CONTENTS OF SUPPLEMENTARY INFORMATION

Summary of Results

I. Key technical advances of the ChIA-PET method (Supplementary Fig. 1a)

1. ChIA-PET is an unbiased, whole-genome, and *de novo* approach for long-range chromatin interaction analysis.
2. A ChIA-PET experiment is capable of providing two global datasets: the protein factor binding sites (equivalent to ChIP-sequencing) and the interactions among the binding sites.
3. ChIA-PET involves ChIP to reduce the complexity for genome-wide analysis and adds specificity to chromatin interactions bound by specific factors of interest.
4. ChIA-PET also involves other technical improvements including the use of sonication to "shake off" random chromatin attachment from specific chromatin interaction complexes, thus decreasing interaction noise, and the use of linker barcodes to discriminate chimeric ligations, therefore, diminishing false positives of chromatin interaction.
5. ChIA-PET is compatible with tag-based next-generation sequencing approaches such as Roche 454 pyrosequencing, Illumina GA, ABI SOLiD, and Helicos' Heliscope.
6. ChIA-PET is applicable to many different protein factors involved in transcriptional regulation or chromatin structural conformation.

II. Key findings (Supplementary Fig. 1b)

1. The majority high-confidence ER-αBS are involved in ER-α-bound chromatin interactions, wherein distal ER-αBS loop toward gene promoters through connection with proximal ER-αBSs.
2. Genes with promoters near interaction anchor regions (anchor genes) are more likely to be active in transcription and up-regulated by oestrogen and ER-α binding. Distant genes could be brought together by such interactions for coordinated transcription activation.
3. ER-α-bound interactions could have heterogeneous looping structures for different functions. Small gene-centric loops could enhance transcription efficiency, and large multi-genic loops could partition genes residing in loop regions (loop genes) away from interaction anchor centers in transcriptional hubs.

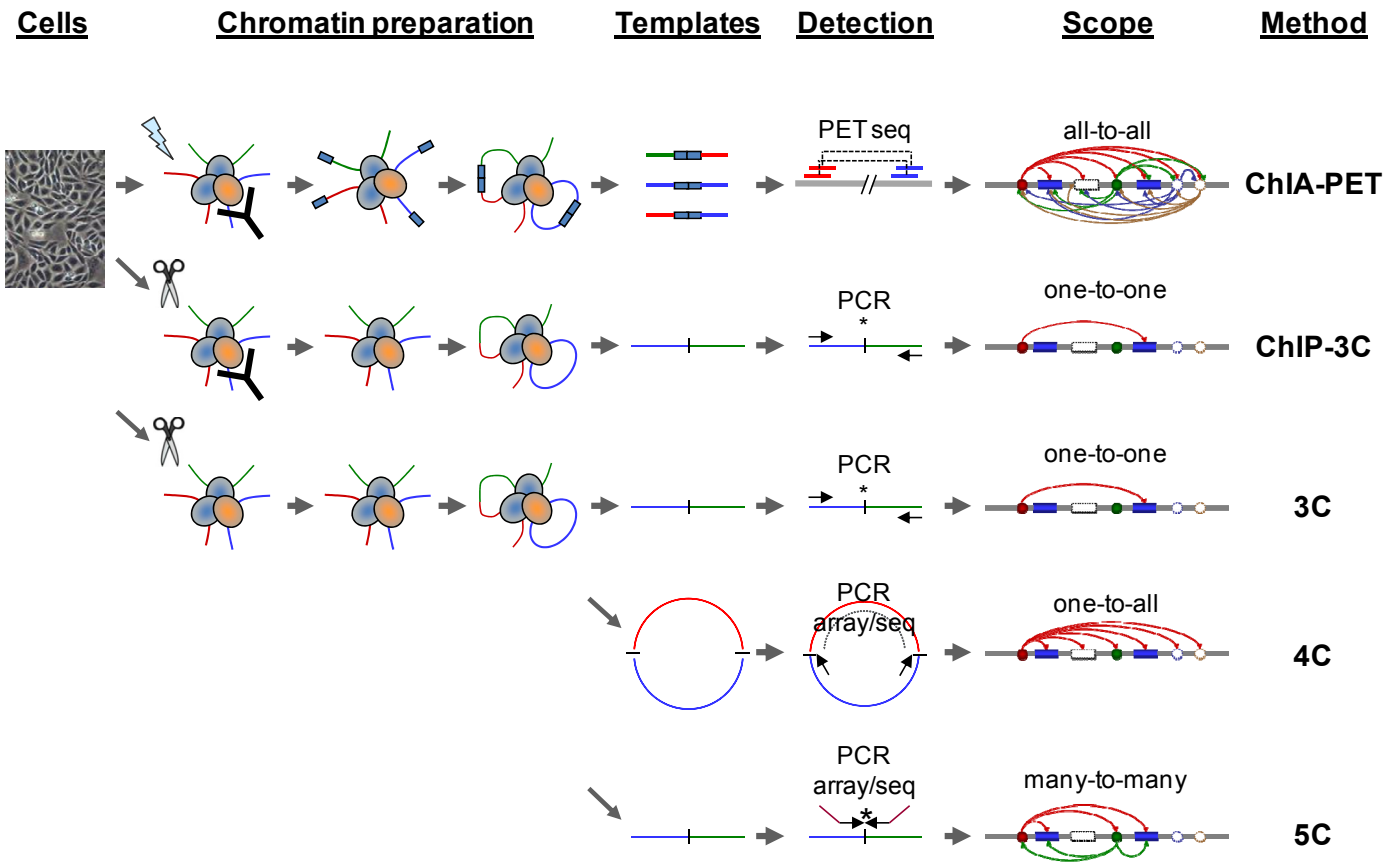
III. A model of chromatin interactions for ER-α function (Supplementary Fig. 1b)

In this model, we hypothesize that ER-α protein dimers bind primarily to distal regulatory elements and initiate long-range chromatin interactions involving promoter regions of target genes. These interactions form DNA loop structures with multiple ER-α binding at the anchoring center. Multiple small and gene-centric loops could package genes near the anchoring center in a tight sub-compartment of chromatin looping structures, which could increase the local concentration of ER-α proteins, and therefore, attract and retain more molecules of cofactors as well as transcriptional machinery (such as RNAPII proteins) for enhanced transcriptional activation. This topological structure could also provide transcription efficiency, allowing RNAPII to cycle through the tight circular gene templates. The large interaction loops, however, are more likely to link together distant genes near the anchor sites at both ends of the loop for coordinated regulation, and separate the genes residing in long loops away from the active ER-α regulation.

IV. Conclusions

1. ChIA-PET can simultaneously identify protein binding sites and chromatin interactions between the protein binding sites in a whole-genome, *de novo*, and unbiased manner.
2. Long-range chromatin interaction is a primary mechanism for ER-α to function in connecting distal regulatory elements together with gene promoters for transcriptional regulation.

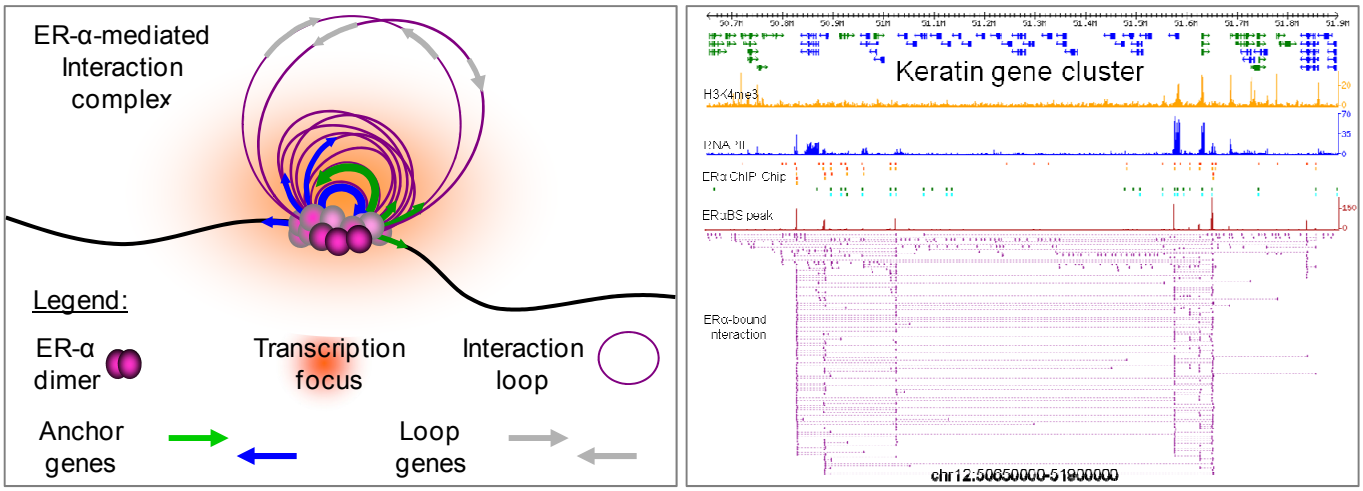
Note: A **Glossary** is provided on the last page.



Supplementary Figure 1. Summary of results

a. Key technical advances of ChIA-PET compared to other methods

ChIA-PET is a whole genome approach for long-range chromatin interaction analysis, and is capable of providing "all-to-all" data consisting of global protein factor binding sites (equivalent to ChIP-sequencing) and the interactions among the binding sites. ChIA-PET is applicable to many different protein factors involved in transcriptional regulation or chromatin structural conformation. The ChIA-PET method involves sonication (lightning bolt) as opposed to restriction enzyme digestion (scissors) to "shake off" random chromatin attachments from specific interaction complexes, ChIP to reduce complexity for genome-wide analysis and add specificity to chromatin interactions bound by specific factors of interest, the use of linker barcodes to discriminate chimeric ligations as well as identify samples. ChIA-PET is a paired tag method and hence is compatible with tag-based next-generation sequencing approaches such as Roche 454 pyrosequencing, Illumina GA, ABI SOLiD, and Helicos' Heliscope.



Supplementary Figure 1. Summary of results

b. The ER-α chromatin interaction model

Left: We speculate that ER-α protein dimers bind to distal regulatory elements and initiate long-range chromatin interactions involving promoter regions of target genes. These interactions form DNA loop structures with multiple ER-α binding at the anchoring center. Multiple small and gene-centric loops could package genes near the anchoring center in a tight sub-compartment, which could increase the local concentration of ER-α proteins, and therefore, attract and retain more molecules of cofactors as well as transcriptional machinery (such as RNAPII proteins) for enhanced transcriptional activation. This topological structure could also enhance transcription efficiency, allowing RNAPII to cycle the tight circular gene templates. The large interaction loops, however, are more likely to link together distant genes at either end of the loop residing near anchor sites for coordinated regulation, and separate the genes in long loops from the active ER-α regulation. **Right:** Most of the characteristics featured in the model of ER-α-bound chromatin interactions are exemplified in the complex interaction at the keratin gene cluster region in Chr12:50650000-51900000 (also in **Fig. 1c** in the paper).

Supplementary Methods

I. Cell culture and oestrogen treatment

MCF-7 cells were grown to at least 80% confluency in DMEM/F12 (Invitrogen/Gibco) supplemented with 5% Fetal Bovine Serum (FBS) (Invitrogen/Gibco), penicillin (Invitrogen), streptomycin (Invitrogen), and gentamycin (Invitrogen). In preparation for the 17 beta-estradiol ("oestrogen", E2, Sigma) treatment, cells were grown in hormone-free media: they were washed with PBS and incubated in phenol red-free medium (Invitrogen/Gibco) supplemented with 5% charcoal-dextran stripped FBS (Hyclone), penicillin, streptomycin, gentamycin, and L-glutamine (Invitrogen) for a minimum of 72 hours. Hormone-depleted cells were treated with oestrogen to a final concentration of 100 nM for 45 min before the ChIP procedure. The control cells were treated with an equal volume and concentration of vehicle, ethanol (ET, Merck), for 45 min. For a ChIA-PET experiment, we routinely use approximately 10^8 cells from 6 150-mm diameter cell culture plates.

II. Chromatin immunoprecipitation (ChIP)

ChIP protocol was performed as described previously[1]. Briefly, we used 1% formaldehyde to crosslink the cells, and sonication to break the chromatin fibers. ER-α specific antibody (HC-20, Santa Cruz or Mab-NRF3A6-050, Diagenode) was used to enrich ER-α bound chromatin fragments. IgG specific antibody (sc-2027, Santa Cruz) were also used for ChIP analyses. ChIP material bound on the antibody beads was subjected to ChIA-PET library construction.

III. ChIA-PET library construction and sequencing

IHM001 ChIA-PET data. MCF-7 cells were grown to at least 80% confluence. In preparation for the oestrogen treatment, cells were grown in hormone-free media for a minimum of 72 hours. Hormone-depleted cells were treated with oestrogen at a final concentration of 100 nM for 45 min before the ChIP procedure. We used two independent MCF-7 chromatin preparations in this ChIA-PET analysis. ChIP protocol was performed as described previously[1]. We used 1% formaldehyde to crosslink the cells, and sonication to break the chromatin fibers. ER-α specific antibody (HC-20, Santa Cruz) was used to enrich ER-α bound chromatin fragments. The DNA fragments present in chromatin fragments were end-repaired using T4 DNA polymerase (NEB), followed by overnight ligation of biotinylated half-linkers that contain a flanking MmeI site (IDT), using T4 DNA ligase (Fermentas) at 16ºC, with mixing. The linker added DNA fragments were then phosphorylated with T4 polynucleotide kinase (NEB), and followed by a second ligation reaction overnight at 22ºC under dilute conditions on the ChIP beads (< 0.2 ng DNA per µl reaction). The cross-links in the DNA/protein complexes were then reversed by incubation at 65ºC overnight with 0.2% SDS (Ambion) and proteinase K (Ambion), and the DNA fragments were purified by phenol/chloroform isopropanol precipitation. The nick in one strand of the DNA at the ligation site was subsequently repaired by incubation with *E. coli* DNA ligase (NEB) and *E. coli* DNA polymerase I (NEB) at 16ºC overnight. The purified DNA was then digested by MmeI (NEB) for at least 2h at 37ºC to release the tag-linker-tag structure (paired end tag, PET). The biotinylated PETs were then immobilized on streptavidin-conjugated magnetic Dynabeads (Invitrogen) and the ends of each PET structure were then ligated to an adapter by T4 DNA ligase (Fermentas) at 22ºC overnight with mixing followed by 20 rounds of PCR to amplify the PETs. This PCR product was the template for Roche 454 pyrosequencing analysis. From the two independent MCF-7 chromatin preparations, we generated two ER-α ChIA-PET library datasets in parallel: IHM001H (0.9 million raw PETs) and IHM001N (0.9 million raw PETs). Subsequently, we subjected the PET templates to Illumina GAII for large scale paired end sequencing analysis, to give a combined final dataset (IHM001F) with 32 million raw PET sequences. A reference figure is presented as **Supplementary Fig. 2,** and statistics

can be found in **Table 1**. All oligonucleotide sequences of linker and adapter sequences used in ChIA-PET are listed in **Supplementary Table 10**.

IHH015F ChIA-PET data. For a biological replicate, we repeated the ER-α ChIA-PET experiment using a different batch of MCF-7 cells and another antibody against ER-α, Mab-NRF3A6-050 ("F3A6") antibody (Diagenode)[2] (also called AC-066-100), and generated a new dataset (IHH015) with 20 million raw PET sequences, which after processing resulted in 5.5 million uniquely mapped PETs. IHH015 was essentially prepared in a similar manner to IHM001, with a few novel modifications for a gentler ChIA-PET preparation to result in the retention of more chromatin interactions: ChIP enrichment was performed using a new, different antibody against ER-α. Also, after overnight ligation of biotinylated, barcoded half-linkers, the chromatin was eluted off the beads with 1% SDS, followed by Triton X-100, and ligated overnight at 22ºC under ultra-dilute conditions (< 0.01 ng DNA per µl reaction). The cross-links in the DNA/protein complexes were reversed by incubation at 37ºC overnight with proteinase K (Ambion), and the DNA fragments were purified by phenol/chloroform isopropanol precipitation. No nick repair was performed, as the ligation step was deemed to have been sufficient to repair nicks, and the DNA was reverse cross-linked at the gentler temperature of 37 ºC. The MmeI-digested, biotinylated PETs were then immobilized on streptavidin-conjugated magnetic Dynabeads (Invitrogen) and the ends of each PET structure were then ligated to an adapter by T4 DNA ligase (Fermentas) at 22ºC overnight with mixing followed by 22 rounds of PCR to amplify the PETs. This PCR product was the template for 454 pyrosequencing as a quality control and Illumina GAII paired end sequencing analysis. A reference figure is presented as **Supplementary Fig. 2,** and statistics can be found in **Table 1**. All oligonucleotide sequences of linker and adapter sequences used in ChIA-PET are listed in **Supplementary Table 10**.

IgG ChIA-PET library (IHM062). For a genome-wide negative control, we used an IgG antibody instead of ER-α antibody in the ChIP step for a mock ChIA-PET experiment. The rest of the procedure is the same as for IHM001 ChIA-PET experiment. We conducted a full run of Roche 454 pyrosequencing, and generated a total of 0.52 million raw PET sequences. A reference figure is shown as **Supplementary Fig. 2,** and statistics can be found in **Table 1**. Discussion can be found in the main text and **Supplementary Note 1**. All oligonucleotide sequences of linker and adapter sequences used in ChIA-PET are listed in **Supplementary Table 10**.

IV. Linker nucleotide barcoding for ChIA-PET library construction

Linker oligonucleotide sequences are used in ChIA-PET library construction to connect proximity ligated DNA fragments. We designed two different linkers (A and B) with specific nucleotide barcodes (CG or AT) for each of the two linker sequences (**Supplementary Fig. 3**). In the first ligation step, chromatin was divided into two aliquots with linker A added to one aliquot and linker B added to the other aliquot. After linker ligation and removal of free linkers, the two aliquots were mixed together for proximity ligation. The ligation products were analyzed by PET sequencing. The PET sequences with AA (CG/CG) and BB (AT/AT) linker barcode composition are considered to be possible intra-complex ligation products, while the PET sequences with AB (CG/AT) linker composition are considered to be derived from chimeric ligation products between DNA fragments bounded in different chromatin complexes. In a separate biological repeat of the ER-α ChIA-PET experiments (the IHH015 set of experiments), we created a library with A and B linker sequences. To specifically and equally compare the AB linker PET sequences with the AA and BB linker PET sequences for their mapping characteristics, we collected all uniquely mapped AB linker PET sequences (IHH015C; 1.8 million PETs), and created an AA and BB linker PET dataset that has almost equal numbers of PETs (IHH015M; 2.0 million PETs) for comparison. The results of this analysis are

described in the text and **Supplementary Note 2**. All oligonucleotide sequences of linker and adapter sequences used in ChIA-PET are listed in **Supplementary Table 10**.

V. ChIP-PET library construction and sequencing

As a ChIP-PET library does not use proximity ligation to capture the relationship of DNA fragments tethered by chromatin complex, we do not expect to see many inter-ligation PETs. If we do see any, these inter-ligation PETs are most likely due to non-specific ligations. Therefore, ChIP-PET can be used as another genome-wide control to ChIA-PET. We modified the original ChIP-PET method[3] and developed the cloning-free version of ChIP-PET method.

Cloning-free ChIP-PET library (IHM043). We used the cloning-free ChIP-PET method and constructed a ChIP-PET library from the same (different batch) MCF-7 chromatin materials used for ChIA-PET analysis. Specifically, in the ChIP-PET procedure, after the first ligation (ligation of the half-linkers) and the phosphorylation of chromatin with T4 polynucleotide kinase, cross-links were reversed by incubation at 65ºC overnight with 0.2% SDS (Ambion) and proteinase K (Ambion). DNA was purified by phenol/chloroform isopropanol precipitation. Subsequently, overnight dilute ligation was performed on the ChIP DNA with T4 DNA ligase (Fermentas) at 22ºC without agitation. Nick repair was then performed by incubation with *E. coli* DNA ligase (NEB) and *E. coli* DNA polymerase I (NEB) at 16ºC overnight, followed by DNA purification which included a Plasmid-Safe Enzyme (Epicenter) step for removing uncircularized products. MmeI digestion and subsequent steps were performed as per the ChIA-PET protocol. For the ER-α ChIP-PET library, we conducted 4 GSFLX runs and generated a total of 2.82 million raw sequences for further analysis. This library is called IHM043. A reference figure is shown as **Supplementary Fig. 2**. Statistics are shown in **Table 1**. Discussion can be found in the main text and **Supplementary Note 1**. All oligonucleotide sequences of linker and adapter sequences used in ChIA-PET are listed in **Supplementary Table 10**.

Cloning-based ChIP-PET library data (SHC007). We previously generated ER-α ChIP-PET library data using the cloning-based ChIP-PET method[1]. We reprocessed this dataset (635K raw PET sequences) by mapping it to hg18 genomic assembly and performed downstream analyses with the same ChIA-PET pipeline used to analyze other ChIA-PET libraries. Statistics are shown in **Table 1**. Discussion can be found in the main text and **Supplementary Note 1**.

A summary of the libraries is provided in **Table 1**.

VI. Other genome-wide datasets generated in this study for ChIA-PET characterization

In order to characterize the ER-α-bound chromatin interactions in transcriptional regulation, we generated additional genome-wide datasets from MCF-7 cells under oestrogen induction conditions as supporting information:

H3K4me3 ChIP-Seq data. H3K4me3 antibody (ab8580, Abcam) was used to generate ChIP-enriched DNA fragments for Illumina GAII single read sequencing analysis. The H3K4me3 ChIP-seq data was mapped to hg18 genome using Illumina's ELAND program, and enrichment peaks for H3K4me3 binding were identified using ChIP-seq peak calling algorithm as previously described[4]. This library was normalized to a background library, and 46K H3K4me3 binding sites were identified in the MCF-7 genome from this dataset (threshold = 4). This dataset characterizes the promoter status of genes in MCF-7 cells during oestrogen induction, and was used to annotate the genes involved in ER-α-bound chromatin interactions.

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 GAII single read sequencing analysis. We generated a total of 2.87 million unique sequences for further 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 described[4]. 13K RNAPII binding sites were identified in the MCF-7 genome from this dataset. This dataset provided a profile of gene transcription activities, which was used characterize transcription status of genes involved in ER-α-bound interactions.

Genomic DNA-PET of 10 Kb insert data. It is known that the MCF-7 genome contains many structural rearrangements[5]. Therefore, ChIA-PET data generated from this genome for detecting long-range interactions could be complicated by genome structural differences between this and the reference genome (hg18). To avoid such complications, we constructed genomic DNA-PET libraries with insert sizes around 10 Kb in span. We generated 35 million "DNA-PET" sequences, which is a 100-fold physical coverage of the MCF-7 genome. This dataset provides digital karyotyping information regarding deletions, inversions, translocations, and insertions in the MCF-7 genome, and identifies rearranged genomic regions. From this dataset, we obtained a list of 97 high-confidence structural variations, and used this information to filter out inter-ligation PET clusters located in these genome aberration regions, and therefore reduce false positive calls. The detailed results of these datasets will be further described in a separate publication.

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-treated MCF-7 cells (mock-treated 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 2.0). 3 replicates were performed for each time point. The data was analyzed using two different time-course differential expression analysis methods: Pooled Variance Meta Analysis and LIMMA and ranked by their scores. The top 5,000 probes or ~10% of all probes were obtained from each ranking and combined resulting in ~7,500 probes. The set was further filtered using mean inclusive Data-driven Smoothness Enhanced Variance Ratio Test (dSEVRaT) with dSEVRaT score > 200 resulting in ~3700 probes. Up and down regulation for each gene was decided based on their trend using hierarchical clustering carried out using Eisen software[6] (http://rana.lbl.gov/EisenSoftware.htm). Further details and results will be described in a subsequent publication.

VII. PET sequence data processing, mapping, and statistical analyses

PET extraction and mapping. The pipeline for processing PET sequences is described in Chiu et al., 2006[7]. Briefly, the raw PET sequence reads were processed through the 'PET-Tool' program[7] for extraction of PET sequences. Based on sequence content, redundant sequences were collapsed into a non-redundant PET sequence set, which were mapped to the reference human genome sequence (hg18) using ELAND[4]. Based on the mapping coordinates of the uniquely mapped PETs, PET sequences aligned to the exact same genome location ($\pm$2 base) were considered to be PCR replicates derived from the same original proximity ligation product with minor variances in nucleotide content due to variations intrinsic to MmeI restriction enzyme digestion[8], adapter ligation, or sequencing errors. This measure counters the biases arising from PCR amplification, and this representation of PET sequences best reflects the original complexity and content of proximity ligation products. The uniquely aligned PETs were subjected to further analysis, whereas the ambiguously mapped PETs (mapped to multiple locations, partially mapped or not mapped at all) based on the current algorithm were not analyzed.

PET classification. Based on mapping characteristics, each of the uniquely aligned PET sequences can be classified by whether it was derived from one DNA fragment or two DNA fragments. Briefly, if the two tags of a PET were mapped on the same chromosome with the genomic span in the range of ChIP DNA fragments (less than 3 Kb), with expected self-ligation orientation and on the same strand, we considered that this PET was most likely derived from a self-ligation of a single ChIP DNA fragment[1], and therefore called the PET a "**self-ligation PET**". If a PET did not fit into these criteria, we considered that the PET most likely resulted from a ligation product between two DNA fragments; therefore we called the PETs "inter-ligation PETs". The two tags of the "inter-ligation PETs" do not have fixed tag orientations, might not be found on the same strands, might have any genomic span, and might not map to the same chromosome. If the two tags of an "inter-ligation" PET mapped in same chromosome but with a span > 3 Kb in any orientation, these PETs were called "**intrachromosomal inter-ligation PETs**". If the two tags mapped in the same chromosome with spans of less than 3 Kb but not with expected orientation or to the same strands, these PETs were called "different orientation inter-ligation PETs". The number of PETs in this category is very small. PETs which mapped to different chromosomes were called "**interchromosomal inter-ligation PETs**".

