαBS (5′ or 3′ to the gene promoter) are looped to anchor genes through connections with proximal ERαBS. Many interaction regions include multiple genes, such as the keratin gene cluster (Fig. 1c) and *NR2F2* locus (Fig. 1d), whereas others include only single genes, such as *SIAH2* (Fig. 1b). Distal ERαBS are stronger than proximal ERαBS, which is the inverse of RNAPII marks that are stronger at gene promoters than distal regions (Supplementary Fig. 16; examples in Fig. 1 and Supplementary Fig. 17). These observations suggest that direct ERα binding might be primarily initiated at one or multiple distal sites which then subsequently recruit other binding sites as anchors to form an interaction complex to ultimately engage the transcriptional machinery at gene promoters.

In addition, we also found 296 interaction regions with no associated anchor genes. While 41 regions contain loop genes, the remaining 255 have no associated UCSC genes assigned to them. Although some interaction regions could be noise or non-functional, some interactions are near gene promoters just outside the 20 Kb cutoff, and further sequencing might extend the interaction data to the promoters. The presence of H3K4me3, RNAPII marks, and RT-qPCR data at the interaction anchor sites suggests that some interactions could be involved in regulating yet-to-be identified transcripts, such as computationally predicted genes and non-coding RNA species (Supplementary Fig. 18). Alternatively, such interactions could be associated with maintaining chromatin structures or other unknown functions.

To understand if genes associated with ERα-bound interactions are oestrogen-regulated, we analyzed expression profiles of several interaction-associated genes by RT-qPCR over a time course of oestrogen induction (Supplementary Materials and Methods). All anchor genes examined are up-regulated by oestrogen induction (Supplementary Fig. 8). We extended our analysis to all interaction-associated genes using whole genome gene expression microarrays (Fig. 4b). Most “anchor genes” are up-regulated (60%), particularly at early time points, as compared with “loop genes” (48%), indicating that “anchor genes” are significantly associated with gene up-regulation (2-tailed p-value = 1.25e-16; Fig. 4C; Supplementary Text VII; Supplementary Table 9; Supplementary Fig. 19). Also, RNAPII marks are more associated with “anchor genes” (39%) than “loop genes” (26%) (2-tailed p-value = 1.00e-19). Conversely, genes assigned to ERαBS not involved in interactions (on the basis that the gene promoters are within 20 Kb to no-interaction ERαBS) have very similar expression profiles to the background control (all UCSC genes not associated with interactions), indicating genes associated with no-interaction ERαBS are less activated compared with genes associated with interaction ERαBS (significance tests in Supplementary Text VII). Hence, some stand-alone ERαBS could be noise, while others could involve non-looping mechanisms such as the recruitment of secondary coactivators for downstream functions6.

Intriguingly, within the anchor gene category, many (495 of 1,575 =31%) gene entries have 5′ and 3′ ends within interaction boundaries. Such entries, called “enclosed anchor genes”, frequently occupy the entirety of short interaction loops, engage multiple anchor sites around or within the gene, tend to have intense RNAPII marks covering the entire gene

(examples in Fig. 2b and c and Supplementary Fig. 20), and are preferentially associated with RNAPII marks and gene up-regulation as indicated by expression microarrays (Supplementary Text VII; Supplementary Table 9).

Collectively, our data shows an association between chromatin interactions and gene transcriptional activation: enclosed anchor genes are closely correlated with up-regulation as measured by gene expression microarray data and RNAPII ChIP-Seq peaks, followed by non-enclosed anchor genes, less for loop genes, and much less for genes not associated with interactions. These results suggest that gene-centric interaction structures may enclose a compartment for concentrating ERα and transcription-related proteins at target genes.

ERα-bound interactions may coordinate transcriptional regulation for multiple genes involved in the same functional pathways. At the keratin gene cluster interaction loci (Fig. 1c), enclosed anchor genes such as *KRT8* and *KRT18* are actively transcribing as evidenced by RNAPII and H3K4me3 marks, while the loop genes such as *KRT72* and *KRT75*, which are mainly keratins expressed in hair cells that do not play a role in mammary epithelial cells such as MCF-7, are mostly inactive (Supplementary Text VIII). Another example is the complex interaction that encompasses these 3 genes, *FOS*, *JDP2*, and *BATF* (Fig. 4c), which encode the dimerization partners of *JUN* to form the AP-1 transcription factors. AP-1 is important in regulating oestrogen receptor dependent transcription by functioning either as a DNA tethering partner or as an ERα co-factor30. In this complex interaction, *FOS* and *BATF* are enclosed anchor genes, and up-regulated as shown by RNAPII marks and RT-qPCR; whereas *JDP2* is a loop gene and down-regulated as shown by RT-qPCR and reduced RNAPII occupancy. Interestingly, the promoter of *JDP2* is marked by H3K4me3, a common feature found in many loop genes (Supplementary Table 9). *JDP2* and other loop genes could be “poised” for activation if they escape from the interaction loop. Hence, long-range transcriptional regulation by ERα may be a fine-tuning mechanism that evolved to differentially regulate specific sets of related genes.