Identification of ER-α ChIP enriched peaks (binding sites). We used self-ligation PET sequences as the basis for identifying ER-α ChIP enriched sites, because the self-ligation PETs provide the most reliable mapping (20+20 bases) to the reference genome and best represent the ChIP fragments by providing two defined end points. The ChIP enrichment peak-finding algorithm was described previously[1]. Briefly, we called a peak a binding site if there were multiple self-ligation PETs overlapping in that region. Multiple self-ligation PETs would not occur by random chance. We applied a Monte Carlo simulation to estimate the random background of PET-derived virtual DNA overlaps, and used the estimated background noise to calculate the false discovery rate of ER-α binding sites identified by the ChIA-PET data. We further filtered out binding sites involved in satellite regions using RepeatMasker[9], as these have been found to be non-specifically pulled down by ChIP procedures. Moreover, we removed binding sites which were present in regions with severe structural variations as found by the 10 Kb insert data library. The numbers of self-ligation and inter-ligation PETs (within $\pm$250 bp) are reported at each site. Together, the total number of self-ligation and inter-ligation PETs equals the ChIP enrichment at that site, as all PETs are derived from chromatin fragments that have undergone the ChIP enrichment procedure. This number is therefore called the "ChIP enrichment count". The list of all binding sites is given in **Supplementary Table 1**. In total, we have 14,560 binding sites in IHM001F and 6,665 binding sites in IHH015F with false discovery rates (FDR) < 0.01. As MCF-7 cancer cells are known to contain amplicon regions in their genomes, we marked ER-αBS based on whether they overlapped amplicon regions as identified by array CGH[10]. Because these binding sites could still play important roles in MCF-7 cells, we did not discard them, and still used them in statistical and other analyses. However, these binding sites have a note in **Supplementary Table 1** to say that they can be found in MCF-7 amplicon regions.

Reproducibility of ER-α ChIP enriched peaks (binding sites). A reproducibility analysis was performed for the pilot library datasets IHM001H and IHM001N, in which the same antibody were used for ChIP. The results are shown in **Supplementary Fig. 3**. For analyses of the two datasets of ER-αBS identified by IHM001F and IHH015F, the reproducibility analysis method is described below and the results are shown in **Supplementary Fig. 13**. The differences of binding sites mostly arise from low ChIP-enrichment binding sites, and could be due to technical variations, MCF-7 batch-to-batch variations, or in the case of differences between IHH015F and IHM001F, because different antibodies against ER-α were used.

To further measure the reliability of the ER-α definition for each ChIA-PET replicate, we compared genomic coordinates of the sites to published ER-αBS and ER-αBS from the other ChIA-PET replicate library (IHM001F and IHH015). We used 50 bp binned intervals around each ChIA-PET site to count the number of ER-αBS from ChIP-chip[11], ChIP-seq[2], and replicate ChIA-PET libraries located in the genomic interval. The distance between sites was measured between borders of ChIA-PET site and borders ChIP-seq or ChIP-chip peak, correspondingly. Then, we took the total number of such neighboring sites at the interval to count the fraction of ChIA-PET ER-αBS and presented the percentages as profiles. For overlapping borders we counted distance between sites as 0. This result is presented in **Fig. 2a** and more details in **Supplementary Fig. 13a**.

Identification of ER-α-bound chromatin interactions using inter-ligation PETs. To determine if an "inter-ligation PET" represents a true and specific interaction event between two DNA fragments that were bound together in close spatial proximity by an ER-α-bound protein complex, we made the following two assumptions. First, if an interaction between two DNA regions is specific, it would be enriched by the ChIP procedure, and hence it would be detected frequently in multiple times in the ChIA-PET data; while if it is non-specific and occurs by random chance, it would be sampled much less frequently than real interactions and at the level expected by chance. The frequency of ChIA-PET data for a particular region can be measured by the digital count of the overlapped inter-ligation DNA fragments inferred by PETs. As each inter-ligation PET was derived from two ChIP DNA fragments, the majority of which were less than 1,500 bp in size (**Supplementary Fig. 2**), we extended the mapped 20 bp tags to 1,500 bp along the reference genome to represent the virtual DNA fragments in pairs, and then we defined peaks (anchors) and valleys from the profile, and then we counted the number of PETs that connect two peaks (anchors) to each other. The number of PETs therefore, is the measurement of the frequency of an interaction between two regions. Hence, using this method, we can distinguish real interaction signals (multiple overlapping inter-ligation PETs) from random background noise (singleton inter-ligation PETs).

We were particularly concerned that the ChIP-enriched loci with more DNA fragments would have a higher chance of inter-ligating with each other at random, resulting in multiple inter-ligation PETs that are actually noise. To detect interactions with significantly more inter-ligation PETs than random expectation, and take into account the ChIP-enrichment bias, we formulated a statistical analysis framework to calculate the probability for the formation of any inter-ligation PETs between two regions should the ligations between DNA fragments occurred based on random chance. We modeled the non-specific interactions such that each DNA fragment has an equal chance to interact with and be ligated to any other fragments. For the null hypothesis, we assume that in the ER-α ChIP-enriched DNA fragment population, each DNA fragment has an equal chance to interact with and be ligated to any other fragments randomly. The expected level of interaction frequency between any two genomic loci and the *p*-value of any two genomic loci having the observed frequency of inter-ligation PETs were computed based on the null hypothesis. As this model gives each DNA fragment in the ChIP-enriched pool an equal chance to randomly interact, the computed significance would also be normalized against the enrichment effect by ChIP that could potentially bias the observed distribution of inter-ligations.

More formally, consider a library with N inter-ligation PETs, the total number of sampled DNA fragments is $2N$. We denote R_A and R_B as representing two DNA regions with c_A and c_B virtual DNAs, where $c_A, c_B << N$. Under the random model, the number of inter-ligation PETs that link R_A and R_B, denoted $I_{A,B}$, approximately follow a hypergeometric distribution:

$$\Pr(I_{A,B} \mid N, c_A, c_B) = \frac{\binom{c_A}{I_{A,B}}\binom{2N - c_A}{c_B - I_{A,B}}}{\binom{2N}{c_B}}$$

By this, we are able to compute a *p*-value to test if $I_{A,B}$ is over-represented. This *p*-value can then be computed upon all pairs of loci with observed DNA fragments. We converted this *p*-value into false discovery rates. Given a cut-off threshold , T, of hypergeometric *p*-value, we are able to calculate the False Discovery Rate, which is the fraction of the clusters with *p*-value below T under the empirical random model.

In addition, we removed interactions which were involved in MCF-7 chromosomal aberrations using the 10 Kb DNA-PET sequences, as well as interactions that involved satellite regions using Repeatmasker[9].

In total, we identified 1,475 inter-ligation PET clusters with 3 or more overlapping individual PET sequences and FDR < 0.05 in IHM001F and 3,561 in IHH015F **(Supplementary Table 2)**. We further filtered out binding sites involved in satellite regions using RepeatMasker[9], as these have been found to be non-specifically pulled down by ChIP procedures. As MCF-7 is known to have amplicon regions, we identified interactions present in amplicon regions based on whether they overlapped amplicon regions as identified by array CGH[10]. Because these interactions could still play important roles in MCF-7 cells, we did not discard them, and still included them in statistical and other analyses. These interactions are listed in **Supplementary Table 2** and noted they are found in amplicon regions.

The vast majority of interactions identified were intrachromosomal. In total, 24 interchromosomal inter-ligation PET clusters were found in IHM001F while 18 were found in IHH015F. To check whether this method of analysis was valid, two groups of 25 intrachromosomal interactions were randomly chosen from IHM001F and IHH015F for manual curation. Besides finding a few intrachromosomals in amplicon regions, there were no other errors observed. All interchromosomals were further subjected to manual curation, and had amplicons removed, as described in **Supplementary Note 3.** Briefly, all interchromosomals were found to be very weak interactions or noise.

ChIA-PET can discover all types of interactions (chromatin fragments that come together in close proximity). However, it cannot separate between interactions that are solely due to proximity-based random collision between protein-bound sites, interactions that are solely due to specific protein-tethering of the two sites or those that are due to a combination of these two factors, because these two mechanisms can both give rise to fragments coming together in close proximity. As it is known that proximity-based random collision mechanisms becomes much weaker after 10 Kb, and is very weak by 100 Kb, the genomic spans of the interactions is reported in **Supplementary Table 2** to allow discrimination between interactions from these different mechanisms for further studies.

Reproducibility of ER-α-bound chromatin interactions. We overlapped chromatin interactions from the two libraries based on whether they shared both anchor regions (a minimum of 1 bp overlap of the anchor regions was required to say that they shared the anchor regions). We found that chromatin interactions were reproducible, particularly in the top interactions, suggesting that the ChIA-PET method is robust and

reproducible for identifying abundant chromatin interactions, while for weak interactions the technical and biological noise can be high. The results are given in **Supplementary Table 7**.

Also, all interactions were visualized on a whole genome interaction viewer, and both views from IHH015F and IHM001F were found to look very similar (**Supplementary Fig. 12**).

Further clustering of ER-α-bound chromatin interactions into complex interactions. Many of the putative chromatin interactions (duplex interactions involving two anchors) connect to each other by overlapping anchors (anchors can be thought of the base of the loop) (**Supplementary Fig. 6**). Based on such connectivity, multiple individual intrachromosomal interactions as identified in **Supplementary Table 2** were collapsed together into interaction regions **(Supplementary Table 3).** In these interaction regions, if two duplex interactions came together such that they overlapped in at least one anchor, they were collapsed into a single complex interaction. Otherwise, they were left as stand-alone duplex interactions. Complex interactions are thought to be stronger than stand-alone duplex interactions as they involve more counts of inter-ligation PETs, and also structurally, they could involve more ER-α proteins coming together. As MCF-7 is known to have amplicon regions, we identified interaction regions that involved amplicon regions based on whether they overlapped amplicon regions as identified by array CGH[10]. Because these interaction regions could still play important roles in MCF-7 cells, we did not discard them, and still used them in statistical and other analyses. However, these interaction regions have a note in **Supplementary Table 3** to say that they overlap with amplicons. We checked by manual curation a random selection of 25 interaction units in both IHH015F and IHM001F to determine whether interaction regions appeared to be reliable. We did not find any errors.

Again, the genomic spans of the interaction regions are reported in **Supplementary Table 3** to allow discrimination between interactions from proximity-based random collisions or specific protein tethering for further studies.

ChIA-PET data visualization . We adopted the 'Generic genome browser' system[12] and developed the "ChIA-PET Genome Browser" to organize and visualize the ChIA-PET data. The "self-ligation PETs" and the "inter-ligation PETs" are displayed in separate tracks to show transcription factor binding sites and interactions, respectively. This browser also includes a custom 'Whole Genome Interaction Viewer' which provides a macroscopic picture of binding sites and interactions along with a whole genome landscape (http://cms1.gis.a-star.edu.sg/index.php). The username is "guest", and the password is "gisimsgtb".

VIII. ER-αBS association with relevant genomic features

Association of ER-αBS with ER-α-bound chromatin interactions. ER-α binding sites were grouped into categories based on their involvement with interaction characteristics: A. binding sites involved in complex interactions ("complex-interactions") where the ER-αBS lies in an interaction anchor of a complex interaction; B. in stand-alone duplex interactions ("duplex-interactions") where the ER-αBS lies in an interaction anchor of a duplex interaction; C. with "no-interactions" (either singleton inter-ligation PETs that may likely represent weak interactions or random background noise or no inter-ligation PETs at all). If a binding site could be placed into different categories, A was favored over B, which in turn was favored over C.

We associated ER-αBS with FoxA1 binding sites generated using ChIP-chip[11], H3K4me3 binding sites generated using Illumina GAII sequencing (in house, data will be presented in greater detail in a subsequent publication), and RNAPII binding sites generated using Illumina GAII sequencing (in house).

To measure the association of ER-αBS with relevant genomic features associated with active transcription we used RNAPII and H3K4me3 ChIP-seq tags. For each ER-αBS we considered a genomic interval ± 10 Kb around center of the ER-αBS. For each position in the interval we counted the number of ChIP-seq tags for RNAPII and H3K4me3 normalized by the number of control tags from a background input DNA sequencing library (in house Illumina GAII sequencing). Normalization to background sequencing at each position was calculated by using the formula:

$$Y'=(Y+1)/(Z+1)$$

Y' is the normalized number (values at axis Y to the left), Y is the original number of ChIP-seq tags, Z is the number of tags in control library, and 1 is a pseudocount (to avoid division by 0). Binned intervals were used to count the fraction of FoxA1 binding sites neighbouring ChIA-PET ER-αBS.

All graphs can be found in **Fig. 3a**, **Fig. 3b**, **Supplementary Fig. 13** and **Supplementary Fig. 15**.

We used the t-test to investigate the statistical significances of the differences between different groups of ER-αBS with regards to functional marks such as H3K4me3, and RNAPII. We used the binomial test to assess the statistical significances of the differences between different groups of ER-αBS with regards to functional marks such as FoxA1. Analyses can be found in **Supplementary Note 5**.

IX. Gene analyses

First, we defined whether ER-αBS are distal or proximal. To be conservative in our definition of what is proximal to a gene promoter, we defined that if an ER-αBS is within 5 Kb of a gene Transcription Start Site (TSS) from UCSC Genes, hg18[13], then the ER-αBS is "proximal", otherwise it is "distal".

Genes (UCSC Genes, hg18)[13] were assigned to complex and standalone duplex interactions (collectively called "interaction regions"). Some genes have multiple alternative transcripts and thus are reflected in the genome as different gene models (transcriptional units), which are each given a different unique gene ID. These different transcriptional units may share the same gene name, but can have different features, for example, some transcripts might have RNAPII marks but not others. In addition, some transcription start sites from a particular gene might be near the interactions but not other transcription start sites belonging to the same gene. In order to fully capture all features of all transcript units, and obtain the most accurate mapping of interactions to genes, we used all gene IDs as given in the UCSC Genes database. In the text, these different gene models which each have unique gene IDs as given by the UCSC Known Gene database, are called "transcriptional units".

If the TSS of a transcriptional unit was within ± 20 Kb of the middle of any anchor in an interaction unit, the associated gene was assigned as an **"anchor gene"** otherwise it was assigned as a **"loop gene"**. Here, 20 Kb was chosen for the gene assignment after performing manual inspection of several known direct ER-α target genes. We decided that 20 Kb would not be too large a region such that there would be a lot of noise from false positive genes, but would still be sufficient to include most known ER-α target genes.

If the transcriptional unit was within ± 20 Kb of the middle of any anchor in an interaction unit and also had the entire transcriptional unit (5' transcription start site to 3' transcription end site for that particular transcriptional unit) entirely wrapped up within interaction boundaries of the interaction unit, then the associated gene was called an **"enclosed anchor gene"**. Otherwise, if the gene was an anchor gene but not

classified as an "enclosed anchor gene" because none of the associated transcriptional units were entirely wrapped up within the interaction boundaries of the interaction unit, it was called a **"non-enclosed anchor gene"**. The gene was marked as up-regulated or down-regulated based on whether it showed such microarray expression probes. It was also marked as up-regulated or down-regulated, and the digital intensity was given, based on RNA-PET data in oestrogen-treated and untreated MCF-7 cells. The gene was marked as H3K4me3 associated if the promoter (1 Kb upstream and downstream of the gene transcription start site) had such a peak. Similarly, it was marked as RNAPII associated if the promoter (1 Kb upstream and downstream of the gene transcription start site) had such a peak. A few genes may have multiple transcriptional units wherein one transcriptional unit is defined as "anchor" and one transcriptional unit is defined as "loop": this situation occurs when one transcriptional unit for one gene had a TSS within 20 Kb of an anchor, whereas another transcriptional unit for the same gene had a TSS that was not within 20 Kb of an anchor. All loop, non-enclosed anchor and enclosed anchor genes are given in **Supplementary Table 8**. A summary of the analyses is presented in **Supplementary Table 9**.

We also compared our data with "stand-alone" ER-αBS that do not have chromatin interactions. If the TSS of a transcriptional unit was within ± 20 Kb of the middle of the ER-αBS, then the associated gene was assigned to the ER-αBS. This data was also analyzed and included in the summary in **Supplementary Table 9.**

In statistical testing, we compared different classes of genes to see if there were any differences in the gene characteristics using Fisher's Exact Tests. More details can be found in **Supplementary Note 6**.

Gene expression visualization and analysis. Gene transcriptional units in different categories were clustered using *Cluster* version 2.11 (http://rana.lbl.gov/eisen/?page_id=42) and visualized using *TreeView* version 1.60 (November 2002) (http://rana.lbl.gov/eisen/?page_id=42)[6]. If two or more probes could be assigned to the same transcriptional unit, one probe was chosen randomly. The results are shown in **Fig. 4b** and **Supplementary Fig. 18**.

X. Validation experiments

Selected ChIP binding sites were subjected to validation analyses by ChIP-qPCR (**Supplementary Fig. 7**). Selected sites of chromatin interactions identified by ChIA-PET data were subjected to validation analyses using a variety of methods including ChIP-3C (**Supplementary Fig. 5** and **Supplementary Fig. 8**), 3C (**Fig. 1b; Fig. 5; Supplementary Fig. 9**), 4C (**Fig. 1c**; **Supplementary Fig. 10**), and FISH (**Fig. 1d**; **Supplementary Fig. 11** and **Supplementary Table 4**). All validated interaction sites are listed in **Supplementary Table 5**. Gene activation and gene repression marks were validated by RT-qPCR (**Fig. 5** and **Supplementary Fig. 8**). All oligonucleotide sequences of primers used in validation experiments are listed in **Supplementary Table 10**. BAC clones used in FISH experiments are also included in **Supplementary Table 10**.

ChIP-qPCR. ChIP-qPCR experiments were performed against ER-α, unphosphorylated RNAPII (8WG16, Covance), and serine-5 phosphorylated RNAPII (ab5131, Abcam). ChIP material was prepared from MCF-7 cells induced with oestrogen for 45 min ("oestrogen-treated"), as well as negative control MCF-7 cells induced with an equal volume of ethanol for 45 min ("ethanol-treated"), as described earlier. ChIP material was reverse cross-linked under conditions of 1% SDS and 65°C, and purified using a PCR purification kit (Qiagen). Real-time PCR quantification was performed as described earlier. The control primer used was from Zhao et al., 2007[14]. All experiments were repeated at least twice. The serine-5 ChIP of RNAPII was also sequenced by Illumina GAII single-read sequencing, and examples of this data, as well as a scanning

ChIP-qPCR experiment, are shown in the *GREB1* locus (**Fig. 5**). All oligonucleotide sequences of primers used in validation experiments are listed in **Supplementary Table 10**.

Chromatin Immunoprecipitation Chromosome Conformation Capture (ChIP-3C). ChIP-3C was performed as described previously[15] with modifications. Briefly, chromatin immunoprecipitation was performed overnight as described in the ChIP protocol. Beads were then washed twice with PBS, and restriction enzyme digestion was performed overnight in 100 µl of 1x buffer at 37ºC with nutation (all from NEB). The beads were then spun down, and the buffer removed. A further restriction digest was performed with fresh buffer and enzyme at 37ºC for half a day. The beads were then spun down, and the buffer removed. The beads were then washed 3x with PBS, and ligation was performed using 1x ligation buffer and T4 DNA ligase (NEB) in 100 µl at 16ºC. A further ligation was performed by adding 100 µl of fresh buffer and enzyme to the mixture and incubating at 16ºC for half a day. 100 µl of Elution Buffer containing 1% SDS was then added to the beads, and the beads were incubated at 65ºC for at least 6 hours. The supernatant was purified with a PCR purification kit (Qiagen). Primers and restriction enzymes for the ChIP-3C procedure were chosen based on the ChIA-PET sequences. All primers and restriction enzymes had to be within a region of ±100-500 bp from the targeted ER-α binding site peak. Primers (1stBase) were designed using Primer3 software available from: http://frodo.wi.mit.edu/primer3/ [16]. PCR products were amplified with AccuPrime Taq High Fidelity DNA Polymerase (Invitrogen) using an MJ thermocycler (GMI). The PCR program used was (1) 94°C for 2 min, (2) 94°C for 30s, (3) 56-60°C for 40s, (4) 68°C for 40s (5) 68°C 5 min, (6) 4°C forever. Steps (2) to (4) were run for 35-47 cycles. PCR products were run on a 1% agarose gel with ethidium bromide. PCR products were sequenced to verify the long-range ligation product. Each validation experiment was repeated at least twice for confirmation. All oligonucleotide sequences of primers used in validation experiments are listed in **Supplementary Table 10**.

Chromosome Conformation Capture (3C). 3C was performed as described previously[17] with modifications for qPCR. Briefly, $1X10^7$ MCF-7 cells were treated with ethanol or E2 for 45mins and crosslinked with 1% formaldehyde for 10 min. The reaction was stopped by the addition of 125mM glycine for 5 min. Nuclei were resuspended in 500 µl of 1.2 x restriction enzyme buffer before incubation at 37°C for 1 hr with 7.5 µl of 20% SDS followed by an additional 1 hr incubation with 50 µl of 20% Triton X-100. Samples were then incubated with 400 units of restriction enzyme at 37°C overnight. After restriction enzyme digestion, 40 µl of 20% SDS was added to the digested nuclei and incubated at 65°C for 10 min. 6.125 ml of 1.15x ligation buffer and 375 µl 20% Triton X-100 was added to dilute the total DNA to a concentration of ~ 2.5ng/ul to favour intramolecular ligation. The diluted sample was incubated at 37°C for 1 hr before the addition of 2000 units of T4 DNA ligase (NEB) at 16°C for 4 hr. Samples were finally de-crosslinked at 65°C overnight before phenol-chloroform extraction and ethanol precipitation. Samples were further purified by QIAquick spin columns and total DNA concentration quantified using Nanodrop.

Gene locus for 3C experiments were chosen based on their ChIA-PET interactions. All primers had to be within a region of 25-150 bp from the restriction enzyme digestion site and are unidirectional from the 5' side of the restriction fragment. Primers were designed using Primer3 software available from: http://frodo.wi.mit.edu/primer3/

To compare signal intensities obtained with different primer sets in a quantitative manner, a control template containing all possible ligation products in equimolar amounts was used to correct for the PCR efficiency of each primer set. For this purpose, we used a BAC spanning each locus of interest and digested it with the corresponding restriction enzyme before ligation and column purification. The ligated control fragments

were diluted to appropriate concentration and mixed with genomic DNA to obtain a final working concentration of total DNA (~ 25-50 ng/μl) that was similar to that of the 3C templates.

The linear range of amplification for the 3C template was determined by serial dilution. An appropriate amount of DNA within the linear range was subsequently used for the experiments. The linear range of the control template was determined with a serial dilution of the random ligation mix made in the same amount. Titration curves for control and 3C templates were generated using two primers pair combinations (detecting the positive interaction and a negative interaction). A relative value for the amount of PCR products were calculated for each dilution point by expressing the 2^{-ct} of each point as a fraction of the positive interaction primer pair at its highest total DNA concentration. The relative values were then plotted against the total DNA concentration. For both primer pair combinations in each titration curve, the control template has similar relative values and actual Ct value between the primer pairs at each concentration do not differ by more than 1.5 cycles. Other primer pairs used for each locus also do not differ by more than 1.5 cycles at each given concentration within the linear range. The 3C template yields significantly more PCR products for the primer pair that detects for the positive interaction. The same template concentrations within the linear range were subsequently used for all repeats.

Quantitative real time PCR was carried out with SYBR green master mix on the ABI PRISM 7900. All dissociation curves for the primer pairs produced only a single peak in their melting curves with the control template. Semi-quantitative PCR of these primers pairs using the control template re-confirmed that there was only a single PCR product of the correct size when visualized on a 2% agarose gel. The dissociation curves of all 3C templates were checked against that of the confirmed control templates ran concurrently each time to ensure they give the same single peak at the correct melting temperature. 3C templates which did not give the same single peak were expressed as zero in the final results. The identities of the PCR products were also confirmed through direct sequencing. To obtain data points for **"normalized relative interaction"** in the final results, Ct values of 3C template were first normalized with values from an internal primer to account for quantity. Next, the values were normalized with values for each corresponding primer pair to account for relative primer efficiency. A similar method of normalization was previously used[18].

The equation used was: Normalized Relative Interaction: $2^{-\Delta\Delta Ct} = [(Ct_{\text{interaction}} - Ct_{\text{internal control}})\text{3C template} - (Ct_{\text{interaction}} - Ct_{\text{internal control}})\text{ BAC template}]$

Each qPCR was carried out in duplicates and 3C validations were repeated between four to six times independently for each locus. 3C results are shown in **Fig. 1b, Fig. 5** and **Supplementary Fig. 9**. All oligonucleotide sequences of primers used in validation experiments are listed in **Supplementary Table 10**.

Circular Chromosome Conformation Capture (4C). We developed a new sonication-based method for performing Circular Chromosome Conformation Capture (4C)[19]. Briefly, MCF-7 cells were treated as mentioned in the ChIP protocol up to the crosslinking step with 1% formaldehyde. An additional centrifugation step was performed to further clarify the supernatant by removing cellular debris. Aliquots were removed and diluted 10 times with Tris-HCl buffer (Qiagen, Buffer EB) containing 1x Protease Inhibitor Cocktail (Roche). The chromatin was incubated for 1h at 37°C. 1 % (final concentration) Triton X-100 was added and the chromatin material was allowed to stand for a further hour at 37°C. End-blunting was performed at room temperature for 45 min, using the End-It DNA End-Repair Kit (Epicentre). The chromatin samples were diluted to 10 ml with sterile water containing 1 x Complete Protease Inhibitor Cocktail, and we performed ligation by adding 1000 units of T4 DNA ligase (Fermentas) and letting the reaction stand at 16°C overnight. 0.15 μg/μl (final concentration) of Proteinase K (Invitrogen) was added,

and the chromatin material was reverse cross-linked at 65°C overnight. The DNA was purified by phenol extraction and isopropanol precipitation, and treated with RNase A (Qiagen) at 37°C for 30 min. Non-circularized DNA was digested away by incubation with Plasmid-safe DNase (Epicentre) at 37°C overnight, and the DNA was re-purified by phenol extraction and isopropanol precipitation.