To functionally determine whether some ERα-associated interaction regions are dependent on ERα, we used siRNA to knock down the level of ERα protein in MCF-7 cells (Supplementary Materials and Methods) and then measured if the interactions and gene transcription were affected. ERα-specific siRNA (siERα) efficiently reduced the amount of ERα protein, and effectively abolished the interactions as demonstrated by a set of 3C assays at the *GREB1* locus (Fig. 5). Furthermore, siERα blocked *GREB1* transcription as determined by RT-qPCR. Similar results were also shown at the *TFF1* site previously31. Together, these data suggest that at least some of the regulatory long-range chromatin interactions identified by ERα ChIA-PET data are mediated by ERα.

Discussion

We demonstrated the ChIA-PET mapping strategy is an unbiased, whole genome approach for *de novo* analysis of chromatin interactions, and hence constitutes a major technological advance in our ability to study higher-order organization of chromosomal structures and functions. The ChIA-PET interaction data greatly increase the accuracy of assigning distal TFBS to target genes, and globally addresses the 3-dimensional chromatin interaction

mechanism by which distal TFBS regulate transcription. We postulate the following primary mechanism for ERα function: ERα protein dimers are recruited to multiple and primarily distal ERαBS, which interact with one another and possibly with other factors such as FoxA1 and RNAPII to form chromatin looping structures around target genes; such topological architectures may partition individual genes into sub-compartments of nuclear space such as interaction anchor-associated genes and interaction loop-associated genes for differential transcriptional activation or repression. We further speculate that tightly enclosed chromatin interaction centers could help achieve and maintain high local concentration of transcription components for efficient cycling of transcriptional machinery on target gene templates (Summary of results in Supplementary Information; more discussion in Supplementary Text IX).

We anticipate that this first-ever global chromatin interactome map and the ChIA-PET assay will constitute a valuable starting point for future studies into the 3-dimensional architecture of transcription biology in whole genome contexts.

Methods Summary

MCF-7 cells grown in hormone-depleted medium were treated with 17 beta-estradiol ("oestrogen", E2) for 45 minutes before cross-linking with 1% formaldehyde for 10 min. ChIA-PET libraries were constructed by first performing ChIP using HC-20 antibody (Santa Cruz) or Mab-NRF3A6-050 antibody (Diagenode)10 against ERα. DNA fragments in ChIP complexes were then ligated to biotinylated half-linkers (linker ligation) containing flanking MmeI restriction sites. The complexes were further ligated under dilute conditions (proximity ligation). Paired-End Tags (PETs) were extracted from the ligation products by MmeI digestion. Released biotinylated PETs were purified by streptavidin-coated magnetic beads, ligated to adaptors, and PCR-amplified. Gel-purified amplicons of PET templates were sequenced by Roche/454 and/or Illumina paired-end sequencing. PET sequences were mapped to the human reference genome (hg18). Binding sites and interactions were identified using overlap PETs readout. To correct for ChIP enrichment bias, we formulated a statistical analysis framework to calculate the probability for the formation of inter-ligation PETs between two regions if ligations between DNA fragments occur by chance. Interactions were further collapsed into complex interactions if they shared interaction anchors. UCSC Genes were assigned to interaction regions if they were within 20 Kb of interaction regions. To functionally characterize ERα-bound interactions and associated genes, we conducted gene expression microarray experiments in time-course with and without E2 treatment, and generated genome-wide maps of H3K4me3 (ab8580, Abcam) and RNAPII (serine-5 phosphorylation antibody, ab5131, Abcam) ChIP-Seq data using Illumina single-read sequencing. Interaction-associated genes were annotated with expression microarray data and RNAPII and H3K4me3 ChIP-Seq peaks. Validation experiments included ChIP-qPCR, 3C, ChIP-3C, 4C, FISH, and RT-qPCR. For siRNA studies, ERα ON-TARGETplus SMARTpool siRNA (Dharmacon) was transfected into MCF-7 cells using Lipofectamine 2000 (Invitrogen). Sequences used in experiments are listed in Supplementary Table 10.

Methods

Further details of the methods are presented in the online Supplementary Information.