At least 100 ng of DNA template was used for the PCR reactions. The DNA samples were amplified using nested inverse PCR. Primers (1st Base) had to be within 100 bp of the targeted ER-α binding site peak and were designed using Primer3 software available from: http://frodo.wi.mit.edu/primer3/ [16]. The RepeatMasker track[9] in the UCSC Genome Browser (http://genome.ucsc.edu/)[20] was used to ensure that the primers did not lie in repeat regions. An MJ thermocyler (GMI) and the high-fidelity DNA polymerase Phusion (Finnzymes) were used for the PCR reactions. The PCR program used for first-round amplification was: (1) 98°C for 30 s; (2) 25 cycles of 98°C for 10 s, 70°C for 30 s and 72°C for 30 s; (3) 72°C for 10 min; and (4) 4°C forever. The PCR program used for second-round amplification was: (1) 98°C for 30 s; (2) 25 cycles of 98°C for 10 s and 72°C for 1 min; (3) 72°C for 10 min; and (4) 4°C forever. 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.

Roche 454 GSFLX sequencing generated 0.5 million sequences. All the sequences were mapped to reference genome to identify the target regions in relation to the bait region and were filtered to remove redundant clonal amplifications (repeated sequences). The majority of the sequences were either mapped randomly along the genome as potential non-specific 4C products, or mapped to the "bait" region within 1 Kb, suggesting self-ligation products in the 4C experiment. To validate the ChIA-PET data, we specifically looked at the KRT gene cluster site where the "bait" region is. Moving right from the "bait" region, we identified overlapping sequence clusters that correlated very well with the locations of the interaction sites identified by ChIA-PET data. The 4C data showed very clean background. From the bait region to the first and the second interaction sites (about 200 Kb distance), there are no background sequences in the intervals. Considering that we used sonication method for preparing the chromatin materials for our 4C analysis, which is different from the standard 3C and 4C protocols, this result suggests that the sonication method is very efficient in "shaking off" non-specific chromatin fragments randomly attached to the specific chromatin interaction complexes. We expect to obtain discrete interaction peaks formed by clusters of inter-ligation sequences from sonicated material, because we expect real interactions to been captured by proximity ligation process. While the detached non-specific chromatin fragments would still be present in the DNA pool, they would not be amplified by the 4C PCR detection method.

All interactions found (interactions are defined as having 2 or more overlapping unique PETs) are shown in the figures. There were no other interactions detected.

The 4C result is presented in **Fig. 1c** and **Supplementary Fig. 10**. All oligonucleotide sequences of primers used in validation experiments are listed in **Supplementary Table 10**.

Fluorescence in-situ hybridization (FISH). MCF-7 nuclei were harvested by treating cells with 0.75 M KCl for 20 min at 37°C. The cells were fixed in Methanol/Acetic acid (3/1), and nuclei were dropped on slides for FISH. Following overnight culture in LB media, BAC DNA was extracted with Nucleobond PC500 (Macherey-Nagel), and then labeled by nick translation in the presence of biotin-16-dUTP or digoxigenin-11-dUTP using Nick translation system (Invitrogen). In presence of 1µg/µl of Cot1DNA (Invitrogen), DNAs BAC clones were resuspended at a concentration of 5 ng/µl in hybridization buffer (2xSSC, 10% dextran sulfate, 1X PBS, 50% formamide). Prior to hybridization, MCF-7 nuclei slides were treated with proteinase

K (Sigma) at 37°C for 2 min followed by 2 1xPBS rinses (5 min at room temperature) and dehydratation through ethanol series (70%, 80% and 100%). Denaturated probes were applied to these pretreated slides and codenaturated at 75°C for 5min and hybridized at 37°C overnight. Two posthybridization washes were performed at 45°C in 2xSSC/50% formamide for 7 min each followed by 2 washes in 2xSSC at 45°C for 7 min each. After blocking, the slides were revealed with avidin-conjugated fluorescein isothiocyanate (FITC) (Vector Laboratories, CA) for biotinylated probes and anti-digoxigenin- Rhodamine for digoxigenin-labeled probes (Roche). After washing, slides were mounted with vectashield (Vector Laboratories, CA) and observed under an epifluorescence microscope (Nikon). Between 100-200 interphase nuclei were analyzed for each mix of probes. Fusion and colocalization spots were counted in each nucleus. Results are tabulated in **Supplementary Table 4** and are presented in **Fig. 1d** and **Supplementary Fig. 11**. More details about the interactions are given in **Supplementary Table 5**. BAC clones used in FISH experiments are also included in **Supplementary Table 10**.

Fisher's exact test was used to evaluate whether the number of fusions were significantly higher when comparing the various types of cells, in both NR2F2 and GATA3.

NR2F2

For FISH studies, we chose one of the longest intrachromosomal interaction complexes, chr15:93128663-94685818, which is about 1.5 Mb in genomic span. This interaction involves many genes, including *NR2F2*, *AK000872*, *AK307134*, *AK057337*, and *BC040875*. For convenience, we refer to this interaction as the "*NR2F2* interaction". BAC probes P1, P2, and P3 were chosen from the list of available BACs from CHORI. P1 and P2 span a region of about 756 Kb, and do not involve interactions. This is the "negative control" region. P2 and P3 span a region of about 966 Kb, and involve interactions. This is the "experimental" region. More details can be found in **Supplementary Tables 4** and **5**. The results are shown in **Fig. 1d**.

Comparing control probes (P1/P2) with experimental probes (P2/P3) in ethanol-treated (ET) cells, there is a very significant (2-tailed p-value = $2.4e^{-14}$) enrichment in the number of fusions when experimental probes are used, indicating the interaction is present in ethanol-treated cells. Comparing control probes (P1/P2) with experimental probes (P2/P3) in oestrogen-treated (E2) cells, there is an extremely significant (2-tailed p-value = $3.3e^{-59}$) enrichment in the number of fusions when experimental probes are used, indicating the interaction is present in oestrogen-treated cells. Comparing control probes (P1/P2) in ethanol-treated (ET) cells with control probes (P1/P2) in oestrogen-treated (E2) cells, there is a very weakly significant difference between the two datasets (2-tailed p-value = 0.044). The control site is therefore weakly oestrogen-dependent. By contrast, comparing experimental probes (P2/P3) in ethanol-treated (ET) cells with control probes (P2/P3) in oestrogen-treated (E2) cells, there is a significant difference between the two datasets (2-tailed p-value = $9.9e^{-12}$). The experimental site is therefore strongly oestrogen-dependent – that is, the interaction is present in more of the oestrogen-treated cells than the ethanol-treated cells.

GATA3

We also chose another intrachromosomal interaction complexes, chr10:8024711-9274086, which is about 1.2 Mb in genomic span. While the automated interaction complex clustering does not take into account inter-ligation singletons, if inter-ligation singletons on the edges are included, the interaction is a bit bigger, on the order of 1.6 Mb. This interaction involves many genes, including *GATA3*. For convenience, we refer to this interaction as the "*GATA3* interaction". BAC probes P1, P2, and P3 were chosen from the list of available BACs from CHORI. P1 and P2 span a region of about 1.6 Mb, and do not involve

interactions. This is the "negative control" region. P2 and P3 span a region of about 1.6 Mb, and involve interactions. This is the "experimental" region. More details can be found in **Supplementary Tables 4** and **5**. The results are shown in **Supplementary Fig. 11a**.

Comparing control probes (P1/P2) with experimental probes (P2/P3) in ethanol-treated (ET) cells, there is a very significant (2-tailed *p*-value = $2.5e^{-9}$) enrichment in the number of fusions when experimental probes are used, indicating the interaction is present in ethanol-treated cells. Comparing control probes (P1/P2) with experimental probes (P2/P3) in oestrogen-treated (E2) cells, there is an extremely significant (2-tailed *p*-value = $6.8e^{-19}$) enrichment in the number of fusions when experimental probes are used, indicating the interaction is present in oestrogen-treated cells. Comparing control probes (P1/P2) in ethanol-treated (ET) cells with control probes (P1/P2) in oestrogen-treated (E2) cells, there is an insignificant difference between the two datasets (2-tailed *p*-value = 0.56). The control site is therefore not oestrogen-dependent. By contrast, comparing experimental probes (P2/P3) in ethanol-treated (ET) cells with control probes (P2/P3) in oestrogen-treated (E2) cells, there is a significant difference between the two datasets (2-tailed *p*-value = 0.00046). The experimental site is therefore oestrogen-dependent – that is, the interaction is present in more of the oestrogen-treated cells than the ethanol-treated cells.

Interchromosomal interactions

Additionally, we chose 2 FISH interactions from IHM001F on the basis that interactions with ER-αBS on both sides are likely to be real. These interactions are: chr7:2692293-2698011and chr12:120670670-120673138, as well as chr9:113812545-113817679 and chr14:90811137-90814459. These interactions have PET counts of 2, and are not reproducible in IHH015F. As we chose interactions to test by FISH validation early in the analysis and sequencing process, when the library sequencing depth was lower, we did not manage to pick better candidates with PET counts of 3 and ER-αBS on both sides. Further work could be done to explore the interchromosomals better, especially interchromosomal interactions with 3 or more PETs and ER-αBS on both sides. However, we believe that our existing FISH data confirms that the vast majority of interchromosomal interactions are most likely to be noise. The counting results are shown in **Supplementary Table 4**. No significance testing was done as the overlap percentages were extremely low. Some examples of microscopic views are shown in **Supplementary Fig. 11b**.

We also tested the *TFF1-GREB1* interaction that had been previously shown in a different paper[21]. While we could neither detect the interaction in ChIA-PET nor validate it by FISH, one explanation could be that different cellular treatments were performed on the MCF-7 cells.

More details on these interactions can be found in **Supplementary Table 5**.

RT-qPCR. Total RNA was prepared from MCF-7 cells induced with oestrogen for 0, 3, 6, 12 and 24 hours using an RNA purification kit (Qiagen), following the manufacturer's protocols. 1 μg of total RNA was incubated with 50 ng of random primer (Roche) at 70ºC for 10 min and then cooled on ice for 1 min. To the mixture, first strand buffer (Clontech) was added to a final concentration of 1x, DTT (Clontech) was added to 0.01 M, dNTP mix (Invitrogen) was added to 1 mM, and 1 μl of Powerscript RT enzyme (Clontech) was added. The mixture was heated to 42ºC for 90 min, and heat inactivated at 70ºC for 15 min. Real-time quantitative PCR was performed using an ABI Real-time PCR 7500 system. PCR was performed with a 10 μl reaction volume consisting of substrate, 0.5 μM of primer pairs (1stBase) and 1x SYBR Green PCR Master Mix (ABI). Reactions were incubated at 95ºC for 10 min, and then 40 cycles (95ºC for 15 s, 60ºC for 1 min) were carried out. Fluorescence was acquired at the end of each cycle at 60°C during the amplification step. The control pair of primers used was that of 36B4 (ribosomal protein mRNA). All experiments were

repeated at least twice. Results are shown in **Fig. 4c**, **Fig. 5**, and **Supplementary Fig. 8**. All oligonucleotide sequences of primers used in validation experiments are listed in **Supplementary Table 10**.

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 ER-α 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 with TRI® Reagent (Sigma) and purified using QIAGEN RNeasy. The RNA was reverse transcribed with oligo $(dT)_{15}$ primer (Promega), dNTP Mix, and M-MLV RT (Promega). Real-time PCR quantification was performed as described earlier. All experiments were repeated at least twice. Results are shown in **Fig. 5**.

Supplementary Notes

Supplementary Note 1. Comparison of ChIA-PET pilot libraries with control libraries

Analyses of one ER-α ChIA-PET pilot library dataset, the IHM001H ChIA-PET dataset, which consists of 293,754 uniquely mapped PET sequences, revealed 103,740 self-ligations, 21,521 of which overlapped to give 3,405 binding sites (FDR < 0.01) with satellites and structural variation regions filtered out, as well as 17,718 intrachromosomal inter-ligation PETs, 585 of which overlapped to give 215 putative intrachromosomal interactions containing 2 or more overlapping PETs (FDR < 0.05) with satellites and structural variation regions filtered out, and 172,296 interchromosomal inter-ligation PETs, 0 of which overlapped to give interactions with FDR < 0.05.

Analyses of another ER-α ChIA-PET pilot library dataset, the IHM001N ChIA-PET dataset, which consists of 271,648 uniquely mapped PET sequences, revealed 78,706 self-ligations, 15,888 of which overlapped to give 2,701 binding sites (FDR < 0.01) with satellites and structural variation regions filtered out, as well as 16,677 intrachromosomal inter-ligation PETs, 463 of which overlapped to give 176 putative intrachromosomal interactions containing 2 or more overlapping PETs (FDR < 0.05) with satellites and structural variation regions filtered out, and 176,265 interchromosomal inter-ligation PETs, 0 of which overlapped to give interactions with FDR < 0.05.

Analyses of one control pilot library dataset, the IHM043 ChIP-PET dataset (reverse cross-linking performed before ligation, and hence should only show interactions that are a result of structural variations that bring together two different regions into one ChIP-enriched region), which consists of 745,251 uniquely mapped PET sequences, revealed 634,993 self-ligations, 8,538 of which overlapped to give 1,158 binding sites (FDR < 0.01) with satellites and structural variation regions filtered out, as well as 7,386 intrachromosomal inter-ligation PETs and 102,872 interchromosomal inter-ligation PETs. Putative "interactions" were searched for using the same ChIA-PET methodology: 2 or more overlapping PETs, FDR < 0.05, and with satellites and structural variation regions removed. 4 intrachromosomal inter-ligation PETs overlapped to give 2 putative intrachromosomal interactions containing 2 overlapping PETs (FDR < 0.05) with satellites and structural variation regions filtered out. Analysis of these 2 "interactions" indicated that one 1 interaction has a genomic span that was less than 5 Kb, suggesting it could be a result of extra-long self-ligation PETs, and the other has a genomic span of over 10 Mb, suggesting it could be non-specific. 2 interchromosomal inter-ligation PETs overlapped to give 1 interchromosomal "interaction".

Analyses of the previously generated SHC007 ChIP-PET dataset[1], which was reprocessed using the ChIA-PET analysis pipeline (reverse cross-linking performed before ligation, and hence should only show interactions that are a result of structural variations that bring together two different regions into one ChIP-enriched region), which consists of 214,668 uniquely mapped PET sequences, revealed 192,511 self-ligations, 2,980 of which overlapped to give 489 binding sites (FDR < 0.01) with satellites and structural variation regions filtered out, as well as 2,196 intrachromosomal inter-ligation PETs and 19,691 interchromosomal inter-ligation PETs. Putative "interactions" were searched for using the same ChIA-PET methodology: 2 or more overlapping PETs, FDR < 0.05, and with satellites and structural variation regions removed. 0 intrachromosomal and interchromosomal interactions were found.

Analyses of one control pilot library dataset, the IgG ChIA-PET dataset (IgG binds nonspecifically to the genome, and a small-scale dataset is not expected to reveal any chromatin interactions as the complexity of mock-ChIP-enriched ChIA-PET datasets is expected to be very high), which consists of 217,708 uniquely mapped PET sequences, revealed 40,847 self-ligations, 0 binding sites (FDR < 0.01), as well as 11,254 intrachromosomal inter-ligation PETs, 0 of which overlapped to give putative intrachromosomal interactions containing 2 or more overlapping PETs and FDR < 0.05. Also, the library contained 165,607 interchromosomal inter-ligation PETs, 0 of which overlapped to give putative intrachromosomal interactions containing 2 or more overlapping PETs and FDR < 0.05.

Taken together, these analyses indicate that the prevalent chromatin interactions identified by ER-α ChIA-PET data are not due to technical artifacts of random ligations or false positives as a result of mapping errors and structural variations in MCF-7.

Reference data can be found in **Table 1**.

Supplementary Note 2. Analysis of chimeric ligations

A major concern in proximity ligation-based analyses such as 3C and 4C for studying chromatin interactions is the level of non-specific chimeric ligations between different chromatin fragment complexes during the proximity ligation reaction. This concern is further amplified if ChIP is used to enrich specific protein-bound chromatin interaction nodes. A particular issue here is how to measure the chimeric ligations and what impact such non-specific ligations will have on our data analysis. To specifically identify cases of non-specific random ligation, we took advantage of the ChIA-PET method's linker sequence. This linker sequence is used to connect proximity ligated DNA fragments. We designed two different linkers (A and B) with specific nucleotide barcodes (CG or AT) for each of the two linker sequences. As illustrated in **Supplementary Fig. 4**, this design can be used for investigating chimeric ligation rates between different chromatin complexes. One linker can be ligated to one aliquot, and another linker can be ligated to another aliquot of the same chromatin samples. After linker ligation and removal of free linkers, the two aliquots are mixed together for proximity ligation. The ligation products are analyzed by PET sequencing. The PET sequences with AA (CG/CG) and BB (AT/AT) linker barcode composition are considered to be possible intra-complex ligation products, while the PET sequences with AB (CG/AT) linker composition are considered to be derived from chimeric ligation products between DNA fragments bounded in different chromatin complexes.

In a separate biological repeat of the ER-α ChIA-PET experiment (the IHH015 set of experiments), we created a library with A and B linker sequences. Despite the use of large volumes (10 ml) for the dilute proximity chromatin ligation reaction, 454 pyrosequencing analysis of this particular library revealed that for this particular proximity ligation experiment, the chimeric AB linker PET sequences are 39%, while the AA

and BB linker PET sequences are 61%. To specifically and equally compare the AB linker PET sequences with the AA and BB linker PET sequences for their mapping characteristics, we collected a subset of PETs with uniquely mapped AB linker PET sequences (IHH015C; 1.8 million PETs), and created an AA and BB linker PET dataset that has almost equal numbers of PETs (IHH015M; 2.0 million PETs) for comparison.

Analyses of the AA and BB linker PET dataset, IHH015M, which consists of 2,049,719 uniquely mapped PET sequences, revealed 953,384 self-ligations, 60,568 of which overlapped to give 3,921 binding sites (FDR < 0.01) with satellites and structural variation regions filtered out, as well as 129,492 intrachromosomal inter-ligation PETs, 18,426 of which overlapped to give 2,186 putative intrachromosomal interactions containing 3 or more overlapping PETs (FDR < 0.05) with satellites and structural variation regions filtered out, and 966,843 interchromosomal inter-ligation PETs, 458 of which overlapped to give 46 interchromosomal interactions containing 3 or more overlapping PETs (FDR<0.05) with satellites and structural variation regions filtered out. Manual curation on the interchromosomal interactions indicated that the majority were in amplicon regions, and after curation to remove interactions that involve amplicon regions[10], 3 were left.

By contrast, analyses of the chimeric AB linker PET dataset revealed 15,490 self-ligations, 144 of which overlapped to give 35 binding sites (FDR<0.01) with satellites and structural variation regions filtered out, as well as 98,805 intrachromosomal inter-ligation PETs, and 1,676,419 interchromosomal inter-ligation PETs. However, after satellites and structural variation regions were filtered out, 0 interactions were left.

One finding from this analysis is that the AA and BB PET dataset showed that 47% are self-ligations, while in the AB PET dataset, only 0.9% are self-ligations. This data is understandable, considering that the vast majority of self-ligation should occur within fragmented chromatin complexes, rather than between complexes as chimeric ligation noise. Therefore, the self-ligation percentage is an indication that the AA and BB PETs consist of many intra-complex ligation products while the AB PETs are mostly inter-complex ligation products. We also observed that the vast majority of AB PETs are inter-chromosomal (94%), while only 47% of AA and BB PETs are inter-chromosomal. This ratio of extremely low self-ligations and high inter-chromosomals indicates that most AB PETs are derived from chimeric ligation products, supporting our design for identifying chimeric ligations using the nucleotide barcode scheme.

Hence, these results validate our view that chimeric ligations tend to distribute randomly throughout the genome, and do not cluster together often. Taken together, the linker barcode data provided convincing evidence that the chimeric ligation products, although unavoidable, would not lead to high levels of false positives because of our data analysis method for identifying true chromatin interactions (which are indicated by clusters of multiple inter-ligation PETs). In other words, the prevalent chromatin interactions identified by ER-α ChIA-PET data are not due to technical artifact of random ligations, and most likely are due to proximity ligations of DNA fragments tethered together in same chromatin complexes.

Reference data can be found in **Table 1**.

Supplementary Note 3. Interchromosomal inter-ligation PET clusters in ChIA-PET data

In IHM001F, only 17 of 28 (61%) interchromosomal inter-ligation PET clusters (3 or more overlapping inter-ligation PETs) were not in amplicon regions[10] (as opposed to 1,244 of 1,451= 86% intrachromosomal interactions in IHM001F). In IHH015F, only 7 of 28 (25%) inter-chromosomal inter-ligation PET clusters were not in amplicon regions (as opposed to 3,187 of 3,543 = 90% intrachromosomal interactions in IHH015F).

The remaining interchromosomal inter-ligation PET clusters that were not in amplicon regions were pooled, and of these, we removed interchromosomal inter-ligation PET clusters appeared to be within chromosomal aberration regions that fell below the threshold used in the first-pass removal of interactions, or appeared to be the result of mapping to duplicated regions in the genome.

Of the remaining 20 interchromosomal inter-ligation PET clusters (Supplementary Table 2), 15 had inter-ligation PET counts of 3, 2 had inter-ligation PET counts of 4, and 3 had inter-ligation PET counts of 5, 6, and 8. Only 2 of 20 interchromosomal inter-ligation PET clusters (10%) had 1 binding site on each of the two anchor regions (that is, 2 binding sites). By contrast, 1338 of 1451 = 92% intrachromosomal interactions in IHM001F had 1 binding site on each of the two anchor regions and 2268 of 3543 = 64% intrachromosomal interactions in IHH015F had 1 binding site on each of the two anchor regions. It should be noted that no putative interaction was reproducible.

On the assumption that interactions with binding site support are likely to be real, we chose 2 putative interchromosomal clusters from IHM001F that had good binding sites on both sides of the interaction for FISH testing (**Supplementary Table 5**). Because we chose sites for FISH validation early in the analysis/sequencing process when the sequencing depth was not so high, we missed some interactions that are PET-3 with 1 binding site on each of the two anchor regions. Both putative interchromosomal clusters had inter-ligation PET counts of 2, and were not reproduced in IHH015F. The interactions tested were: Chr7:2693519-2696330 and Chr12:120670670-120672906, as well as Chr9:113813488-113817420 and Chr14:90810438-90813466. Neither of the 2 putative interchromosomal clusters could be validated, giving further support to the notion that the weak, poorly reproduced interactions found by ChIA-PET on ER-α are noise. We also tested the previously identified TFF1-GREB1 interaction, but could not validate it by FISH nor find it in our ChIA-PET data, possibly due to different cell treatments[21].

All interchromosomal PET clusters with 2 or more inter-ligation PETs from IHH015F and IHM001F are listed in **Supplementary Table 2**.

It should be noted that interchromosomals are much more likely to be noise than intrachromosomals as a result of the nature of interchromosomal identification: A random PET has a low chance of giving rise to an intrachromosomal, but a higher (21 times, because there are 21 more chromosomes that the random mapping could go to) chance of giving rise to an interchromosomal.

Supplementary Note 4. IHH015F library data statistics

We generated a large ER-α ChIA-PET dataset with 20 million (IHH0015F) PET sequences using Illumina GAII paired end sequencing (**Table 1**; **Supplementary Methods**) for comprehensive analysis of ER-α binding and chromatin interactions in oestrogen treated MCF-7 cells. Of the uniquely mapped PET sequences (6.1 million), 1.8 million PETs (30%) were self-ligation PETs. 6.2% self-ligation PETs formed overlapping PET groups, 6,665 putative ER-αBS (FDR< 0.01, PET count per ER-αBS ≥5, **Supplementary Table 1**). Of the inter-ligation PETs, 0.35 million (5.7% of uniquely aligned PETs) were intrachromosomal and 3.9 million (64%) 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 Methods**), we identified a large set of 3,543 intrachromosomal and a small set of 4 interchromosomal overlapping clusters consisting of 3 or more inter-

ligation PETs (FDR < 0.05). These represent paired inter-ligating ChIP fragments which indicate potential distant chromatin interactions bound by ER-α (**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 (2,644/3,826=69%)-involve self-ligation PET-defined ER-αBS (FDR< 0.01).

Of the 6,665 putative ER-αBS (FDR< 0.01, PET count per ER-αBS ≥5), 4,552 (PET count 5-49) are considered low-enrichment, and 2,113 (PET count ≥50) are high-enrichment ER-αBS. The high-enrichment ER-αBS are much more reliable than the low-enrichment sites (**Supplementary Fig. 13**), and are much more frequently involved in interactions (79% are associated with interactions) than the low-enrichment ER-αBS (only 29%). These results suggest that high-confidence and strong ER-αBS are more likely to be involved in chromatin interactions than weaker ER-αBS.
To further understand ER-αBS in relation to ER-α target genes, we analyzed how many ER-αBS are proximal or distal to gene promoters, based on a cut-off of 5 Kb from Transcription Start Sites (TSS) of UCSC Known Gene database. Of the 2,987 ER-αBS involved in chromatin interactions, 515 (17%) are proximal and 2,472 (83%) are distal to TSS (**Supplementary Fig. 14**). We also observed the same ratio for the ER-αBS not involved in interactions: 668 (18%) are proximal and 3,010 (82%) are distal. Therefore, the vast majority of ER-αBS are distal to gene TSS, in agreement with previous studies[1,2,22].

The ER-αBS were highly reproducible, especially in the top 100 strongest ER-αBS in IHH015F, which showed 99% overlap with IHM001F (**Supplementary Table 6**).

Many putative intrachromosomal interaction sites were reproducible, particularly in the top 100 of abundant chromatin interactions in IHH015F, 86 of which could be found in IHM001F (**Supplementary Table 7**). It should be noted that we found very few putative interchromosomal interactions, despite the fact that the majority of inter-ligation PETs were interchromosomal. This indicates that while the ChIA-PET method can accurately map interchromosomal inter-ligation PETs, in the case of ER-α ChIA-PET analysis, they appear to be noise. Moreover, none of the putative interchromosomal interactions observed in one library were reproducible in the other library, and hence likely to be either very weak interactions or noise (**Supplementary Note 3**). As such, ER-α appears to function by primarily intrachromosomal mechanisms, and therefore, we focused on intrachromosomal interactions for further analyses.

3,120 duplex interactions were further assembled into 519 complex interactions in IHH015F. The remaining interactions (423) are stand-alone duplex interactions. Collectively, we identified 942 ER-α-bound chromatin interaction regions from IHH015F (**Supplementary Table 3**). Many chromatin interaction regions were reproducible, particularly in the top 100 of abundant chromatin interactions in IHH015, 95 of which could be found in IHM001F (**Supplementary Table 7**).