ChIA-PET library construction and sequencing

Cells were grown in hormone-free media for a minimum of 72 hours. Hormone-depleted cells were treated with 17 beta-estradiol ("oestrogen", E2, Sigma) at a final concentration of 100 nM for 45 min before the ChIP procedure. ChIP protocol was performed with ERα specific antibody (HC-20, Santa Cruz) as described previously8 or Mab-NRF3A6-050 antibody (Diagenode)10. The DNA fragments were end-repaired, followed by overnight ligation of biotinylated half-linkers that contain a flanking MmeI site. The linker added DNA fragments were then phosphorylated and followed by a second ligation reaction overnight under dilute conditions (< 0.2 ng DNA per ml reaction). Cross-links were reversed and the DNA fragments were purified. Nicks were repaired and the DNA was digested by MmeI for at least 2h at 37°C to release the tag-linker-tag PET structure. The biotinylated PETs were then immobilized on streptavidin-conjugated magnetic Dynabeads and the ends of each PET structure were then ligated to an adapters followed by PCR to amplify the PETs which were then sequenced by Roche 454 sequencing and Illumina Paired End Sequencing.

H3K4me3 ChIP-Seq data

H3K4me3 antibody (ab8580, Abcam) was used to generate ChIP-enriched DNA fragments for Illumina single read sequencing analysis. The H3K4me3 ChIP-Seq data was mapped to hg18 genome, and enrichment peaks for H3K4me3 binding were identified using ChIP-Seq peak calling algorithm as previously described32.

RNAPII ChIP-Seq data

RNA Polymerase II (RNAPII) serine 5 phosphorylation antibody (ab5131, Abcam) was used to generate RNAPII ChIP-enriched DNA fragments for Illumina single read sequencing analysis. The RNAPII ChIP-Seq data was mapped to the human genome (hg18) using Illumina's ELAND program, and the enrichment peaks RNAPII binding were identified using ChIP-Seq peak calling algorithm as previously described32.

Microarray gene expression data to identify oestrogen-regulated genes

A comprehensive dataset of time-course microarray experiments was performed to investigate the effects of oestrogen treatment on gene expression profiles and identify oestrogen responsive genes. Oestrogen treated (10 nM) and DMSO-mock MCF-7 cells (negative control) for 0, 3, 6, 9, 12, 24, and 48 hours were collected for RNA extraction and the labeled probes were hybridized to microarrays (HG-U133 Plus).

Circular Chromosome Conformation Capture (4C)

MCF-7 cells were harvested and cross-linked. Aliquots in SDS-containing buffer were diluted 10 times and Triton X-100 was added. End-blunting was performed. The chromatin samples were ligated in 10 ml at 16 °C overnight followed by reverse cross-linking and purification. The DNA samples were amplified using nested inverse PCR. Primers had to be

within 100 bp of the targeted ERα binding site peak and the resulting amplification product was run in a 6 % PAGE gel, and the fraction of the smear band above about 500 bp in size was excised. The DNA samples were sequenced using a 454 GSFLX long reads kit.

Fluorescence in-situ hybridization (FISH)

MCF-7 nuclei were harvested by treating cells with 0.75 M KCl for 20min at 37°C. The cells were fixed, and nuclei were dropped on slides for FISH. To prepare BAC probes, the BACs were grown; the DNA was harvested, and then labeled by nick translation. In the presence of 1 μg/μl of Cot1DNA, DNAs BAC clones were resuspended at a concentration of 5 ng/μl in hybridization buffer. Prior to hybridization, MCF-7 nuclei slides were treated and dehydratated through ethanol series (70%, 80% and 100%). Denaturated probes were applied to these pretreated slides and codenaturated at 75°C for 5 min and hybridized at 37°C overnight. Two posthybridization washes were performed. After blocking, the slides were revealed with avidin-conjugated fluorescein isothiocyanate (FITC) for biotinylated probes and anti-digoxigenin-Rhodamine for digoxigenin-labeled probes. After washing, slides were mounted and observed under an epifluorescence microscope. Between 100-200 interphase nuclei were analyzed for each mix of probes. Fusion and colocalization spots were counted in each nucleus.

Chromatin Immunoprecipitation Chromosome Conformation Capture (ChIP-3C)

ChIP-3C was performed as described previously14 with modifications. Chromatin immunoprecipitation was performed as described in the ChIP protocol. ChIP beads were washed and digested with restriction enzyme. Beads were washed, and ligation was performed in 100 μl at 16°C. Reverse cross-linking was performed. Primers and restriction enzymes for the ChIP-3C procedure were chosen based on ChIA-PET sequences. All primers and restriction enzymes had to be within a region of ±100-500 bp from the targeted ERα binding site peak.