The genomic spans of most stand-alone duplex interactions (86%) are less than 100 Kb, 14% are in the range of 100 Kb to 1 Mb, and none (0%) are over 1 Mb. Complex interactions extend genomic span by connecting multiple duplex interactions. Hence, most complex interactions (61%) have genomic spans in the range of 100 Kb to 1 Mb, with a few that are over 1 Mb **(Supplementary Fig. 12**). Taken together, the ER-αBS and chromatin interactions identified by ChIA-PET data constitute a whole genome chromatin interaction map bound by ER-α protein.

FoxA1 binding sites are significantly enriched in association with ER-αBS involved in complex- and duplex-interactions as compared to ER-αBS with no-interactions (1-tailed *p*-value = $4.9e^{-10}$; **Fig. 3a** and **Supplementary Fig. 15**), suggesting that FoxA1 could be involved in ER-α-bound chromatin interactions.

Next, we analyzed how many ER-αBS are proximal or distal to gene promoters, based on a cut-off of 5 Kb from the UCSC Known Gene database gene Transcription Start Sites (TSS). We identified 515 (17%) of interaction-associated ER-αBS that are proximal and 2,472 (83%) that are distal. We also identified 668 (18%) of non-interaction associated ER-αBS that are proximal and 3,010 (82%) that are distal. The vast majority of ER-αBS are distal to TSS, in agreement with previous studies.

Subsequently, we examined the 942 ER-α-bound chromatin interaction regions with respect to the looping structure and the function for gene transcription. We envisage that multiple ER-αBS may function as "anchor" regions generating chromatin looping structures in 3-dimensional space (**Fig. 4a**). Genes in proximity 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 known gene database entries[13]. 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**), an optimized parameter that includes many known and validated ER-α target genes. Most interaction regions (607/942=64%) were associated with "anchor genes". 285 of 942 (30%) interaction regions were associated with at least one "loop gene" and interaction regions with longer spans tended to involve more "loop genes". Altogether, 9,547 genes (**Supplementary Tables 3** and **8**) were assigned, including 3,553 "anchor genes" (TSS to interaction anchor within 20 Kb) and 5,994 "loop genes" (TSS >20 Kb away from interaction anchors). We found 274 interaction regions did not have any "anchor" or "loop genes", which we hypothesize could have a structural or different function. Using the same distance parameter (20 Kb), we also assigned 3,925 genes to 3,678 stand-alone ER-αBS that are not involved in interactions.

Within the interaction regions that have at least one anchor gene, there are 1,757 distal ER-αBS and 515 proximal ER-αBS (< 5 Kb to TSS); and all distal ER-αBS are connected to anchor genes. ChIA-PET can thus help to reduce ambiguity in gene assignments to binding sites (examples in **Supplementary Fig. 17**).

In addition, we found 607 of 942 interactions regions had anchor genes associated. Hence, there are 335 interaction regions with no associated anchor genes in IHH015F. While 61 have loop genes associated, the remaining 274 are located in poorly annotated gene desert areas, and so have no associated UCSC Genes assigned to them.

We compared whether anchor genes showed different characteristics from genes associated with stand-alone binding sites. Interestingly, we found that genes associated with interaction-associated ER-αBS are highly up-regulated even at early time points (3h, 6h), suggesting interaction-associated ER-αBS employ looping for early transcriptional effects. By contrast, genes associated with stand-alone ER-αBS do not appear to be active at earlier time-points, but are associated with up-regulation at late time points. We hypothesize that stand-alone ER-αBS, which could have lower levels of ER-α proteins, could require a secondary co-activator for late transcriptional effects **(Supplementary Fig. 20)**.

We investigated if chromatin interactions are functionally involved in transcriptional regulation. "Anchor genes" appeared to be associated with gene up-regulation as indicated by H3K4me3, RNAPII, and expression microarray up-regulated marks, whereas "loop genes" did not appear to be up-regulated as

indicated by these microarray analysis when compared with the background set of reference genes (**Supplementary Note 7**).

Intriguingly, within the anchor gene category, we found that the majority (1,828 of 3,553=52%) of anchor gene entries has 5' and 3' ends within the interaction boundaries. Such entries, called "enclosed anchor genes", frequently occupy the entirety of short interaction loops and are often found to engage multiple anchor sites within the gene structure as well. We observed that the "enclosed anchor genes" tend to have intense RNAPII marks covering the entire gene (data not shown), and are preferentially associated with RNAPII as well as gene up-regulation as assessed by expression microarrays (**Supplementary Note 7**).

Supplementary Note 5. Functional analyses of ER-αBS involved in interactions

We used the t-test to investigate the statistical significances of the differences between different groups of ER-αBS with regards to functional marks such as H3K4me3, and RNAPII. For each site we counted number of extended 200 bp ChIP-seq tags overlapping site and number of control background sequences (input DNA). We compared distributions of normalized numbers of ChIP-seq tags for each group of ER-αBS. We used Fisher's exact test to assess the statistical significances of the differences between different groups of ER-αBS with regards to functional marks such as FoxA1.

RNAPII

In IHM001F, the average normalized value of RNAPII ChIP tags was 2.18 in ER-αBS associated with complex-interactions, 1.64 in ER-αBS associated with duplex-interactions, and 1.30 in non-interacting ER-αBS. The difference between ER-αBS associated with complex-interactions and non-interacting ER-αBS was significant (*p*-value <1E-16) in IHM001F. The difference between duplex-interaction ER-αBS and no-interaction ER-αBS was significant (*p*-value=4.46E-10) in IHM001F. Moreover, the difference between ER-αBS associated with complex-interactions and duplex-interaction ER-αBS was significant (*p*-value=3.73E-08) in IHM001F.

In IHH015F, the average normalized value of RNAPII ChIP tags was 1.73 in ER-αBS associated with complex-interactions, 1.45 in ER-αBS associated with duplex-interactions, and 1.19 in non-interacting ER-αBS. The difference between ER-αBS associated with complex-interactions and non-interacting ER-αBS was significant (*p*-value=1.26E-26) in IHH015F. The difference between ER-αBS associated with duplex-interactions and non-interacting ER-αBS was significant (*p*-value=8.73E-04) in IHH015F. Moreover, the difference between ER-αBS associated with complex-interactions and ER-αBS associated with duplex-interactions was significant (*p*-value=1.24E-03) in IHH015F.

In IHM001F, the average normalized value of RNAPII ChIP tags was 3.31 in proximal, interacting ER-αBS, 1.71 in distal, interacting ER-αBS, 2.29 in proximal, non-interacting ER-αBS, and 1.10 in distal, non-interacting ER-αBS. Proximal ER-αBS were enriched by RNAPII in comparison to distal ER-αBS for interacting as well as non-interacting ER-αBS. The difference between proximal, interacting ER-αBS compared with proximal, non-interacting ER-αBS was significant (*p*-value=6.24E-10). The difference between distal, interacting ER-αBS compared with distal, non-interacting ER-αBS was strongly significant (*p*-value<1E-16**).** Hence, the enrichment of RNAPII in interacting ER-αBS was significantly higher than in non-interacting ER-αBS in IHM001F.

In IHH015F, the average normalized value of RNAPII ChIP tags was 2.71 in proximal, interacting ER-αBS, 2.17 in distal, interacting ER-αBS, 1.53 in proximal, non-interacting ER-αBS, and 1.02 in distal, non-interacting ER-αBS. The difference between proximal, interacting ER-αBS compared with proximal, non-

interacting ER-αBS was significant (*p*-value=2.14E-03). The difference between distal, interacting ER-αBS compared with distal, non-interacting ER-αBS was significant (*p*-value<1E-16**)**. Hence, the enrichment of RNAPII in interacting ER-αBS was significantly higher than in non-interacting ER-αBS in IHH015F.

H3K4me3

In IHM001F, the average normalized value of H3K4me3 ChIP tags was 7.35 in ER-αBS associated with complex-interactions, 6.71 in ER-αBS associated with duplex-interactions, and 7.02 in non-interacting ER-αBS. The difference between ER-αBS associated with complex-interactions and non-interacting ER-αBS was not significant (*p*-value= 5.03E-01) in IHM001F. The difference between ER-αBS associated with duplex-interactions and non-interacting ER-αBS was also not significant (*p*-value=6.14E-01) in IHM001F. Moreover, the difference between ER-αBS associated with complex-interactions and ER-αBS associated with duplex-interactions was not significant (*p*-value= 2.55E-01) in IHM001F.

In IHH015F, the average normalized value of H3K4me3 ChIP tags was 7.20 in ER-αBS associated with complex-interactions, 7.51 in ER-αBS associated with duplex-interactions, and 6.83 in non-interacting ER-αBS. The difference between ER-αBS associated with complex-interactions and non-interacting ER-αBS was not significant (*p*-value=0.43) in IHH015F. The difference between ER-αBS associated with duplex-interactions and non-interacting ER-αBS also was not significant by (*p*-value=0.42) in IHH015F. Moreover, the difference between ER-αBS associated with complex-interactions and ER-αBS associated with duplex-interactions was not significant (*p*-value=0.65) in IHH015F.

In IHM001F, the average normalized value of H3K4me3 ChIP tags was 16.50 in proximal, interacting ER-αBS, 5.24 in distal, interacting ER-αBS, 20.65 in proximal, non-interacting ER-αBS, and 4.26 in distal, non-interacting ER-αBS. Proximal ER-αBS were enriched by H3K4me3 marks in comparison to distal ER-αBS for interacting as well as for non-interacting category The difference between proximal, interacting ER-αBS compared with proximal, non-interacting ER-αBS was not significant (*p*-value=4.44E-02). The difference between distal, interacting ER-αBS compared with distal, non-interacting ER-αBS was significant (*p*-value=1.03E-09). Hence, overall, the enrichment of H3K4me3 in interacting ER-αBS was not significantly higher than in non-interacting ER-αBS in IHM001F.

In IHH015F, the average normalized value of H3K4me3 ChIP tags was 18.00 in proximal, interacting ER-αBS, 4.98 in distal, interacting ER-αBS, 21.61 in proximal, non-interacting ER-αBS, and 3.67 in distal, non-interacting ER-αBS. The difference between proximal, interacting ER-αBS compared with proximal, non-interacting ER-αBS was not significant (*p*-value=1.22E-01). The difference between distal, interacting ER-αBS compared with distal, non-interacting ER-αBS was significant (*p*-value=1.42E-14). Hence, overall, the enrichment of H3K4me3 in interacting ER-αBS was not significantly higher than in non-interacting ER-αBS in IHH015F.

FoxA1

In IHM001F, the fraction of FoxA1 ChIP-chip sites overlapping ER-αBS (in $\pm$200bp proximity) was 0.39 in ER-αBS associated with complex-interactions, 0.41 in ER-αBS associated with duplex-interactions, and 0.27 in non-interacting ER-αBS. The difference between ER-αBS associated with complex-interactions and non-interacting ER-αBS was significant (*p*-value=4.37E-21) in IHM001F. The difference between ER-αBS associated with duplex-interactions and non-interacting ER-αBS was significant (*p*-value=2.48E-18) in IHM001F. The difference between ER-αBS associated with complex-interactions and ER-αBS associated with duplex-interactions was not significant (*p*-value=4.11E-01) in IHM001F.

In IHH015F, the average normalized value of FoxA1 ChIP tags was 0.25 in ER-αBS associated with complex-interactions, 0.22 in ER-αBS associated with duplex-interactions, and 0.15 in non-interacting ER-αBS. The difference between ER-αBS associated with complex-interactions and non-interacting ER-αBS was significant (*p*-value=3.02E-22) in IHH015F. The difference between ER-αBS associated with duplex-interactions and non-interacting ER-αBS was significant (*p*-value=2.28E-05) in IHH015F. The difference between ER-αBS associated with complex-interactions and ER-αBS associated with duplex-interactions was significant (*p*-value=9.60e-02) in IHH015F.

In IHM001F, the fraction of FoxA1 ChIP-chip sites overlapping ER-α sites was 0.31 in proximal, interacting ER-αBS, 0.42 in distal, interacting ER-αBS, 0.20 in proximal, non-interacting ER-αBS, and 0.28 in distal, non-interacting ER-αBS. Overall, distal ER-αBS have higher fraction of FoxA1 sites. We compared interacting versus non-interacting sites separately. The difference between proximal, interacting ER-αBS compared with proximal, non-interacting ER-αBS was significant (*p*-value=4.87E-06). The difference between distal, interacting ER-αBS compared with distal, non-interacting ER-αBS was strongly significant (*p*-value=9.98E-31). Hence, the enrichment of FoxA1 in interacting ER-αBS was significantly higher than in non-interacting ER-αBS in IHM001F.

In IHH015F, the fraction of FoxA1 ChIP-chip sites overlapping ER-α sites was 0.19 in proximal, interacting ER-αBS, 0.25 in distal, interacting ER-αBS, 0.11 in proximal, non-interacting ER-αBS, and 0.16 in distal, non-interacting ER-αBS. The difference between proximal, interacting ER-αBS compared with proximal, non-interacting ER-αBS was significant (*p*-value=9.19E-05). The difference between distal, interacting ER-αBS compared with distal, non-interacting ER-αBS was significant (*p*-value=1.73E-18). Hence, the enrichment of FoxA1 in interacting ER-αBS was significantly higher than in non-interacting ER-αBS in IHH015F.

Taken together, these observations indicate that RNAPII and FoxA1, but not H3K4me3, could predict interactions at distal ER-αBS, and suggest that RNAPII and FoxA1 could participate in tethering chromatin interactions.

Supplementary Note 6. Numbers of genes associated with ER-α-bound interactions

We annotated the interaction regions in relation to UCSC known gene database transcripts (http://genome.ucsc.edu/)[13] . It should be noted that a gene may have multiple transcripts. Reporting transcript as opposed to genes could run the risk of giving an impression of a lot of genes. Hence, while in the text, we reported transcript numbers, here for the sake of completeness, we also report gene numbers.

IHM001F

Transcripts-based calculations

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 also assigned 11,790 genes to 12,126 stand-alone ER-αBS that are not involved in interactions. Within the anchor gene category, we defined (495 of 1,575 =31%) gene entries with 5’ and 3’ ends within interaction boundaries as “enclosed anchor genes”.

Gene-based calculations

Altogether, 807 “anchor genes” and 1,662 “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 also assigned 5,811 genes to 12,126 stand-alone ER-αBS that are not involved in interactions.

Within the anchor gene category, we defined (295 of 807 =37%) gene entries with 5' and 3' ends within interaction boundaries as "enclosed anchor genes".

IHH015F

Transcripts-based calculations

Altogether, 9,547 genes (**Supplementary Tables 3** and **8**) were assigned, including 3,553 "anchor genes" (TSS to interaction anchor within 20 Kb) and 5,994 "loop genes" (TSS >20 Kb away from interaction anchors). Using the same distance parameter (20 Kb), we also assigned 3,925 genes to 3,678 stand-alone ER-αBS that are not involved in interactions. Within the anchor gene category, we defined (1,828 of 3,553=52%) gene entries with 5' and 3' ends within interaction boundaries as "enclosed anchor genes".

Gene-based calculations

Altogether, 1,767 "anchor genes" and 2,756 "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 also assigned 1,799 genes to 3,678 stand-alone ER-αBS that are not involved in interactions. Within the anchor gene category, we defined (972 of 1,767 =55%) gene entries with 5' and 3' ends within interaction boundaries as "enclosed anchor genes".

Supplementary Note 7. Investigations into different functional gene categories

By examining the microarray treeviews, we could see that there was a big enrichment in up-regulated genes (in particular early regulated genes) associated with interaction anchors. In terms of percentages of up-regulated and down-regulated genes (**Supplementary Table 9**), the difference was also very striking. These findings suggest to us that interaction-associated ER-αBS employ looping for early transcriptional effects but stand-alone ER-αBS, which could have lower levels of ER-α proteins, might require a secondary co-activator for late transcriptional effects.

We used Fisher's Exact Test to test whether the differences in levels of H3K4me3, RNAPII, and expression microarray up-regulation marks that could be seen between UCSC Genes, standalone ER-αBS genes, anchor genes, loop genes, and enclosed anchor genes were significant. All anchor, loop and enclosed anchor data can be found in **Supplementary Table 8**. A summary is given in **Supplementary Table 9**.

One puzzling aspect of the data was that H3K4me3 and RNAPII marks did not show any significant differences between genes associated with stand-alone binding sites compared with genes associated with interaction anchors. While some stand-alone ER-αBS could be noise, we also hypothesize that some stand-alone ER-αBS could be still functional but might require secondary coactivators to be transcribed due to low local concentrations of ER-α, H3K4me3 and RNAPII could be present because the genes are poised for late transcriptional effects when secondary co-activators are recruited.

IHM001F

In IHM001F, comparing H3K4me3 marks between UCSC Genes without-interaction genes and loop genes, the 2-tailed *p*-value was 0.33, which is not significant. Comparing H3K4me3 marks between UCSC Genes without interaction genes and anchor genes, the 2-tailed *p*-value was $1.88e^{-2}$, which is significant. Comparing H3K4me3 marks between UCSC Genes without interaction genes and enclosed anchor genes, the 2-tailed *p*-value was 0.18, which is not significant. Comparing H3K4me3 marks between UCSC Genes without interaction genes and non-enclosed anchor genes, the 2-tailed *p*-value is $1.57e^{-4}$, which is significant. Comparing H3K4me3 marks between enclosed anchor genes and non-enclosed anchor genes, the 2-tailed *p*-value is $1.09e^{-3}$, which is significant.

Comparing RNAPII marks between UCSC Genes without interaction genes and loop genes, the 2-tailed *p*-value was $7.63e^{-3}$, which is significant. But more so than this, comparing RNAPII marks between UCSC Genes without interaction genes and anchor genes, the 2-tailed *p*-value was $1.37e^{-36}$, which is extremely significant. Comparing RNAPII marks between UCSC Genes without interaction genes and enclosed anchor genes, the 2-tailed *p*-value was 3.07^{-16}, which is very significant. Comparing RNAPII marks between UCSC Genes without interaction genes and non-enclosed anchor genes, the 2-tailed *p*-value is $8.85e^{-23}$, which is extremely significant. Comparing RNAPII marks between enclosed anchor genes and non-enclosed anchor genes, the 2-tailed *p*-value is 0.24, which is not significant.

Comparing up-regulated marks between UCSC Genes without interaction genes and loop genes, the 2-tailed *p*-value was $5.24e^{-4}$, which is significant. But more so than this, comparing up-regulated marks between UCSC Genes without interaction genes and anchor genes, the 2-tailed *p*-value was $8.52e^{-37}$, which is extremely significant. Comparing up-regulated marks between UCSC Genes without interaction genes and enclosed anchor genes, the 2-tailed *p*-value was $6.23e^{-7}$, which is very significant. Comparing up-regulated marks between UCSC Genes without interaction genes and non-enclosed anchor genes, the 2-tailed *p*-value was $3.99e^{-33}$, which is extremely significant. Comparing up-regulated marks between non-enclosed anchor genes and enclosed anchor genes, the 2-tailed *p*-value is $2.00e^{-2}$, which is significant.

Comparing down-regulated marks between UCSC Genes without interaction genes and loop genes, the 2-tailed *p*-value was $4.20e^{-2}$, which is significant. Comparing down-regulated marks between UCSC Genes without interaction genes and anchor genes, the 2-tailed *p*-value was $1.03e^{-5}$, which is very significant. Comparing down-regulated marks between UCSC Genes without interaction genes and enclosed anchor genes, the 2-tailed *p*-value was 0.65, which is not significant. Comparing down-regulated marks between UCSC Genes without interaction genes and non-enclosed anchor genes, the 2-tailed *p*-value is $7.22e^{-7}$, which is very significant. Comparing down-regulated marks between non-enclosed anchor genes and enclosed anchor genes, the 2-tailed *p*-value is $2.20e^{-2}$, which is significant.

Comparing H3K4me3 marks between UCSC Genes without stand-alone binding site associated genes and stand-alone binding site-associated genes, the 2-tailed *p*-value was $4.29e^{-45}$, which is extremely significant. Comparing RNAPII marks between UCSC Genes without stand-alone binding site associated genes and stand-alone binding site-associated genes, the 2-tailed *p*-value was $2.20e^{-61}$, which is extremely significant. Comparing up-regulated marks between UCSC Genes without stand-alone binding site associated genes and stand-alone binding site-associated genes, the 2-tailed *p*-value was $6.82e^{-15}$, which is very significant.

Comparing down-regulated marks between UCSC Genes without stand-alone binding site associated genes and stand-alone binding site-associated genes, the 2-tailed *p*-value was $1.89e^{-76}$, which is extremely significant.

IHH015F

IHH015F shows somewhat similar trends. Comparing H3K4me3 marks between UCSC Genes without interaction genes and loop genes, the 2-tailed *p*-value was $2.46e^{-8}$, which is very significant. Comparing H3K4me3 marks between UCSC Genes without interaction genes and anchor genes, the 2-tailed *p*-value was $2.60e^{-3}$, which is significant. Comparing H3K4me3 marks between UCSC Genes without interaction genes and enclosed anchor genes, the 2-tailed *p*-value was 0.62, which is not significant. Comparing H3K4me3 marks between UCSC Genes without interaction genes and non-enclosed anchor genes, the 2-tailed *p*-value

is $7.71e^{-7}$, which is very significant. Comparing H3K4me3 marks between non-enclosed anchor genes and enclosed anchor genes, the 2-tailed *p*-value is $8.39e^{-5}$, which is significant.

Comparing RNAPII marks between UCSC Genes without interaction genes and loop genes, the 2-tailed *p*-value was $7.92e^{-17}$, which is very significant. But more so than this, comparing RNAPII marks between UCSC Genes without interaction genes and anchor genes, the 2-tailed *p*-value was $1.96e^{-24}$, which is extremely significant. Comparing RNAPII marks between UCSC Genes without interaction genes and enclosed anchor genes, the 2-tailed *p*-value was $4.19e^{-12}$, which is very significant. Comparing RNAPII marks between UCSC Genes without interaction genes and non-enclosed anchor genes, the 2-tailed *p*-value is $1.03e^{-14}$, which is very significant. Comparing RNAPII marks between non-enclosed anchor genes and enclosed anchor genes, the 2-tailed *p*-value is 0.49, which is not significant.

Comparing up-regulated marks between UCSC Genes without interaction genes and loop genes, the 2-tailed *p*-value was 0.90, which is not significant. Comparing up-regulated marks between UCSC Genes without interaction genes and anchor genes, the 2-tailed *p*-value was $2.30e^{-27}$, which is extremely significant. Comparing up-regulated marks between UCSC Genes without interaction genes and enclosed anchor genes, the 2-tailed *p*-value was $2.78e^{-20}$, which is very significant. Comparing up-regulated marks between UCSC Genes without interaction genes and non-enclosed anchor genes, the 2-tailed *p*-value was $2.48e^{-10}$, which is very significant. Comparing up-regulated marks between non-enclosed anchor genes and enclosed anchor genes, the 2-tailed *p*-value is $7.84e^{-2}$, which is weakly significant.

Comparing down-regulated marks between UCSC Genes without interaction genes and loop genes, the 2-tailed *p*-value was $1.72e^{-5}$, which is significant. Comparing down-regulated marks between UCSC Genes without interaction genes and anchor genes, the 2-tailed *p*-value was $7.35e^{-6}$, which is very significant. Comparing down-regulated marks between UCSC Genes without interaction genes and enclosed anchor genes, the 2-tailed *p*-value was 0.63, which is not significant. Comparing down-regulated marks between UCSC Genes without interaction genes and non-enclosed anchor genes, the 2-tailed *p*-value was $1.45e^{-11}$, which is very-significant. Comparing down-regulated marks between non-enclosed anchor genes and enclosed anchor genes, the 2-tailed *p*-value is $4.70e^{-7}$, which is very significant.

Comparing H3K4me3 marks between UCSC Genes without stand-alone binding site associated genes and stand-alone binding site-associated genes, the 2-tailed *p*-value was $5.41e^{-26}$, which is extremely significant. Comparing RNAPII marks between UCSC Genes without stand-alone binding site associated genes and stand-alone binding site-associated genes, the 2-tailed *p*-value was $1.05e^{-52}$, which is extremely significant. Comparing up-regulated marks between UCSC Genes without stand-alone binding site associated genes and stand-alone binding site-associated genes, the 2-tailed *p*-value was 0.23, which is not significant. Comparing down-regulated marks between UCSC Genes without stand-alone binding site associated genes and stand-alone binding site-associated genes, the 2-tailed *p*-value was $7.83e^{-8}$, which is very significant.

The great differences in the observed 2-tailed *p*-values for up and down-regulation in the stand-alone binding site-associated genes between IHH015F and IHM001F appear because of the differences in the number of transcriptional units in these two libraries. Looking at the percentages (IHH015F 40.8% up-regulated, 59.2% down-regulated and IHM001F 40.1% up-regulated, 59.9% down-regulated), the trends are similar (**Supplementary Table 9**).

As such, investigation of the patterns of H3K4me3, RNAPII and up-regulated expression microarray marks in two different biological replicates of ChIA-PET support the notion that genes associated with oestrogen-

bound chromatin interactions tend to be regulated under oestrogen conditions, in particular anchor genes, and to a lesser extent, loop genes, as compared with background genes that are not associated in any way with chromatin interactions. While chromatin interactions are associated with both gene repression and gene activation, on the balance, it appears that the percentages of up-regulated genes (and hence gene activation) are higher when associated with anchor and enclosed anchor genes, but the percentages of down-regulated genes (and hence gene repression) are higher when associated with loop genes.

Supplementary Note 8. Chromatin interactions at the keratin gene cluster

Keratins play major structural roles in cells[23-25], and mutations give rise to various human hereditary keratin diseases, such as epidermolysis bullosa simplex[26]. Keratins are also known to be involved in signaling and regulatory pathways[26]. Keratins have very distinct expression patterns, and epithelial tumors frequently have the same patterns as the originating cells. This finding has led some genes, including KRT8, KRT18, and KRT7, to be used in immunohistochemistry analyses of cancers to identify tumor origins[26]. Keratins are present in the human genome as two families: type I genes on chr17, and type II genes on chr12[24]. Keratins are unique in that type I genes and type II genes pair up by the formation of a heterodimer between one type I and one type II. Any keratin proteins that deviate from this rule are rapidly degraded[27]. Therefore, gene expression in the keratin gene cluster has to be highly regulated in order to maintain distinct coexpression patterns. We hypothesize that chromatin interactions help in coordinating gene regulation and in maintaining coexpression patterns. We examined MCF-7 human breast adenocarcinoma cells, which are derived from ductal epithelial cells. Of the keratins used in immunohistochemistry diagnosis, breast adenocarcinomas typically express KRT8, KRT18, KRT19, KRT7, and occasionally KRT5, but not KRT20. Analysis of chromatin interactions in the keratin region suggests that chromatin interactions are correlated with gene expression coordination. Both ChIA-PET and 4C data shows that KRT7, KRT8, and KRT18 are all pulled into the "hub" of the same interaction complex. KRT7, 8, and 18 are known to be expressed in breast carcinomas. In particular, KRT8 and KRT18 are tightly coexpressed genes, and the gene products bind tightly to each other. These two genes are connected by many inter-ligations. By contrast, KRT5, 6, 1, 2, and other keratins involved in other aspects such as in hair development for example KRT72 and KRT75, are not expressed, and they are present in the "loop" of the interaction complex. Hence, chromatin interactions in the keratin region may bring together relevant genes into transcriptional foci, and loop out irrelevant genes, in order to achieve tightly coordinated gene expression regulation. The keratin gene region is shown in detail in **Fig. 1c** and **Supplementary Figs. 20e and 20f**.

Supplementary Note 9. Additional discussion

In this study, we demonstrated the ChIA-PET strategy combined with ultra-high-throughput sequencing is an unbiased, whole genome approach for *de novo* analysis of chromatin interactions. This approach represents a major technological advance in our ability to study higher-order organization of chromosomal structures and functions. A single ChIA-PET experiment can provide two global datasets: the protein factor binding sites and the interactions among the binding sites, and hence is conceptually superior to all currently available methodologies for chromatin interactome analysis including the 3C-based methods that are limited in scope, and the recently reported ChIP-Seq[28,29] that provides only binding sites. The ChIA-PET method features sonication and linker barcoding to reduce noise encountered in chromatin interaction analysis, and ChIP to reduce complexity and add specificity to interaction data[30]. One caveat for ChIP-based interaction analysis methods including ChIA-PET and ChIP-3C is that chromatin interactions due to random flexing of DNA polymers in short distances cannot be easily distinguished from specific, protein-bound interactions. Also, to establish whether protein-bound interactions are indeed mediated by the protein factor, functional analyses such as siRNA knockdown of the factor followed by chromatin interaction analyses would be required. How to structurally validate and functionally analyze the chromatin interactions identified by ChIA-PET

experiments on a large scale is a new challenge to the field. Lastly, due to the potential high complexity and heterogeneity of chromatin interactions, deep sequencing for ChIA-PET analysis is necessary. Although we applied 454[31] and Illumina GAII sequencing[28,29] in this study, the PET strategy is suited to next-generation sequencers[32], and ChIA-PET can be coupled with other tag-based sequencing systems, such as ABI SOLiD[33], Helicos[34] and newer sequencing technologies for even deeper library coverage (**Summary of Results in Supplementary Information, Supplementary Fig. 1a**).

Recent genome-wide ChIP studies on many TFBSs including ER-α[1,2,22] have raised numerous questions such as which distal TFBSs are functional and how they enable remote control of gene transcription? ChIA-PET analysis provides plausible answers to these questions and greatly increases the accuracy of gene assignment to TFBS. As illustrated in **Fig. 5** and **Supplementary Fig. 21**, the strongest ER-αBS in the *E2F6* and *GREB1* region is distal to both of the genes, and it is assigned to only *GREB1* because of the ChIA-PET data. In this study, we identified 1,955 high-confidence distal ER-αBS involved in 689 ER-α-bound chromatin interactions associated with 1,575 genes located close to the interaction anchor regions, including many previously characterized ER-α target genes. RNAPII occupancy as well as microarray gene expression data suggest that these "anchor" genes are preferentially up-regulated by oestrogen as compared to "loop" genes buried deep within the loops of chromatin interactions and genes assigned by stand-alone ER-αBS that are not involved in interactions. Therefore, the presence of chromatin interactions at ER-αBS could be a functional measure for better prediction of ER-α target genes. Furthermore, this ER-α-bound chromatin interactome map provides a comprehensive structure basis to understand how ER-α may function in transcription regulation. We postulate the following primary mechanism for ER-α function: ER-α protein dimers are recruited to multiple ER-αBS including distal ER-αBS which can interact with one another and possibly other factors such as RNAPII and FoxA1 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. This model addresses the question of how coordinated transcription may be achieved: transcription factors use chromatin interactions to loop multiple genes, such as *FOS/JDP2/BATF* and keratin genes, into the same transcription foci for coordinated transcription regulation. The size of the loops may also have different functions. Small loops may wrap up entire gene bodies in confined compartment for efficient transcription, and the large loops may push genes away from the center of transcription foci (**Summary of Results in Supplementary Information, Supplementary Fig. 1b**).

An intriguing question arising from this study is why a transcription factor such as ER-α evolved to use such an extensive and intensive chromatin interaction mechanism for transcriptional regulation, or how would this mechanism benefit the cells. Our data suggest that chromatin interactions represent the most efficient use of binding sites constrained by an imposed linear distribution (order) on several levels. First, the obvious redundancy in ER-α binding and interactions could enhance the robustness of ER-α transcriptional control such that mutations at any one interaction site would not entirely eliminate regulatory control. Second, given that speciation may be driven by retrotransposon dispersion through altering regulatory control[35], we speculate that extensive looping may allow for evolutionary binding site experimentation by a retrotransposon dispersal strategy. Third, by bringing two or more binding sites together, chromatin interactions will result in higher concentration levels of transcription factors in that spatial region, thus allowing for greater modulation of genetic responses while reducing the need for the cell to synthesize more transcription factor proteins. Fourth, as a matter of topology, looping and anchor clustering in 3-dimensional space provides greater access of ER-α for direct regulation of the transcriptional machinery, as compared with the proximity constraints of linear DNA. We further speculate that chromatin interaction centers involve many strands of chromatin coming together that could help achieve and maintain high local concentrations of

transcriptional components. Loops that connect gene transcription start and end sites may allow for cycling of the transcriptional machinery in a highly efficient manner. It is now known that ER-α-DNA interactions at a defined ER-αBS oscillate in an on-off state with periodicity, and oscillators use boundaries to change wave direction[36, 37]. Given the extensive system of interaction complexes, ER-α could oscillate between spatially proximate anchors of interaction regions, using the chromatin interaction boundaries to provide oscillation dynamics to ER-α behavior. Thus, the looping and anchor system we hypothesize represents a topological solution to a number of mechanistic observations of this transcription factor. Similar mechanisms may be employed by other transcription factors in mammalian genomes.

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 factor biology in whole genome contexts.

Supplementary Tables

Supplementary Table 1. ER-αBS identified in this study and associated genes (in a separate Excel file)

Supplementary Table 2. ER-α-bound chromatin interaction (duplex interaction) sites identified in this study (in a separate Excel file). The genomic spans of the interactions are also provided.

Supplementary Table 3. ER-α-bound chromatin interaction regions identified in this study (complex interactions and standalone duplex interactions) with associated genes (in a separate Excel file). The genomic spans of the interactions are also provided.

Supplementary Table 4. FISH validation data

***NR2F2* FISH experiment data**				
Type	**Control**	**Control**	**Control**	**Interaction**
Probes	**ET: P2/P3**	**ET: P1/P2**	**E2: P2/P3**	**E2: P1/P2**
Separate	224	212	334	121
Overlap	114	322	87	343
Total	338	534	421	464
Overlap %	33.7%	60.3%	20.7%	73.9%
Normalized probe overlap rate (X)	1X	1.79X	1X	3.57X
***GATA3* FISH experiment data**				
Separate	244	159	236	114
Overlap	156	241	165	294
Total	400	400	401	408
Overlap %	39.0%	60.3%	41.1%	72.1%
Normalized probe overlap rate (X)	1X	1.54X	1X	1.75X
Chr7/Chr12 FISH experiment data				
Separate	308	389	319	388
Overlap	8	13	7	13
Total	316	402	326	401
Overlap %	2.5%	3.2%	2.1%	3.2%
Chr9/Chr14 FISH experiment data				
Separate	468	452	448	444
Overlap	3	11	7	7
Total	471	463	455	451
Overlap %	0.6%	2.4%	1.5%	1.6%

For *NR2F2* and *GATA3*, P1 and P2 are test BAC probes near the two anchors of the interaction complex covering >1 Mb. P3 is a control BAC probe. For the interchromosomal interactions, P1 and P2 are the test BAC probes, one to each chromosomal location. P3 is a different chromosomal location on a different chromosome. The FISH experiments using the combined probes of P1/P2 and P2/P3 were done in ethanol control (ET) and oestrogen treated (E2) MCF-7 cells. Each FISH experiment analyzed 100-200 nuclei, and all spots in each nucleus were counted. The FISH images of the probe pairs show red and green spots when the probes are separated (ET: P3/P2 and E2: P2/P3), and yellow sections between red and green spots when the probes overlap (ET: P2/P1 and E2: P2/P1). The probe overlap rate is normalized using the control probe pair (P2/P3) as the base level of background noise. Both *GATA3* and *NR2F2* show a basal level interaction in oestrogen-untreated conditions, and a further induction of interactions under oestrogen treatment, indicating the chromatin interaction is oestrogen-dependent. The results were significant by Fisher's Exact Test (**Supplementary Methods**). By contrast, the interchromosomal interactions have extremely low levels of overlap and so were not tested by Fisher's Exact Test. The *TFF1-GREB1* interchromosomal interaction (Chr 2 / Chr 12) was assessed by eye, no difference was visible between the different conditions. Given that the other two interchromosomals did not show any differences upon detailed counting, the *TFF1-GREB1* interaction was not counted in full detail.

Supplementary Table 5. Validated chromatin interactions in ChIA-PET data

	Chromatin interactions	PET counts* IHM001F IHH015F	Validation methods	Comments and references
1	***TFF1***[a,b]	3,301 20,929	3C, ChIP-3C	• Shown to be mediated by ER-α through siRNA 3C, ChIP-3C by Pan et al. 2008[38] • ChIP-3C by Carroll et al. 2005[39]
2	***GREB1***[a,b]	8,453 14,158	3C, ChIP-3C	• Shown to be mediated by ER-α through siRNA, 3C, ChIP-3C in this study • 3C by Deschenes et al., 2007[40]
3	***P2RY2***[a]	1,180 10,978	3C, ChIP-3C	
4	***SIAH2***[a]	2,576 10,874	3C, ChIP-3C	
5	**Keratin genes**[a]	9,790 3,922	4C	• 3 interactions were validated
6	***CA12***[b]	198 279	3C	• 3C by Barnett et al. 2008[41]
7	***CTSD***[b]	1,419 9,577	3C	• 3C by Bretschneider et al., 2008[42]
8	***NR2F2***[a]	4,449 320	FISH	• The distance between the two test sites is at the border line (1Mb) for FISH experiment
9	***GATA3***[a]	4,118 2,085	FISH	• The distance between the two test sites for FISH is 1.6Mb.
10	**Chr7 to Chr12**[a]	2 0	FISH	• ChIA-PET data is unreproducible • FISH test sites: chr7:2692293-2698011and chr12:120670670-120673138
11	**Chr9 to Chr14**[a]	2 0	FISH	• ChIA-PET data is unreproducible • FISH test sites: chr9:113812545-113817679 and chr14:90811137-90814459
12	***BX647900***[a]	343 1,193	ChIP-3C	
13	***CAP2***[a]	1,217 3,354	ChIP-3C	
14	***ELOVL2***[a]	500 978	ChIP-3C	
15	***SLC9A3R1***[a]	1,500 2,029	ChIP-3C	

16	***DB554846***[a]	4,449 320	ChIP-3C	DB554846 is a gene prediction.
17	***NRIP1***[b]	8,531 16,084	ChIP-3C	• ChIP-3C by Carroll et al. 2005[39]. NRIP1 is denoted a "loop gene".
18	***NOTCH2***[a]	0 0	3C	• A negative control site which showed no interactions by both ChIA-PET and 3C
19	***GREB1*** **(in chr2)** **to** ***TFF1*** **(in chr22)**[a]	0 0	FISH	• A negative control site which showed no interactions by both ChIA-PET and 3C under these experimental conditions. • Previous results by a different group (Hu et al., *PNAS*, 2008) showed a possible interaction, but this could be due to different treatment in different cells.
20	***P2RY2*** **(in chr11)** **to** ***GREB1*** **(in chr2)**[a]	0 0	ChIP-3C	• A negative control site which showed no interactions by both ChIA-PET and ChIP-3C under these conditions • Was tested due to concerns that sites with high ChIP enrichment might show non-specific interactions

[a] Data shown in this study; [b] Data also shown in other studies

***** Number of inter-ligation PETs in the associated complex or stand-alone duplex interaction from **Supplementary Table 3**. The number on top is the count in IHM001F and the number below is the count for IHH015F.

Supplementary Table 6. Reproducibility of ER-αBS identified by ChIA-PET data

	IHM001F	**IHH015F**
ER-αBS peaks (ChIP enrichment ≥5) ^	13,111	5,409
Overlap of ER-αBS peaks in replicate library	3,785 (28.9%)*	3,791 (70.1%)*
ER-αBS high peaks (ChIP-enrichment ≥50)	2,214	1,847
Overlap of ER-αBS high peaks in replicate library	1,589 (71.8%)*	1,638 (88.7%)*
Overlap of top 100 ER-αBS peaks in replicate library	100 (100%)	99 (99%)

^ ER-αBS peaks are measured by frequency of number of self-ligation PETs, reflecting ChIP enrichment in each PET cluster. Because overlaps between peaks in amplicon regions might be high, we removed peaks in amplicons for this analysis. We also performed the same analysis with all peaks, and found that the numbers did not differ by much.

*Numbers of overlapping peaks between these two libraries might not be equal as one peak in one library could overlap with two or more peaks in another library.

Note: We checked the overlaps of one particular set from one library as stated in the table, with the whole other library – meaning for example that 2,214 of ER-αBS high peaks (ChIP-enrichment ≥50) from IHM001F could be found within all of IHH015F (all ER-αBS peaks (ChIP enrichment ≥5).

Supplementary Table 7. Reproducibility of ER-α-bound chromatin interactions in ChIA-PET library replicates

	IHM001F	**IHH015F**
Duplex interactions sites	1475	3561
Overlap of interaction sites in replicate library	681 (46.2%)*	710 (19.9%)*
Overlap of top 100 most abundant interaction sites in replicate library^	86 (86%)	86 (86%)
Overlap of top 50 most abundant interaction sites in replicate library^	45 (90%)*	46 (92%)*
Complex interactions	274	519
Overlap of complex interactions in replicate library	237 (86.5%)*	333 (64.2%)*
Stand-alone duplex interactions	415	423
Overlap of stand-alone duplex interactions in replicate library	234 (56.4%)*	119 (28.1%)*
Chromatin interaction regions*	689	942
Overlap of interaction regions in replicate library^	471 (68.4%)*	452 (48.0%)*
Overlap of top 100 most abundant interaction regions in replicate library^	94 (94%)*	95 (95%)*
Overlap of top 50 most abundant interaction regions in replicate library^	47 (94%)*	48 (96%)*

*Numbers of overlapping peaks between these two libraries might not be equal as one interaction or interaction region in one library could overlap with two or more interactions or interaction regions in another library.
^Based on number of inter-ligation PETs in the chromatin interaction or chromatin interaction region.
Note: We checked the overlaps of one particular set from one library as stated in the table, with the whole other library – meaning for example that we overlapped the top100 most abundant interaction regions in IHM001F with all chromatin interaction regions from IHH015F.

Supplementary Table 8. Genes associated with ER-α-bound chromatin interactions (in a separate Excel file)

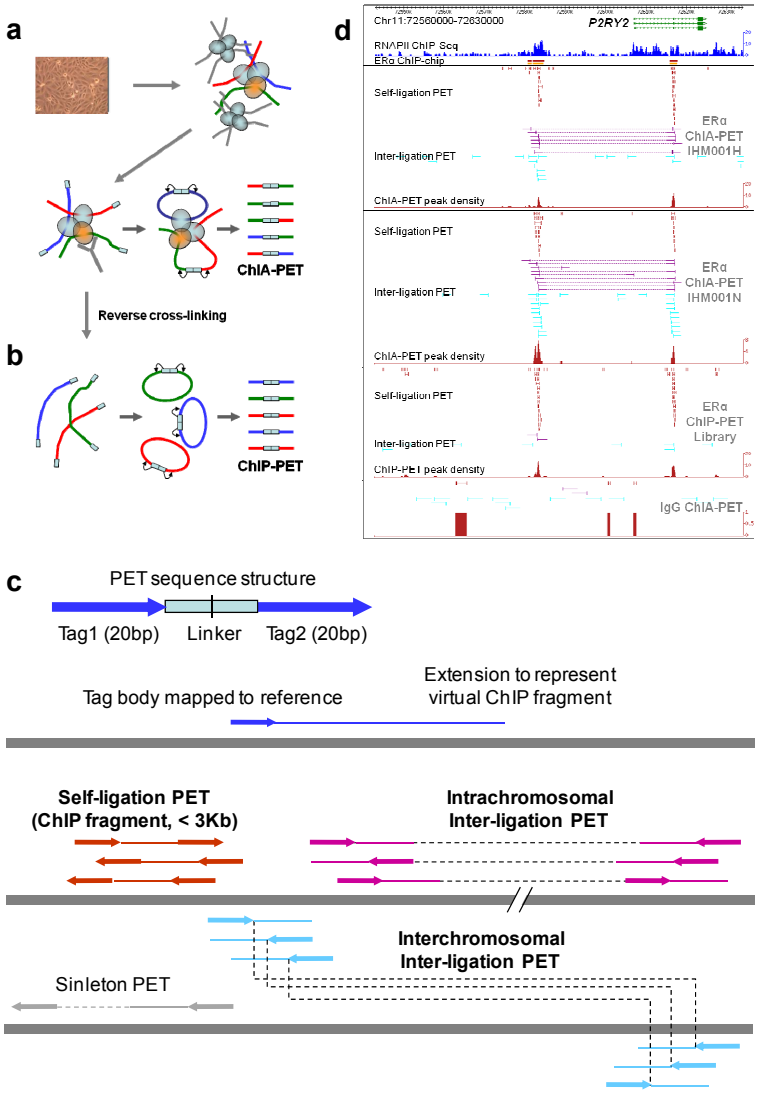
Supplementary Table 9. Association of ER-α-bound chromatin interactions with gene transcription status

IHM001F		Transcription marks		Expression status	Differentially expressed genes	
	Number of TUs^	H3K4me3	RNAPII	Differentially expressed	Up-regulated	Down-regulated
All UCSC Genes*	64,976	32,633 (50.2%)	15,948 (24.5%)	8,276 (12.7%)	3,792 (45.8%)	4,484 (54.2%)
Interaction-associated genes (Anchor + Loop)	5,342	2,698 (50.5%)	1,589 (29.8%)	929 (17.4%)	487 (52.4%)	442 (47.6%)
Loop genes	3,767	1,860 (49.4%)	980 (26.0%)	549 (14.6%)	261 (47.5%)	288 (52.5%)
Anchor genes	1,575	838 (53.2%)	609 (38.7%)	380 (24.1%)	226 (59.5%)	154 (40.5%)
Non-enclosed anchor genes	1,080	605 (56.0%)	407 (37.7%)	288 (26.7%)	170 (59.0%)	118 (41.0%)
Enclosed anchor genes	495	233 (47.1%)	202 (40.8%)	92 (18.6%)	56 (60.9%)	36 (39.1%)
Stand-alone binding site-associated genes	11,790	6,613 (56.1%)	3,606 (30.6%)	2,175 (18.4%)	873 (40.1%)	1,302 (59.9%)
IHH015F						
All UCSC Genes*	64,976	32,633 (50.2%)	15,948 (24.5%)	8,276 (12.7%)	3,792 (45.8%)	4,484 (54.2%)
Interaction-associated genes (Anchor + Loop)	9,547	5,068 (53.1%)	2,831 (29.7%)	1,502 (15.7%)	705 (46.9%)	797 (53.1%)
Loop genes	5,994	3,208 (53.5%)	1,714 (28.6%)	825 (13.8%)	336 (40.7%)	489 (59.3%)
Anchor genes	3,553	1,860 (52.4%)	1,117 (31.4%)	677 (19.1%)	369 (54.5%)	308 (45.5%)
Non-enclosed anchor genes	1,725	962 (55.8%)	552 (32.0%)	355 (20.6%)	163 (45.9%)	192 (54.1%)
Enclosed anchor genes	1,828	898 (49.1%)	565 (30.9%)	322 (17.6%)	206 (64.0%)	116 (36.0%)
Stand-alone binding site-associated genes	3,925	2,291 (58.4%)	1,378 (35.1%)	603 (15.4%)	246 (40.8%)	357 (59.2%)

^ All numbers are given in terms of gene transcriptional units. "Transcriptional unit" is abbreviated as "TU". Only TU present on chromosomes 1-22 and chromosome X are shown. *All UCSC Genes are given as a control. The proportion of interaction-associated genes or stand-alone binding site-associated genes is much less. ¶ The percentages of genes with H3K4me3 and RNAPII marks are based on the number of TU in each category. § The percentages of genes differentially expressed are based on the number of TU in each category. # The percentages of up or down regulated genes are based on the number of differentially expressed genes in each category. Gene transcriptional units (UCSC Genes, hg18)[13] were assigned to interaction complexes and duplexes (collectively called "interaction regions"). The interaction-associated genes (transcriptional units) were further partitioned into different interaction categories, and the numbers of down-regulated, up-regulated, H3K4me3-containing, and RNAPII-containing gene transcriptional units are shown. All loop, non-enclosed anchor and enclosed anchor genes are given in **Supplementary Table 8**.

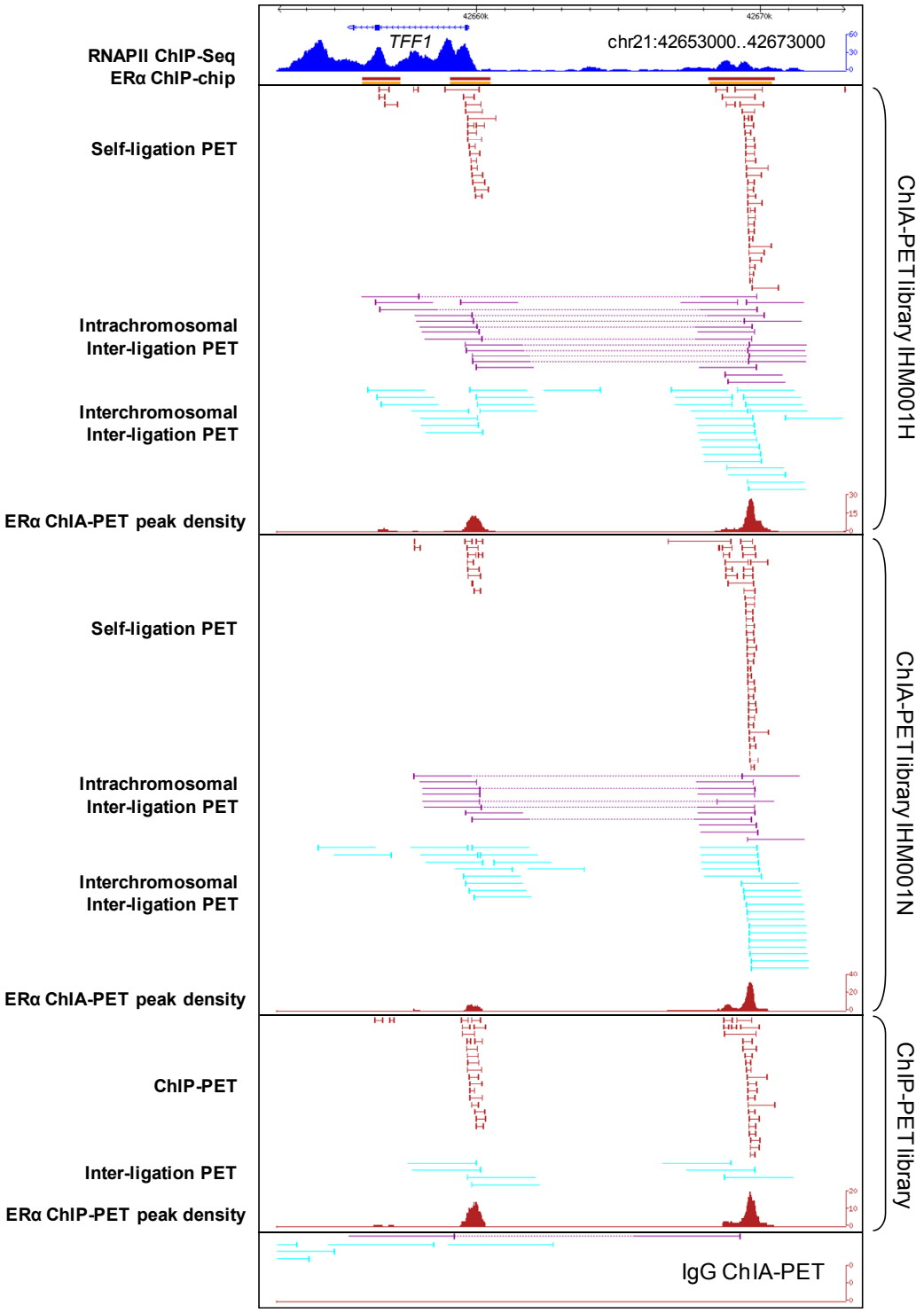
Supplementary Table 10. Sequences and notes (in a separate Excel file)