Chromosome Conformation Capture (3C)

3C was performed as described previously14 with modifications. MCF-7 cells were harvested and cross-linked. Nuclei were resuspended in restriction enzyme buffer and treated with SDS then Triton-X. Samples were then digested with the selected restriction enzyme at 37°C overnight. SDS was used to stop the reactions, which was then sequestered with Triton-X. Ligation was performed at 16°C for 4h in 6.5 ml volume. Samples were reverse-crosslinked and purified. All primers had to be within a region of ±150 bp from the restriction enzyme digestion site.

RT-qPCR

Total RNA was prepared from MCF-7 cells induced with oestrogen for 0, 3, 6, 12 and 24 hours. Reverse transcription was performed with random primers. Real-time quantitative PCR was performed. The control primers used were against 36B4 (ribosomal protein mRNA).

siRNA knockdown

MCF-7 cells were seeded in hormone-depleted medium for 1 day prior to transfection. 100 nM siGENOME Non-Targeting siRNA Pool #1 or ERα ON-TARGETplus SMARTpool siRNA (Dharmacon) was then transfected into MCF-7 cells using Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. 48 hrs following transfection, the cells were treated with either E2 or ethanol for 45 min (for western blot analysis, 3C and ChIP assays) or 8 hrs (for mRNA analysis). Total RNA was isolated and reverse transcribed with oligo $(dT)_{15}$ primer (Promega) and real-time PCR was used for quantification.

Note: Microarrays, 3C, ChIP-3C, RT-qPCR, and siRNA knockdowns were repeated at least twice. All oligonucleotide sequences are listed in Supplementary Table 10.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge the Genome Technology and Biology Group at the Genome Institute of Singapore for technical support; Mr. Atif Shahab, Mr. Chan Chee Seng, and Mr. Fabianus H. Mulawadi for computing support; Drs Shujun Luo and Gary Schroth for Illumina sequencing support; and Drs Wouter de Laat, Bing Ren, and X. Shirley Liu for advice. M.J.F., P.Y.H.H., Y.H., P.Y.T. and Y.K.L. are supported by A*STAR Scholarships. M.J.F. is supported by a L'Oreal-UNESCO For Women In Science National Fellowship. Y.R. and C.L.W. are supported by A*STAR of Singapore and NIH ENCODE grants (R01 HG004456-01, R01HG003521-01, and part of 1U54HG004557-01).

References

1. Fraser P. Transcriptional control thrown for a loop. Curr Opin Genet Dev. 2006; 16:490–5. [PubMed: 16904310]
2. Collas P, Dahl JA. Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. Front Biosci. 2008; 13:929–43. [PubMed: 17981601]
3. Wei CL, et al. A global map of p53 transcription-factor binding sites in the human genome. Cell. 2006; 124:207–19. [PubMed: 16413492]
4. Wold B, Myers RM. Sequence census methods for functional genomics. Nat Methods. 2008; 5:19–21. [PubMed: 18165803]
5. Massie CE, Mills IG. ChIPping away at gene regulation. EMBO Rep. 2008; 9:337–43. [PubMed: 18379585]
6. Carroll JS, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell. 2005; 122:33–43. [PubMed: 16009131]
7. Carroll JS, et al. Genome-wide analysis of estrogen receptor binding sites. Nat Genet. 2006; 38:1289–97. [PubMed: 17013392]
8. Lin CY, et al. Whole-genome cartography of estrogen receptor alpha binding sites. PLoS Genet. 2007; 3:e87. [PubMed: 17542648]
9. Lupien M, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008; 132:958–70. [PubMed: 18358809]
10. Welboren WJ, et al. ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J. 2009; 28:1418–28. [PubMed: 19339991]
11. West AG, Fraser P. Remote control of gene transcription. Hum Mol Genet. 2005; 14 Spec No 1:R101–11. [PubMed: 15809261]
12. Woodcock CL. Chromatin architecture. Curr Opin Struct Biol. 2006; 16:213–20. [PubMed: 16540311]

13. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002; 295:1306–11. [PubMed: 11847345]
14. Hagege H, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc. 2007; 2:1722–33. [PubMed: 17641637]
15. Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet. 2005; 37:31–40. [PubMed: 15608638]
16. Cai S, Lee CC, Kohwi-Shigematsu T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat Genet. 2006; 38:1278–88. [PubMed: 17057718]
17. Zhao Z, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet. 2006; 38:1341–7. [PubMed: 17033624]
18. Ling JQ, et al. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science. 2006; 312:269–72. [PubMed: 16614224]
19. Simonis M, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 2006; 38:1348–54. [PubMed: 17033623]
20. Wurtele H, Chartrand P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. Chromosome Res. 2006; 14:477–95. [PubMed: 16823611]
21. Dostie J, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 2006; 16:1299–309. [PubMed: 16954542]
22. Tiwari VK, Cope L, McGarvey KM, Ohm JE, Baylin SB. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. Genome Res. 2008; 18:1171–9. [PubMed: 18502945]
23. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory interactions in vivo. Nat Genet. 2002; 32:623–6. [PubMed: 12426570]
24. Osborne CS, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet. 2004; 36:1065–71. [PubMed: 15361872]
25. Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. Nat Methods. 2007; 4:895–901. [PubMed: 17971780]
26. Barski A, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007; 129:823–37. [PubMed: 17512414]
27. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev. 2006; 20:2922–36. [PubMed: 17079683]
28. Laganiere J, et al. From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc Natl Acad Sci U S A. 2005; 102:11651–6. [PubMed: 16087863]
29. Hsu F, et al. The UCSC Known Genes. Bioinformatics. 2006; 22:1036–46. [PubMed: 16500937]
30. Kushner PJ, et al. Estrogen receptor pathways to AP-1. J Steroid Biochem Mol Biol. 2000; 74:311–7. [PubMed: 11162939]
31. Pan YF, et al. Regulation of estrogen receptor-mediated long-range transcription via evolutionarily conserved distal response elements. J Biol Chem. 2008
32. Chen X, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008; 133:1106–17. [PubMed: 18555785]

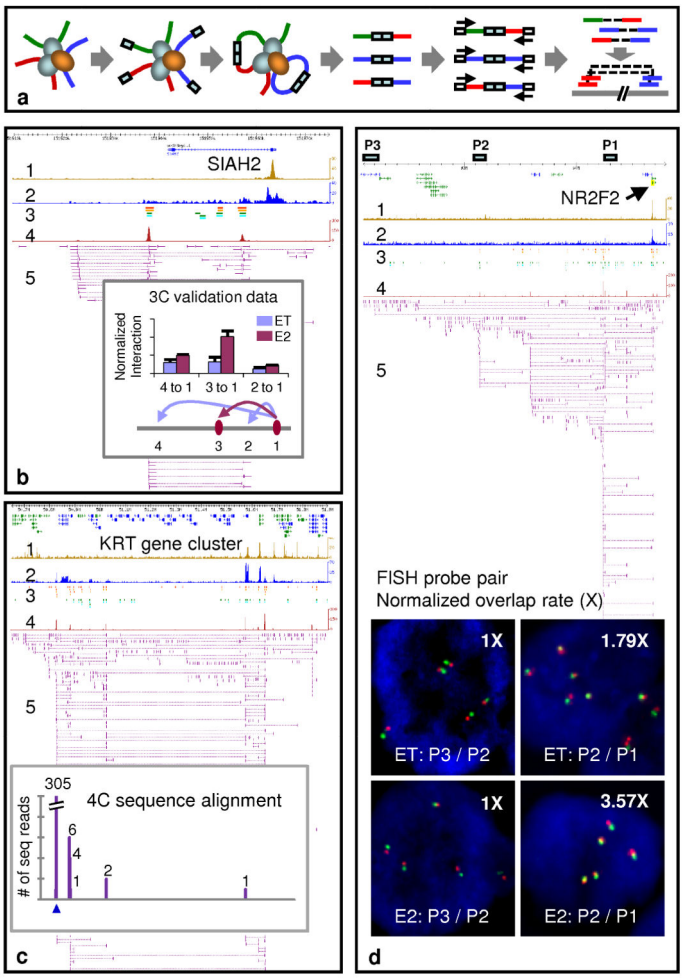

Figure 1. ChIA-PET method with validations

(a) ChIA-PET schematic. DNA fragments in sonicated, ChIP-enriched chromatin complexes were processed by linker ligation, proximity ligation, PET extraction, sequencing, and mapping to reveal interacting loci. **(b)** ChIA-PET browser tracks: 1, H3K4me3 ChIP-Seq; 2, RNAPII ChIP-Seq; 3, ERα (orange) and FoxA1 ChIP-chip (green)9; 4, ERα ChIA-PET density ; 5, Inter-ligation PETs. Inset: 3C validation of interacting ERαBS (purple) and controls (blue) under ethanol (ET) and oestrogen-induction (E2). **(c)** 4C validation, showing 4C bait region (blue) and interaction targets (purple bars). **(d)** FISH validation, showing increased P2/P1 interactions under E2-induction with background normalization (P3/P2). FISH probe genomic locations (P1/P2/P3) are indicated.