Supplementary Figures



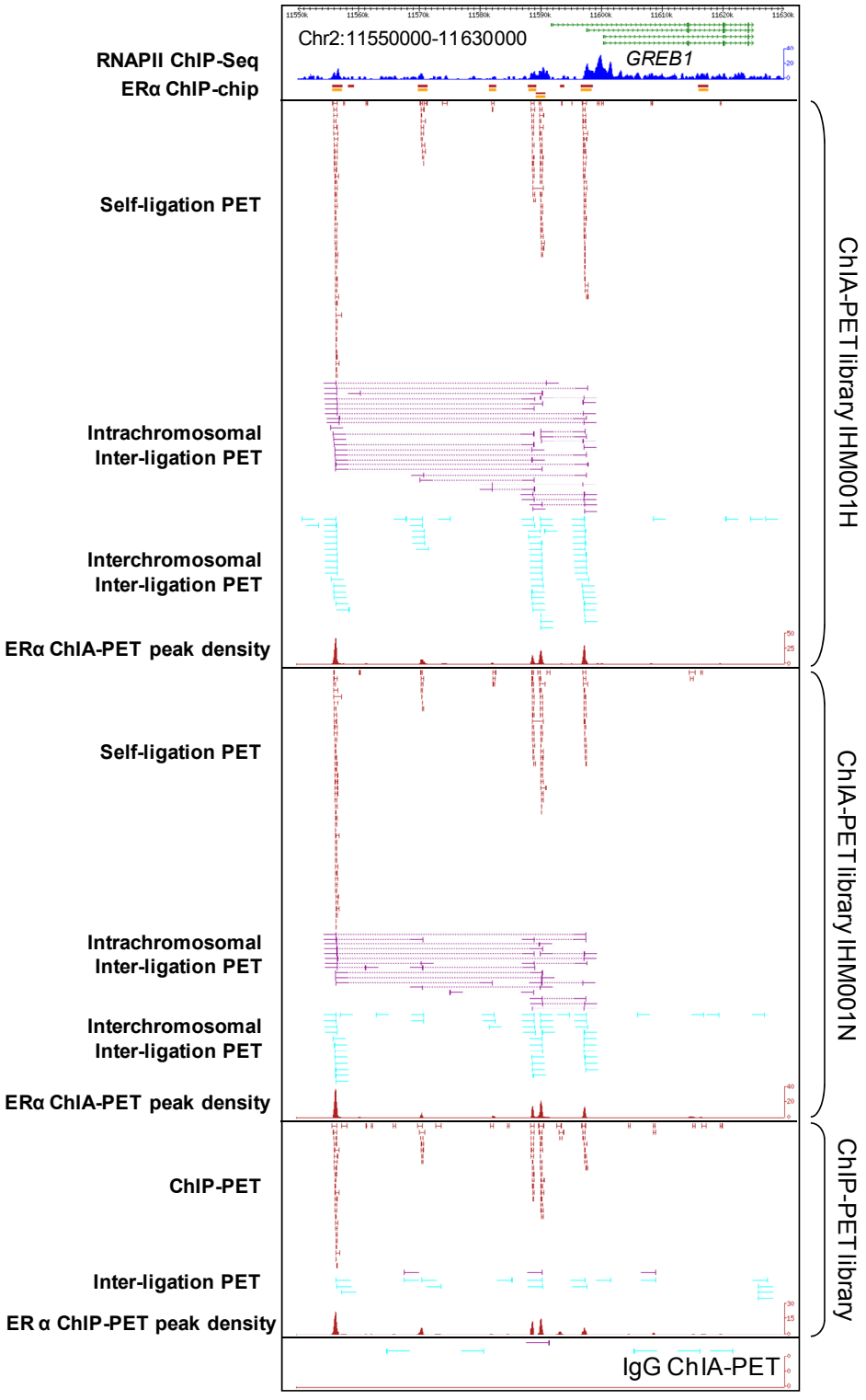
Supplementary Figure 2. ChIA-PET mapping scheme and examples

a. Schematic view of the ChIA-PET method. **b.** The ChIP-PET method, which is a control library to ChIA-PET library. The only difference between ChIA-PET and ChIP-PET is that in ChIP-PET, chromatin complexes are reverse cross-linked and the bound DNA fragments are released, before the second the proximity ligation. **c.** The structure of a ChIA-PET sequence and the mapping of self-ligation PETs and inter-ligation PETs to the reference genome. Overlapping PET clusters are considered putative interaction signals and singleton PETs are noise. **d.** Two replicate ChIA-PET data (IHM001H and IHM001N) mapped at the *P2RY2* locus, showing self-ligation PET and inter-ligation PET (purple for intra- and blue for inter-chromosomal). The ChIP-PET library data showed only self-ligation PET clusters for two ER-αBS and scattered singletons of inter-ligation PET. The IgG ChIA-PET mock library showed only singletons.



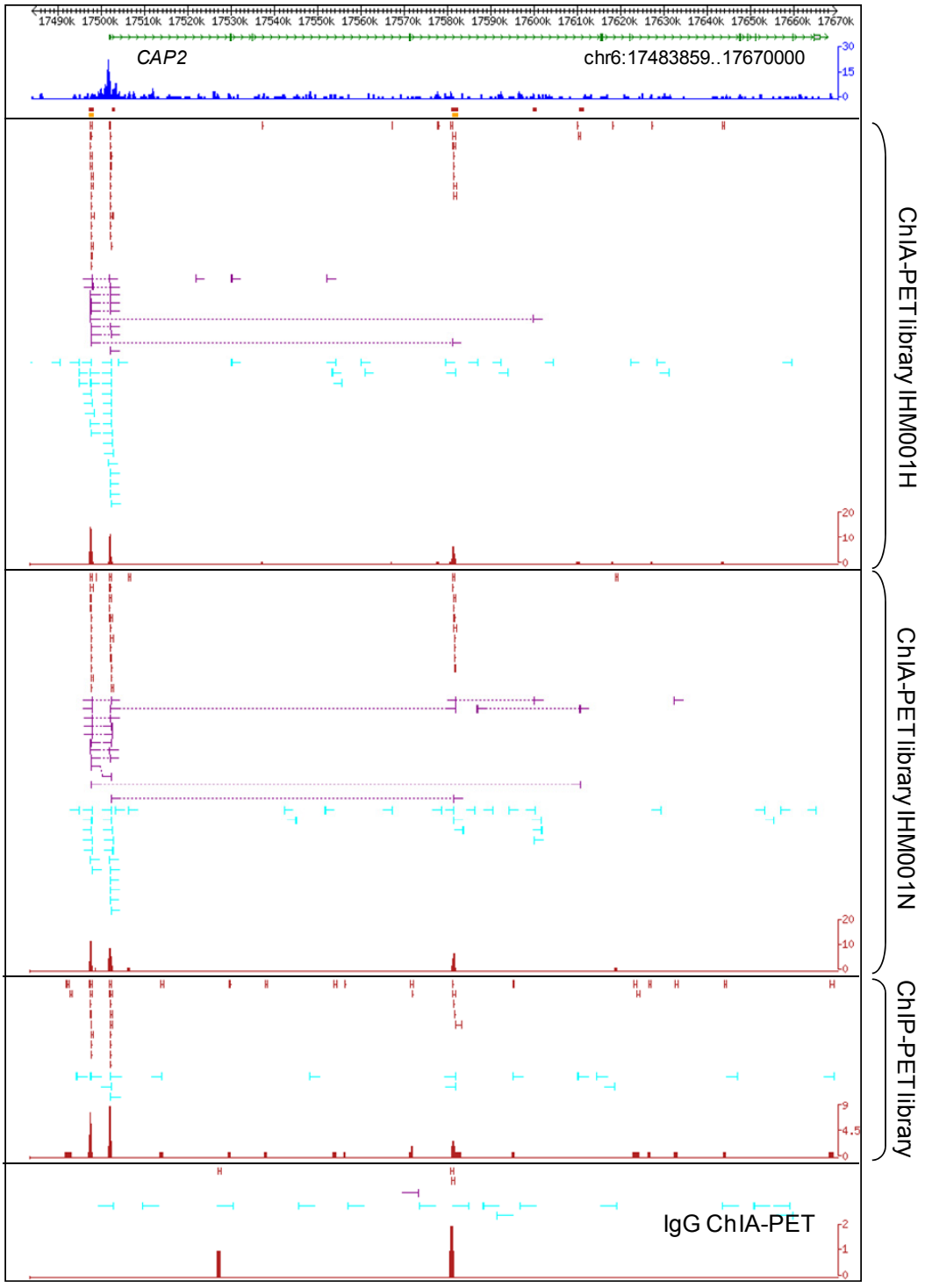
Supplementary Figure 2. ChIA-PET mapping scheme and examples

e. Two replicates of ER-α ChIA-PET data and control libraries mapping at the *TFF1* locus.



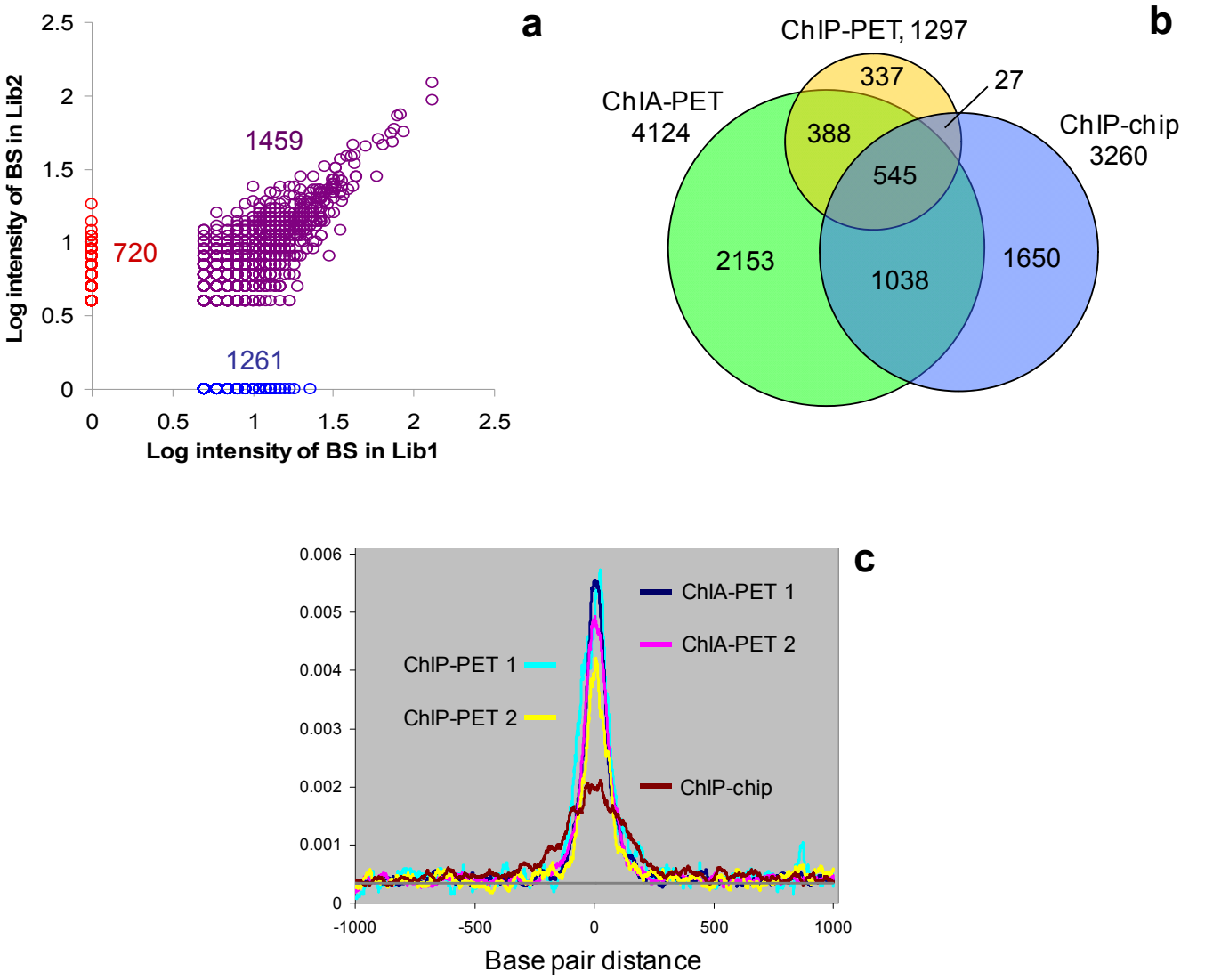
Supplementary Figure 2. ChIA-PET mapping scheme and examples

f. Two replicates of ER-α ChIA-PET data and control libraries mapping at the *GREB1* locus.



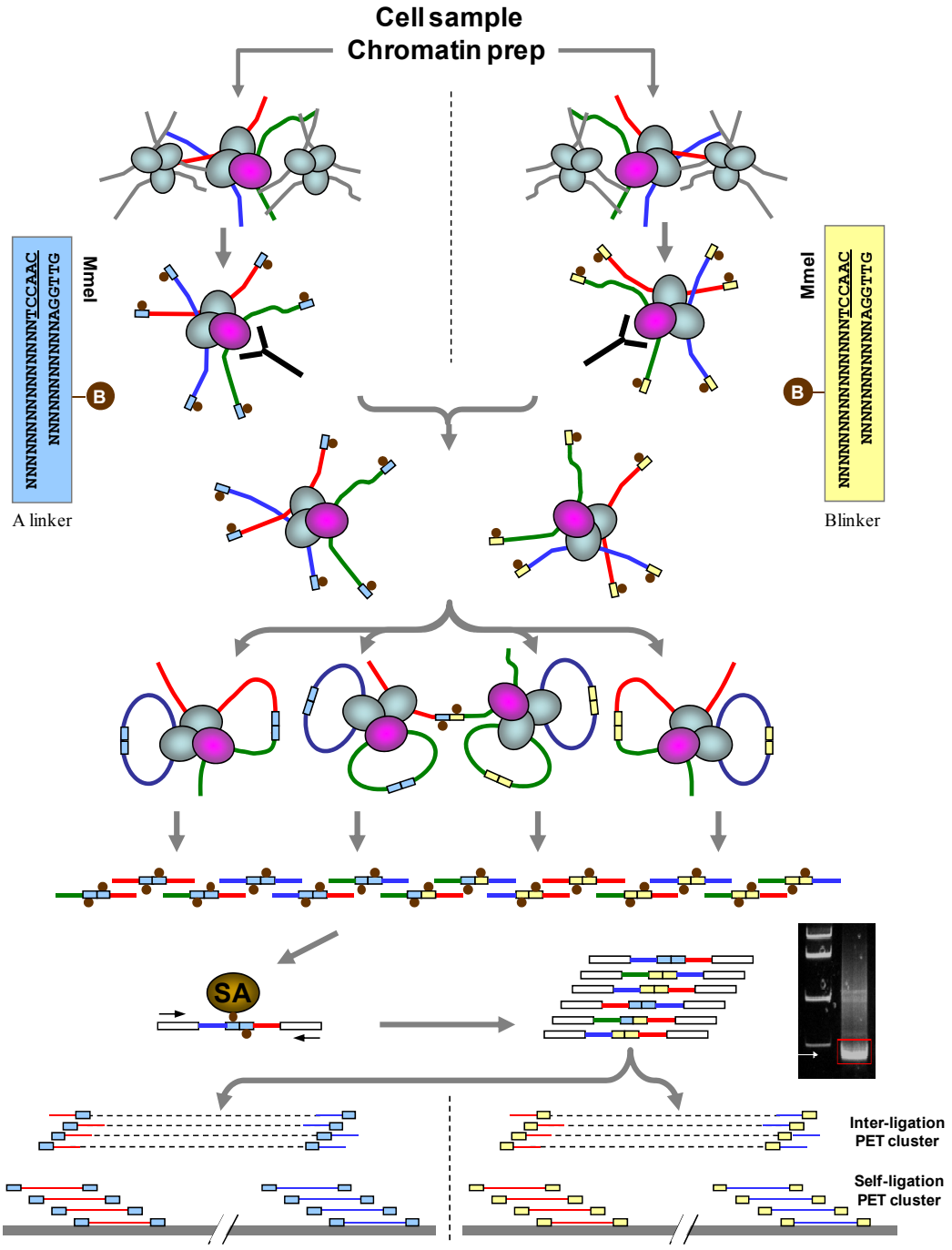
Supplementary Figure 2. ChIA-PET mapping scheme and examples

g. Two replicates of ER-α ChIA-PET data and control libraries mapping at the *CAP2* locus.



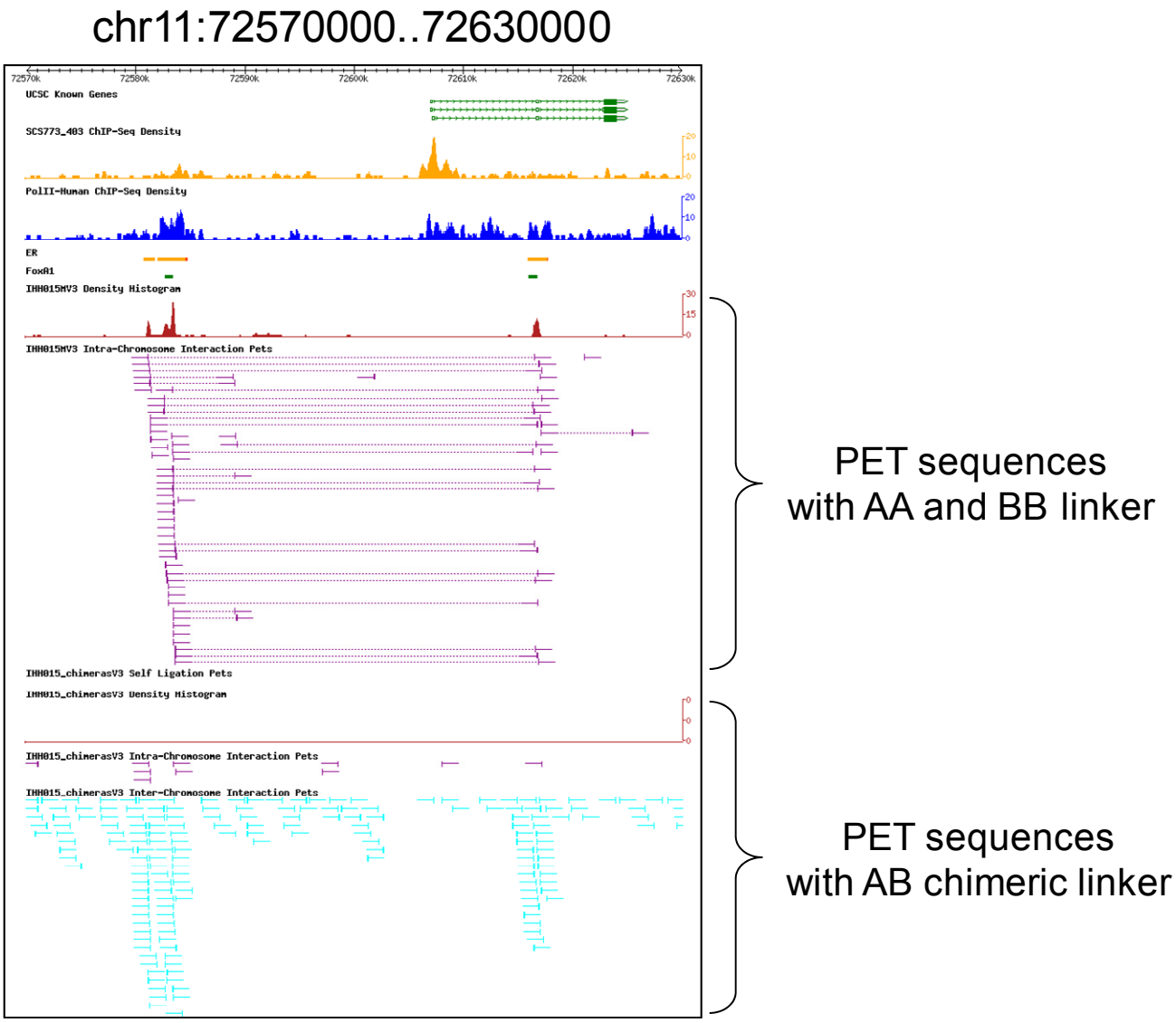
Supplementary Figure 3. Reproducibility of ER-αBS identified by pilot ChIA-PET libraries

a. The ER-α binding sites identified by ER-α ChIA-PET experiments are largely reproducible. The majority of ER-α binding sites found in one library can also be found in the other library ("Lib1" = IHM001H, "Lib2" = IHM001N). The intensities of the ER-α binding peaks identified in both libraries are highly correlated. The correlation coefficient is 0.9. **b.** Venn Diagram of ER-α binding sites found by different studies. The comparison was performed between the ER-α binding sites found by ChIA-PET, ChIP-PET[1], and ChIP-chip[11]. The combined dataset from the two ChIA-PET libraries identified 4,124 binding sites. The combined ChIP-PET library from this study and a previous one[1] found 1,297 binding sites. The ChIP-chip experiment found 3,260 binding sites[11]. Of the 1,297 ChIP-PET binding sites, 9332 (72%) overlapped with the ChIA-PET study, whereas only 27 sites (0.65%) overlapped with the ChIP-chip data solely. Of the 3,260 sites in the ChIP-chip study, about half overlapped with the ChIA-PET data. **c.** Distribution of ERE motif in ER-α binding sites identified by ChIA-PET, ChIP-PET, and ChIP-chip data ("ChIA-PET 1" = IHM001H, "ChIA-PET 2" = IHM001N, "ChIP-PET 1" = IHM043, "ChIP-PET 2" = SHC007[1], "ChIP-chip" = a previously published ChIP-chip study[11]).



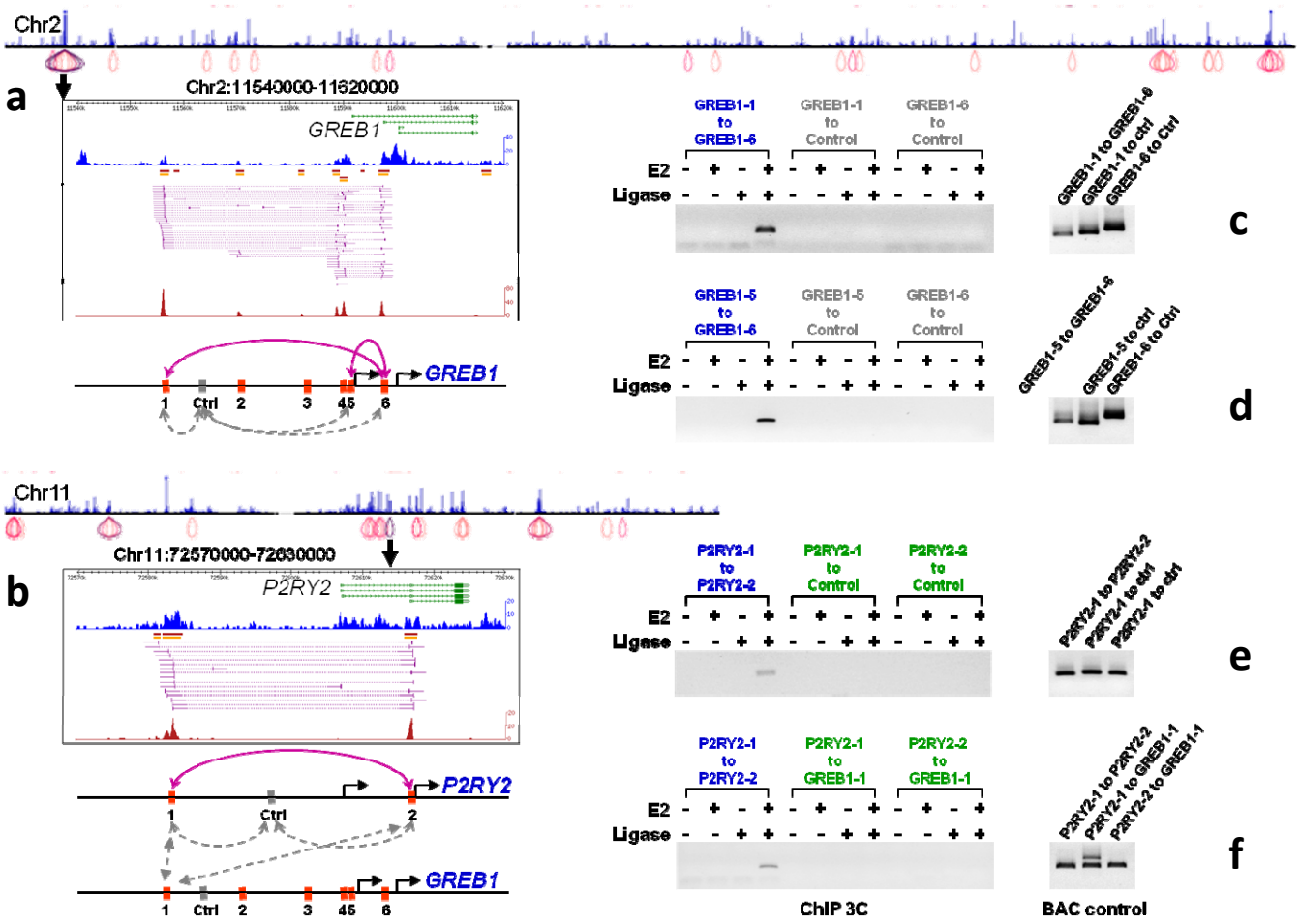
Supplementary Figure 4. Schematic view of linker nucleotide barcoding for ChIA-PET analysis

a. Schematic view of the use of linker nucleotide barcoding for ChIA-PET analysis. Linker A and B are added to different aliquots of same ChIP material. After linker ligation, the two aliquots are mixed for diluted proximity ligation and subsequent PET sequencing analysis. The PET sequences with linker composition of A and B (AB) are considered as having derived from undesired chimeric ligation products between two different chromatin complexes.



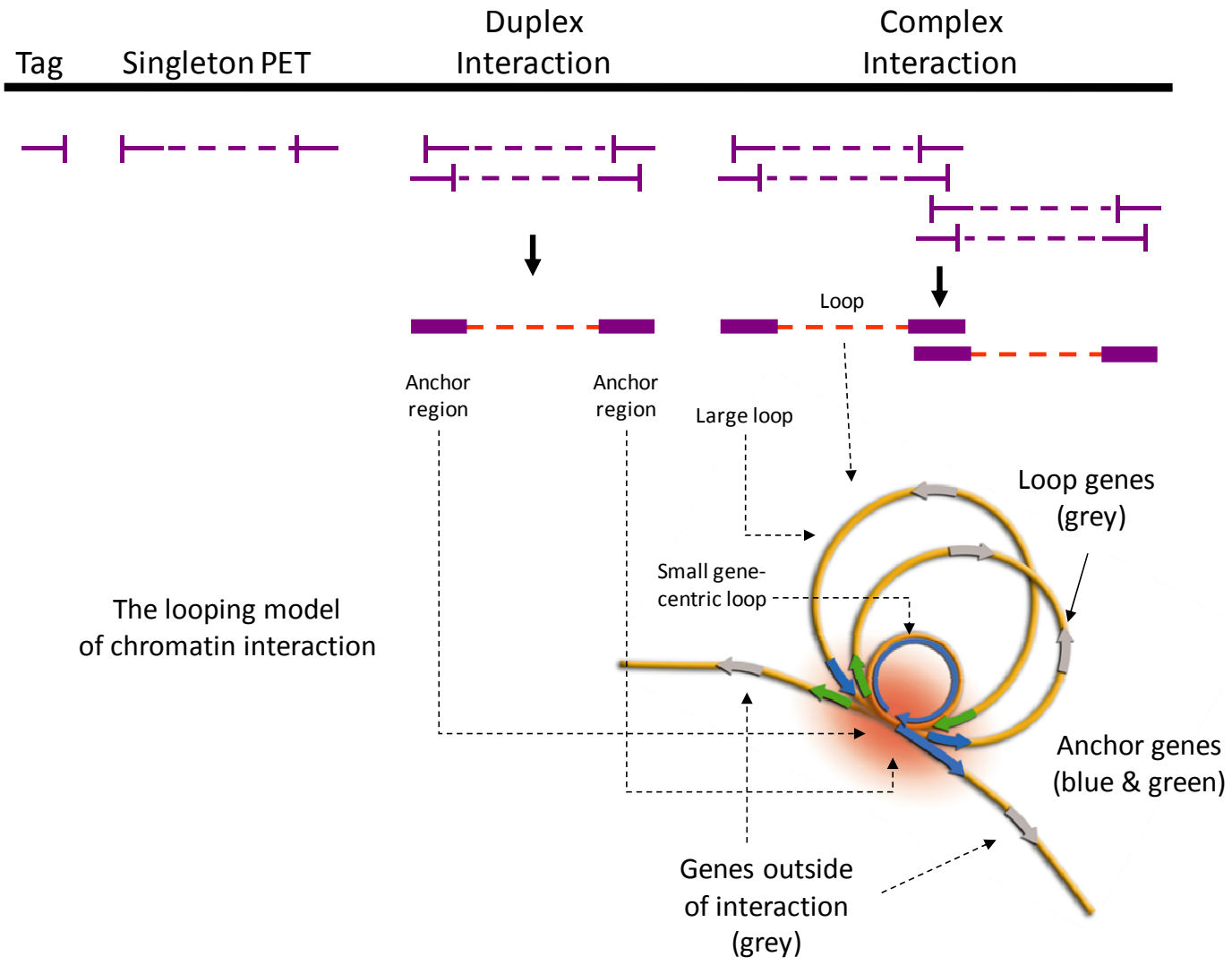
Supplementary Figure 4. Schematic view of linker nucleotide barcoding for ChIA-PET analysis

b. An example of ChIA-PET with specific ligation PETs (AA and BB linker nucleotide barcode composition) and chimeric non-specific ligation PETs (AB linker nucleotide barcode composition) mapped at the *P2RY2* locus. The AA and BB data showed abundant self-ligation PET peaks for two ER-αBS and an inter-ligation PET cluster indicating interactions between the two sites, while the AB data showed no self-ligation PETs and only scattered inter-ligation PETs.



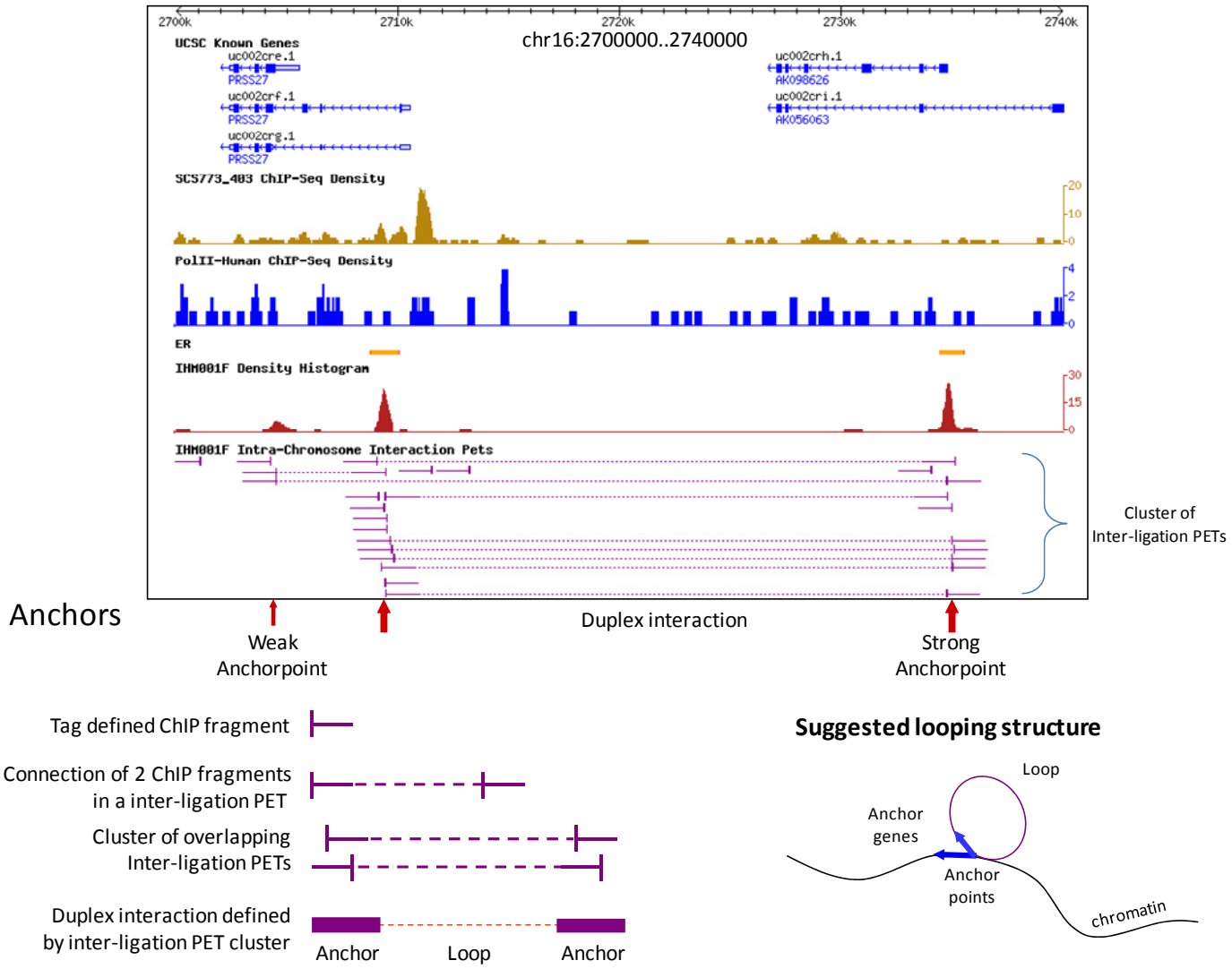
Supplementary Figure 5. ChIP-3C to analyze possible ChIP enrichment biases for interactions

a. ER-α ChIA-PET mapping on chromosome 2. ER-α binding sites are shown as blue vertical bar and ER-α-mediated interactions are shown as purple rings. The zoom-in view on the *KIAA0575* (*GREB1*) locus is shown in an 80 Kb window. Six ER-α binding sites were identified in this region. The binding sites #1, #5 and #6 were selected to represent the long (40 Kb) and short (8 Kb) distance of interactions for ChIP-3C validation tests (purple arrowed lines). In addition, internal non-interacting sites were chosen to be negative controls of ChIP-3C experiments (grey dotted arrowed lines). **b.** ER-α ChIA-PET mapping on chromosome 11 and the zoom-in view on the *P2RY2* locus. Two ER-α binding sites were identified in this region. The binding sites #1 and #2 were tested by ChIP-3C experiments (purple arrowed line). Internal non-interacting sites were included in the ChIP-3C experiments as negative controls (grey dotted arrowed lines). **c.** Result of ChIP-3C analysis between the *KIAA0575* (*GREB1*) binding sites #1 and #6 with negative controls and positive controls. **d.** Result of ChIP-3C between the KIAA0575 (*GREB1*) binding sites #5 and #6 with controls. **e**. Result of ChIP-3C between the *P2RY2* binding sites #1 and #2 with controls. **f.** Result of ChIP-3C between the *KIAA0575* (*GREB1*) and the *P2RY2* loci. Positive controls of 3C PCR reactions using various primers were tested with digested, mixed and ligated BAC clones from the *KIAA0575* (*GREB1*) and *P2RY2* regions.



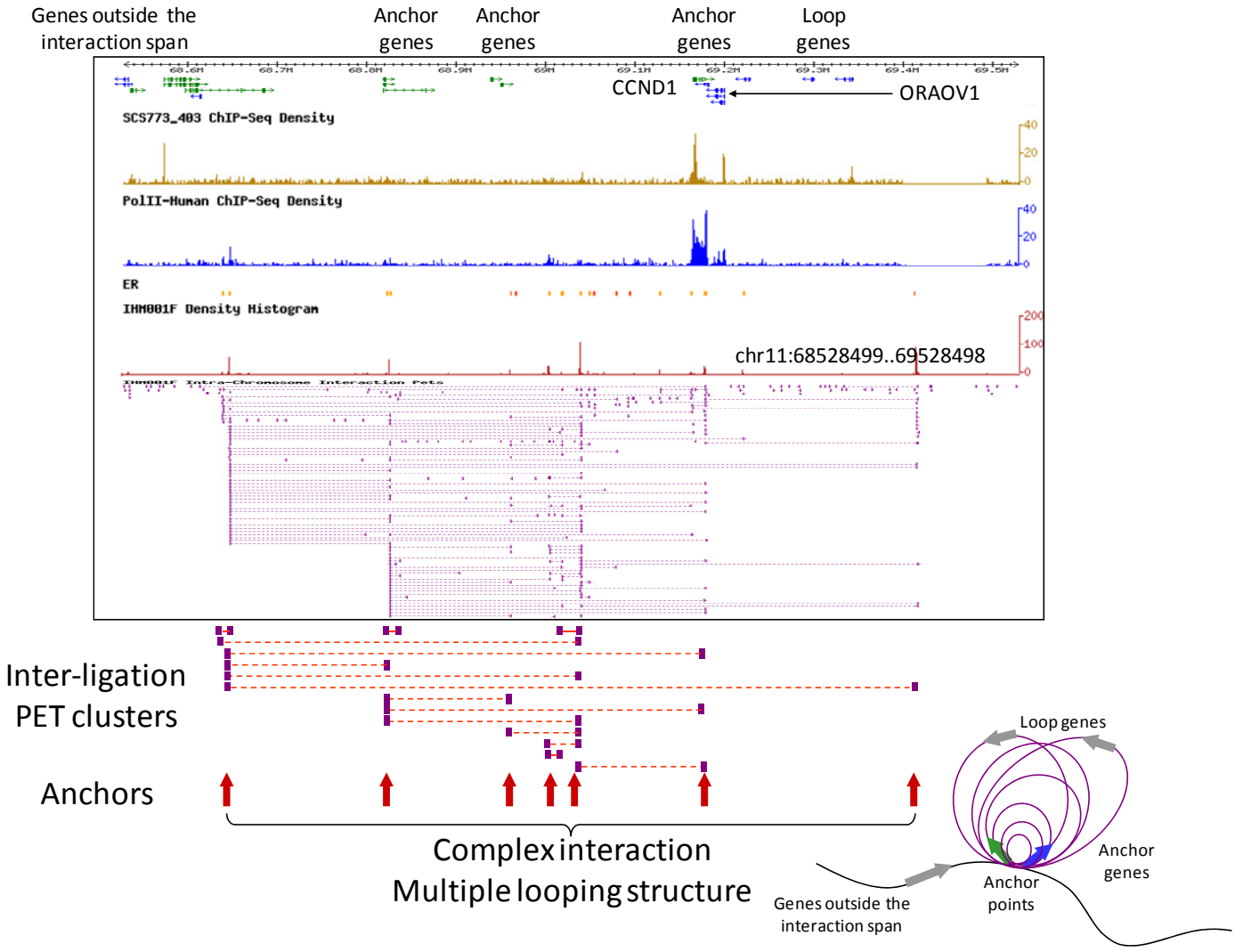
Supplementary Figure 6. Illustration of structural components of ER-α-bound chromatin interactions

a. Structures of tags, PETs, duplex interactions, complex interactions, anchors, loops, and genes associated with interactions.



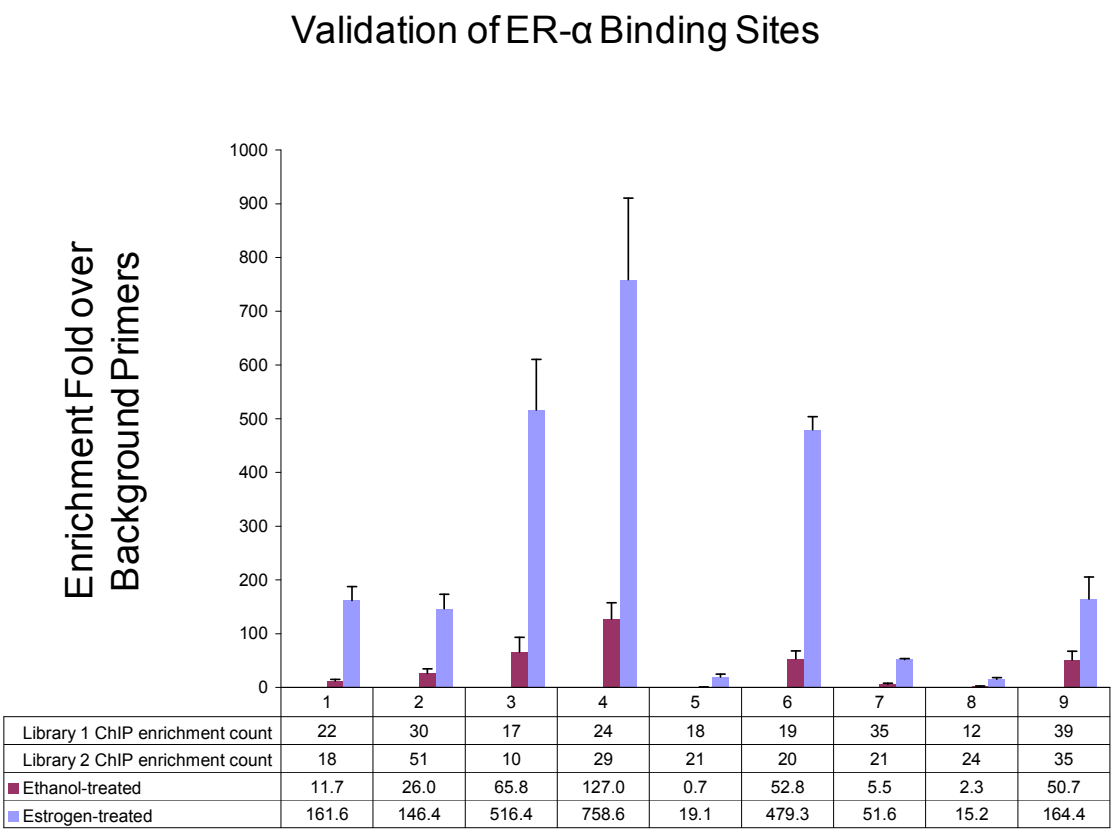
Supplementary Figure 6. Illustration of structural components of ER-α-bound chromatin interactions

b. An example of a duplex interaction structure



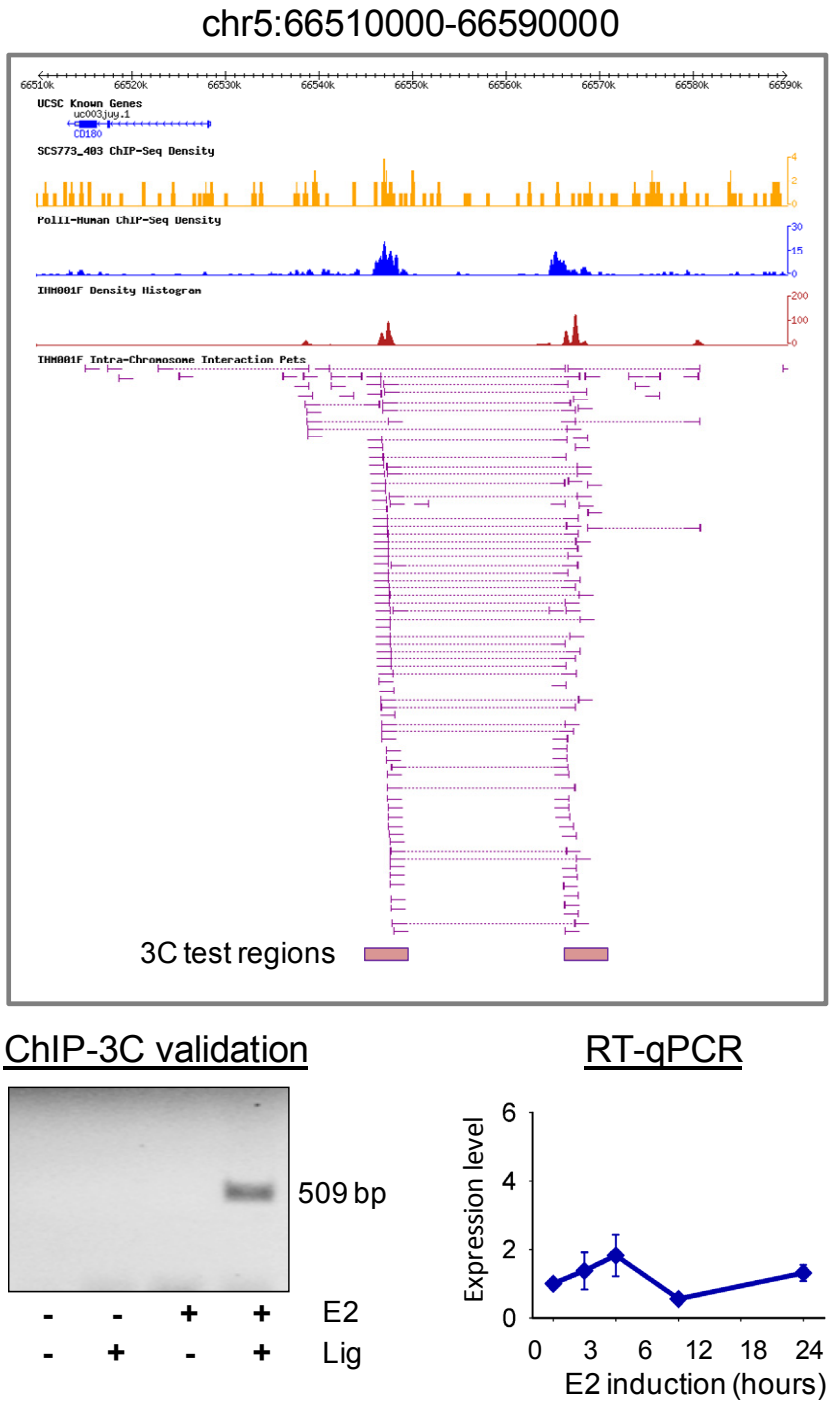
Supplementary Figure 6. Illustration of structural components of ER-α-bound chromatin interactions

c. An example of a complex interaction structure



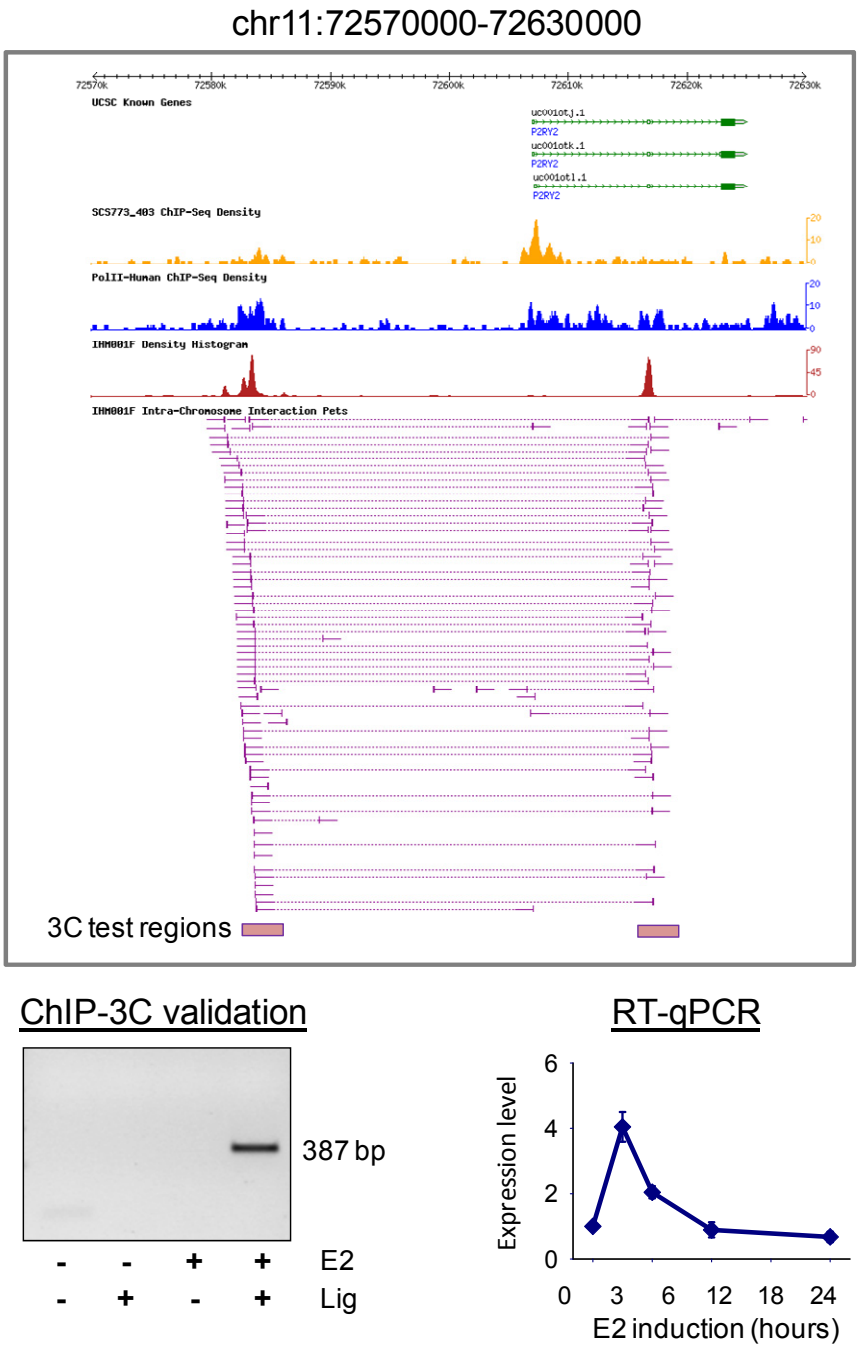
	1	2	3	4	5	6	7	8	9
Library 1 ChIP enrichment count	22	30	17	24	18	19	35	12	39
Library 2 ChIP enrichment count	18	51	10	29	21	20	21	24	35
Ethanol-treated	11.7	26.0	65.8	127.0	0.7	52.8	5.5	2.3	50.7
Estrogen-treated	161.6	146.4	516.4	758.6	19.1	479.3	51.6	15.2	164.4

Supplementary Figure 7. ChIP-qPCR validation of new ER-αBS identified by ChIA-PET in this study



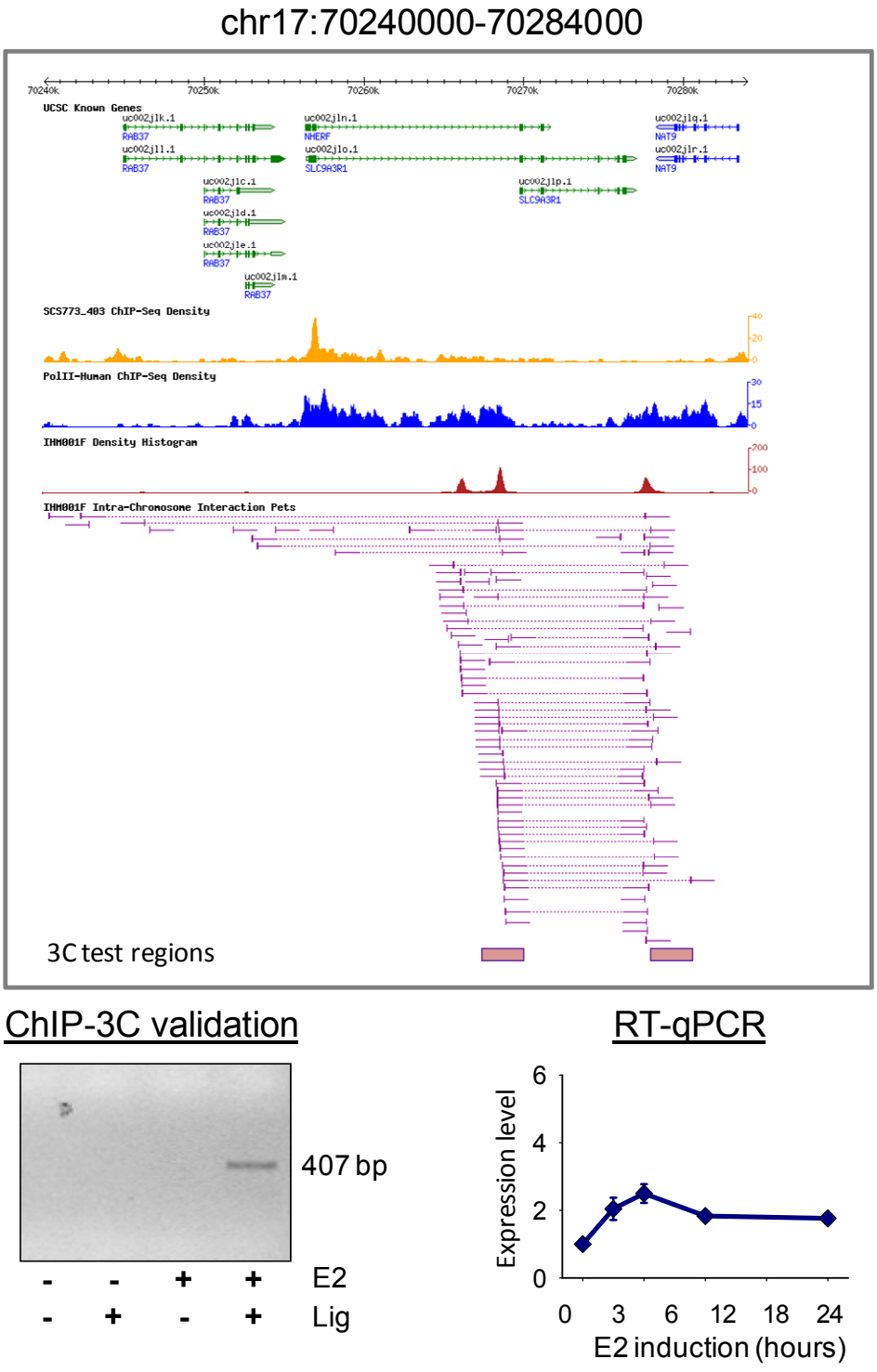
Supplementary Figure 8. ChIP-3C and RT-qPCR validations

a. Screenshot, ChIP-3C and RT-qPCR at *LY64*. ChIP-3C experiments were done with oestrogen treated (E2+) and untreated control (E2-) MCF-7 cells. The results of RT-qPCR on mRNA over a time course of oestrogen induction show that the ChIA-PET associated genes are oestrogen-regulated.