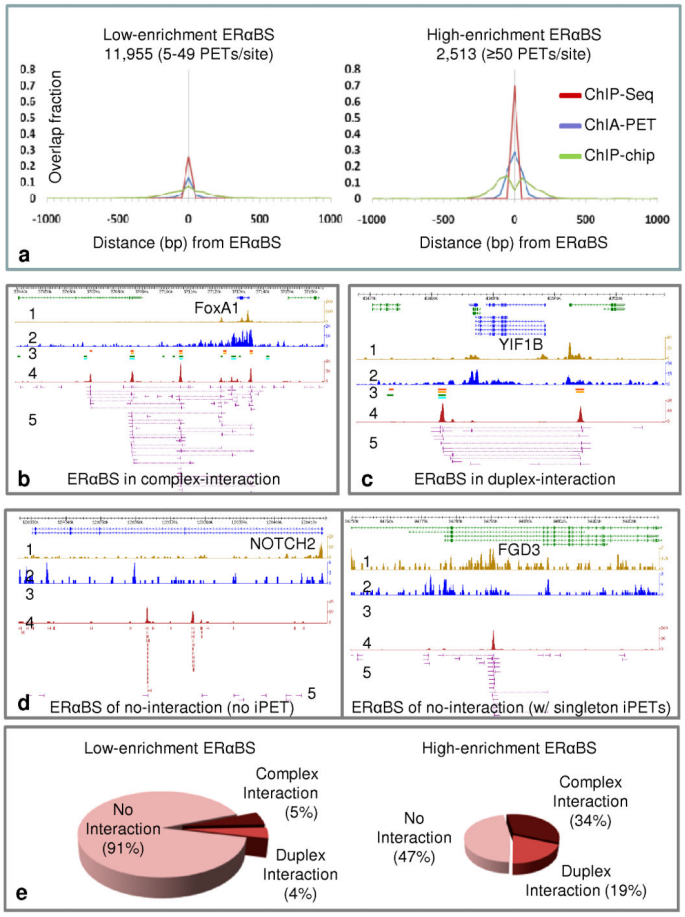

Figure 2. ERαBS reproducibility and association with chromatin interactions

(a) Numbers of ERαBS identified with different ChIP enrichment cutoffs and reproducibility analyses as measured by overlapping with another ChIA-PET dataset (IHH015F), ChIP-Seq10, and ChIP-chip7 data. Examples of ERαBS involved in **(b)** complex-interactions, **(c)** duplex-interactions, and **(d)** no-interactions (singleton inter-ligation PETs only or no inter-ligation PETs). **(e)** ERαBS distribution in different categories of interactions as exemplified in **b-d**.

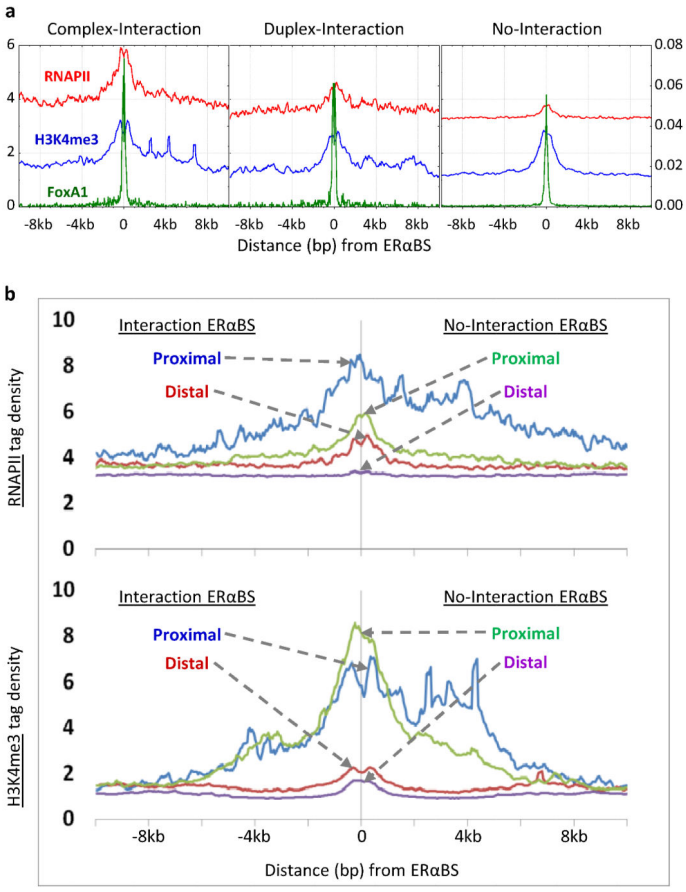

Figure 3. Association of ERα-bound chromatin interactions with functional marks

(a) Association of ERαBS in complex-, duplex-, and no-interaction categories with RNAPII (red), H3K4me3 (blue), and FoxA1 (green) functional marks. **(b)** Association of proximal and distal interacting and non-interacting ERαBS with H3K4me3 and RNAPII functional marks.

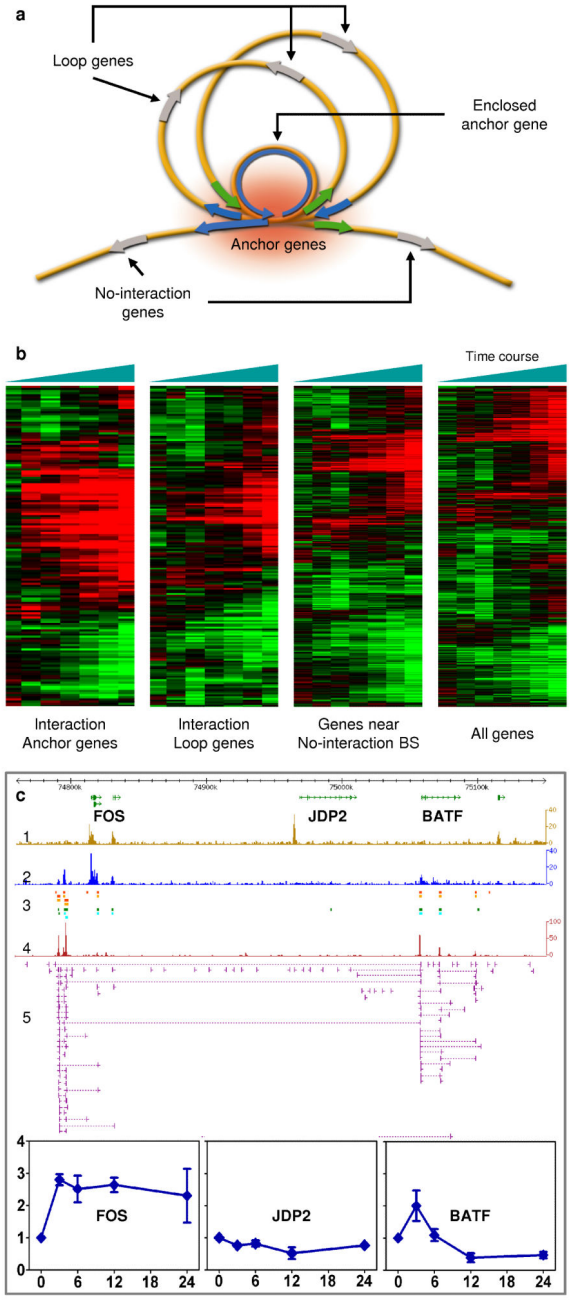

Figure 4. Proposed ERα-bound chromatin interaction and transcription regulation mechanism

(a) Distal ERαBS interact with proximal sites, forming chromatin loops. Anchor genes (green and blue) are close to interaction anchors with concentrated active transcriptional machinery (red shading). Other genes far from interaction centers (grey) are less active. **(b)** Expression microarray data (oestrogen induction from 0 to 48h; red denotes activation; green denotes repression) for interaction anchor genes, loop genes, and genes near no-interaction ERαBS, with all other UCSC Genes29 ("All genes") denoting background. **(c)** ChIA-PET interactions data at the *FOS/JDP2/BATF* loci. Transcription activities are shown

by H3K4me3/RNAPII ChIP-Seq and RT-qPCR analysis (oestrogen induction from 0 to 24h).

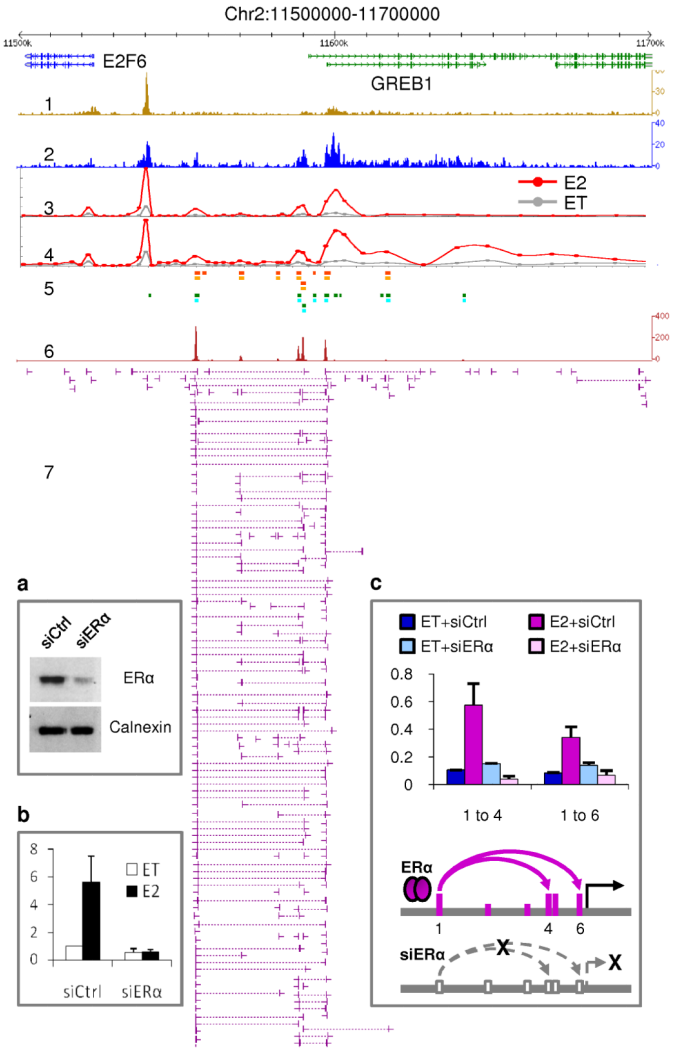

Figure 5. ERα-bound chromatin interactions are required for transcription activation

Genome browser at the *GREB1* locus showing data tracks: H3K4me3 and RNAPII ChIP-Seq (1 and 2); RNAPII ChIP-qPCR scans (3 and 4) using different RNAPII antibodies under oestrogen-induction (E2, in red) and ethanol control (ET, in grey); ERα (orange) and FoxA1 (green) ChIP-chip7 (5); ChIA-PET density (6) and interaction data (7). Inset: siRNA knockdown experiments. MCF-7 cells were transfected with siRNA against ERα (siERα) or control (siCtrl), and then analyzed by **(a)** western blot with ERα and calnexin (control) antibodies; **(b)** RT-qPCR on *GREB1* expression; and **(c)** 3C assays at *GREB1*: siERα knockdown abolishes chromatin interactions and turns off transcription.

Table 1

Summary statistics of library PET sequences

				Self-ligation		Intrachromosome inter-ligation		Interchromosome inter-ligation	
Library Code	**Library identity**	**Total PET**	**Unique PET**	**PET**	***PET clusters**	**PET**	**^PET clusters**	**PET**	**^PET clusters**
Small scale testing of the ChIA-PET method									
IHM001N	ChIA-PET	715,369	271,648	78,706	2,701	16,677	176	176,265	0
IHM001H	ChIA-PET	764,899	293,754	103,740	3,405	17,718	215	172,296	0
IHM043	ChIP-PET	1,118,509	745,251	634,993	1,158	7,386	2†	102,872	1
SHC007	ChIP-PET	361,241	214,668	192,511	489	2,196	0	19,961	0
IHM062	ChIA-PET (IgG)	436,248	217,708	40,847	0	11,254	0	165,607	0
Chimeras analysis									
IHH015M	ChIA-PET (AA+BB)	4,246,429	2,049,719	953,384	3,909	129,492	2,183	966,843	3
IHH015C	ChIA-PET (chimeras)	5,904,476	1,790,714	15,490	35	98,805	0	1,676,419	0
Large scale ChIA-PET analysis									
IHM001F	ChIA-PET	31,828,194	4,638,633	1,249,081	14,560	234,400	1,451	3,155,152	15
IHH015F	ChIA-PET	19,590,581	6,125,099	1,841,684	6,665	348,057	3,543	3,935,358	4

Notes: ChA-PET data mapped at satellites and structural variation sites were removed.

*Self-ligation PET clusters for identifying binding sites (FDR < 0.01, PET count ≥ 5).

^Inter-ligation PET clusters for identifying interactions include at least 2 (small scale) or 3 (chimeras and large scale analysis) overlapping PETs (FDR < 0.05). Interchromosomal interactions were subjected to manual curation.

†One interaction has a genomic span < 5 Kb, suggesting it results from extra-long self-ligation PETs, and the other has a genomic span >10 Mb and PET counts of only 2, and so could be non-specific.