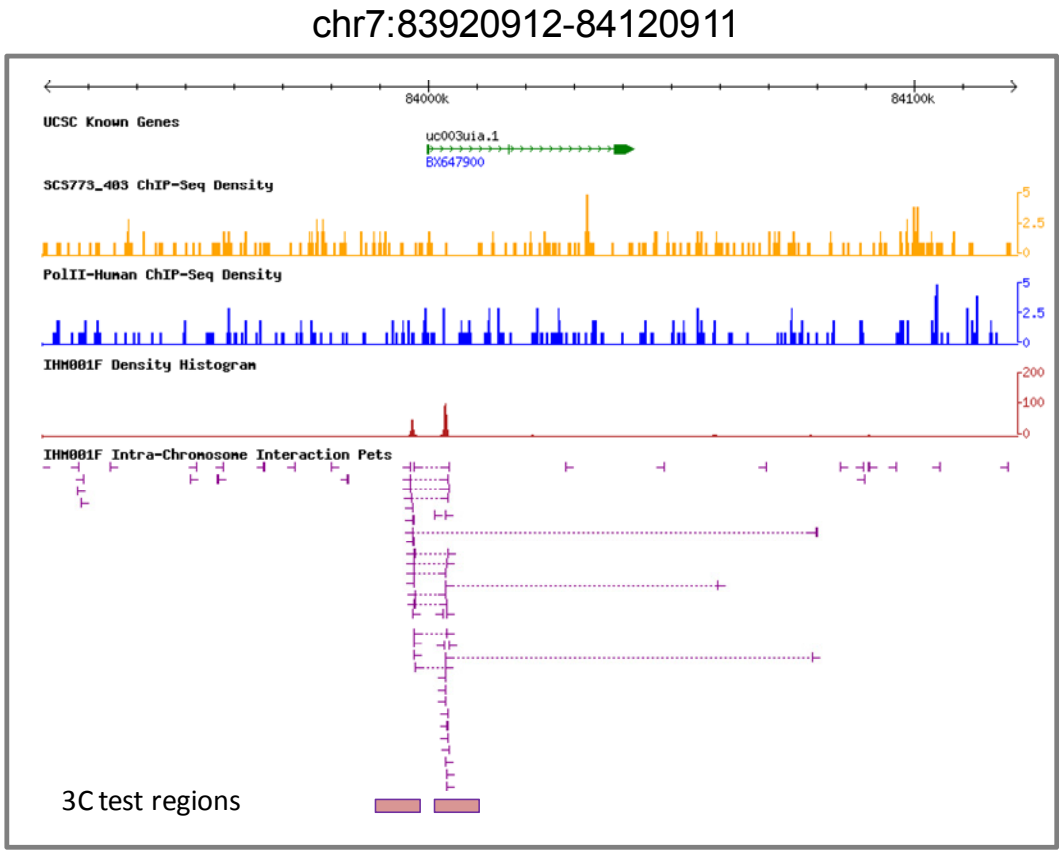
Supplementary Figure 8. ChIP-3C and RT-qPCR validations

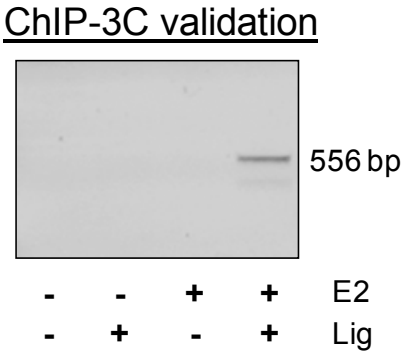
b. Screenshot, ChIP-3C and RT-qPCR at *P2RY2*. ChIP-3C experiments were done with oestrogen treated (E2+) and untreated control (E2-) MCF-7 cells. The results of RT-qPCR on mRNA over a time course of oestrogen induction show that the ChIA-PET associated genes are oestrogen-regulated.



Supplementary Figure 8. ChIP-3C and RT-qPCR validations

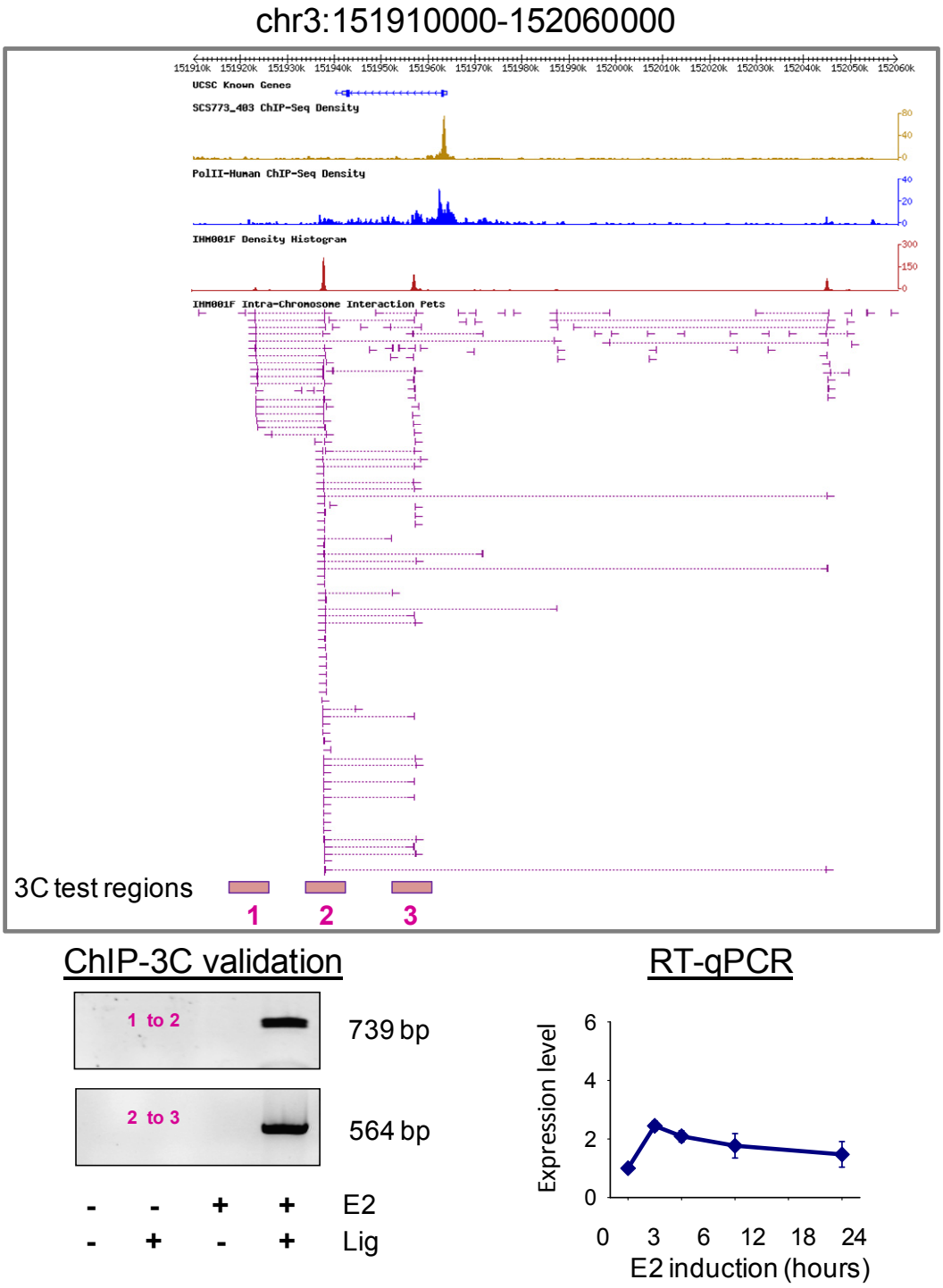
c. Screenshot, ChIP-3C and RT-qPCR at *SLC9A3R1*. ChIP-3C experiments were done with oestrogen treated (E2+) and untreated control (E2-) MCF-7 cells. The results of RT-qPCR on mRNA over a time course of oestrogen induction show that the ChIA-PET associated genes are oestrogen-regulated.





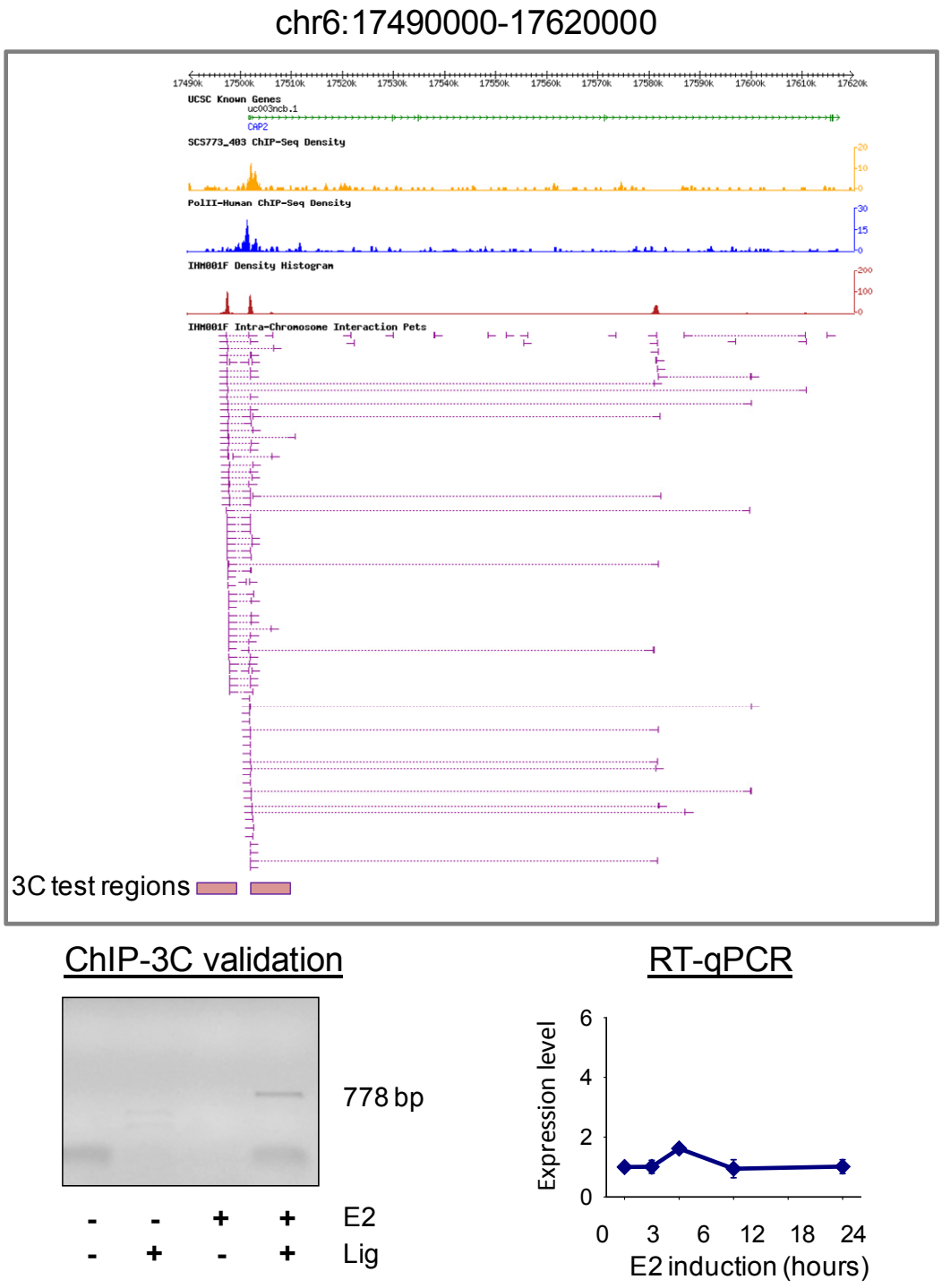
Supplementary Figure 8. ChIP-3C and RT-qPCR validations

d. Screenshot and ChIP-3C at *BX647900*. ChIP-3C experiments were done with oestrogen treated (E2+) and untreated control (E2-) MCF-7 cells. RT-qPCR was difficult to perform on this site. Perhaps, the gene is expressed at a very low level if at all.



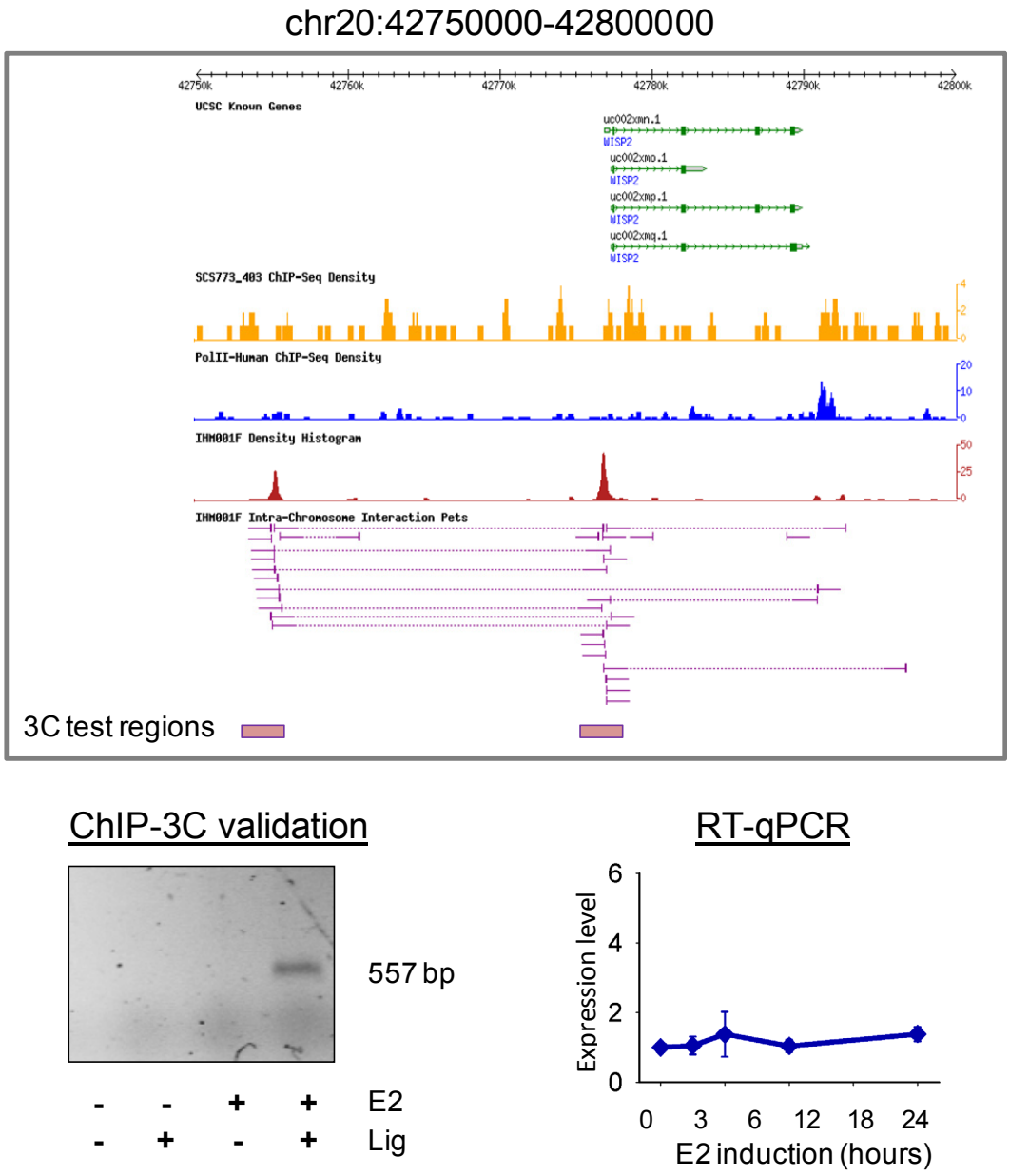
Supplementary Figure 8. ChIP-3C and RT-qPCR validations

e. Screenshot, ChIP-3C and RT-qPCR at *SIAH2*. ChIP-3C experiments were done with oestrogen treated (E2+) and untreated control (E2-) MCF-7 cells. The results of RT-qPCR on mRNA over a time course of oestrogen induction show that the ChIA-PET associated genes are oestrogen-regulated.



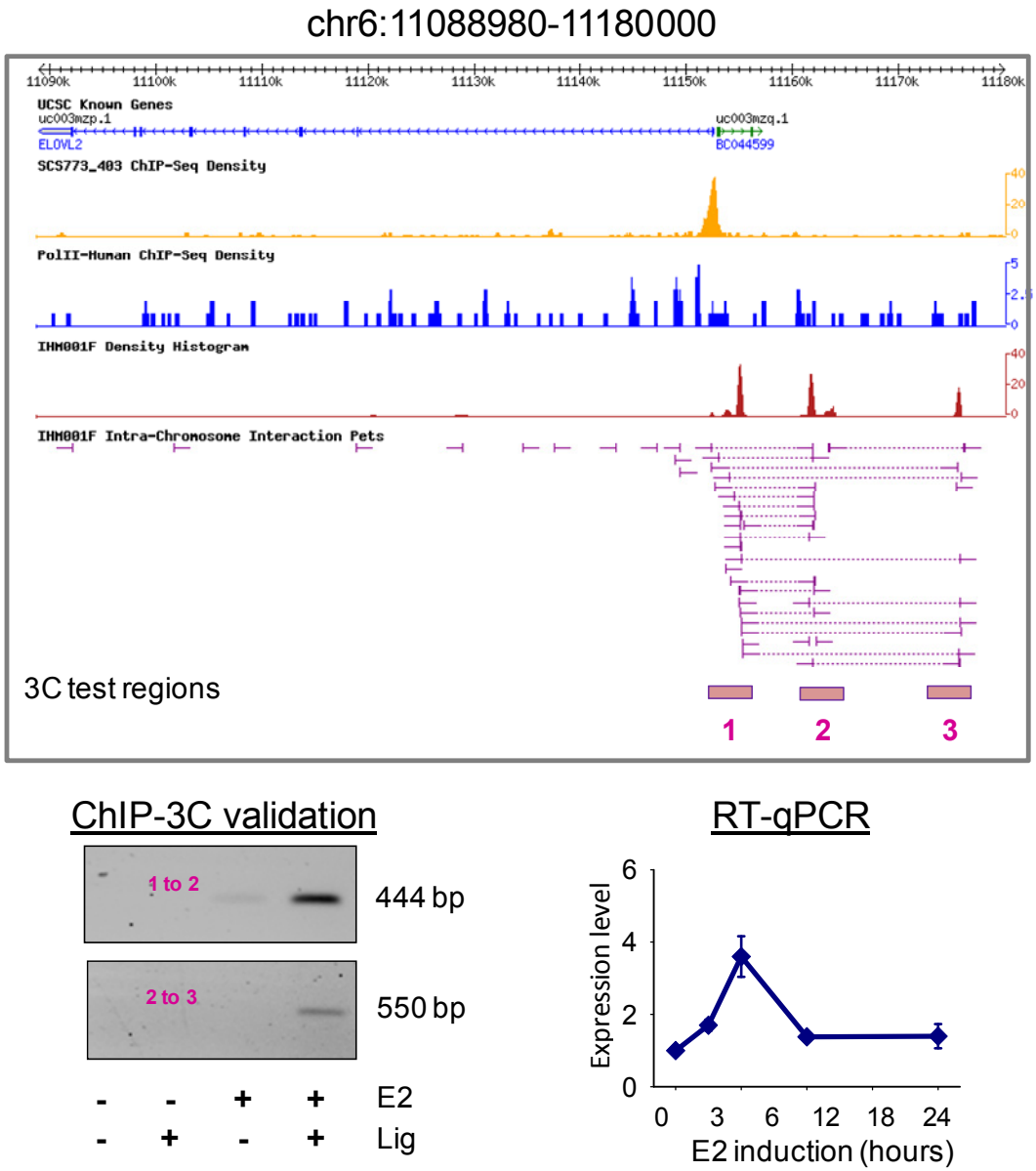
Supplementary Figure 8. ChIP-3C and RT-qPCR validations

f. Screenshot, ChIP-3C and RT-qPCR at *CAP2*. ChIP-3C experiments were done with oestrogen treated (E2+) and untreated control (E2-) MCF-7 cells. The results of RT-qPCR on mRNA over a time course of oestrogen induction show that the ChIA-PET associated genes are oestrogen-regulated.



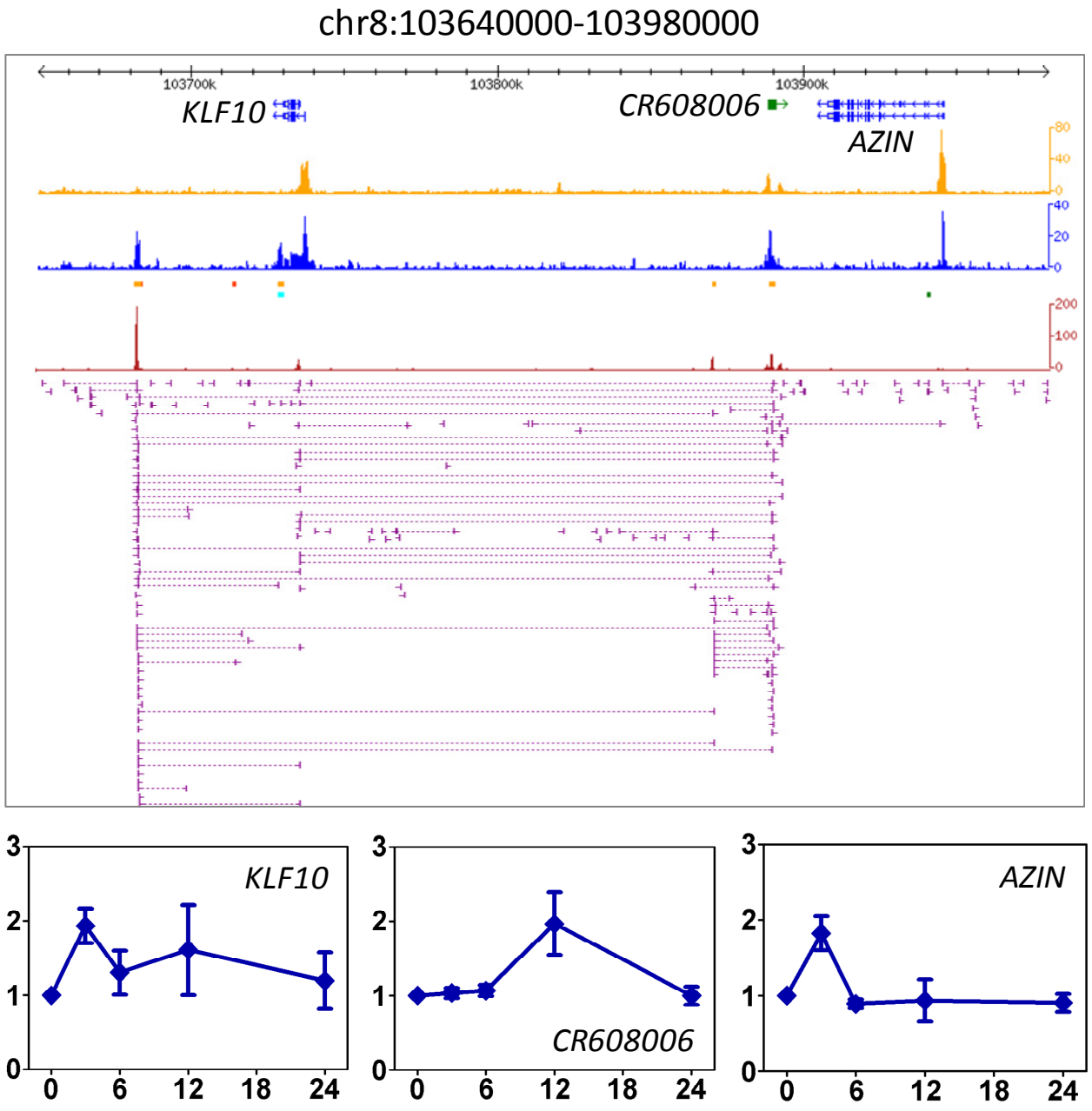
Supplementary Figure 8. ChIP-3C and RT-qPCR validations

g. Screenshot, ChIP-3C and RT-qPCR at *WISP2*. ChIP-3C experiments were done with oestrogen treated (E2+) and untreated control (E2-) MCF-7 cells. The results of RT-qPCR on mRNA over a time course of oestrogen induction show that the ChIA-PET associated genes are oestrogen-regulated. Note: For clarity, UCSC Genes *AK090842* and *CR597563*, which go across the screen, have been removed from the display.



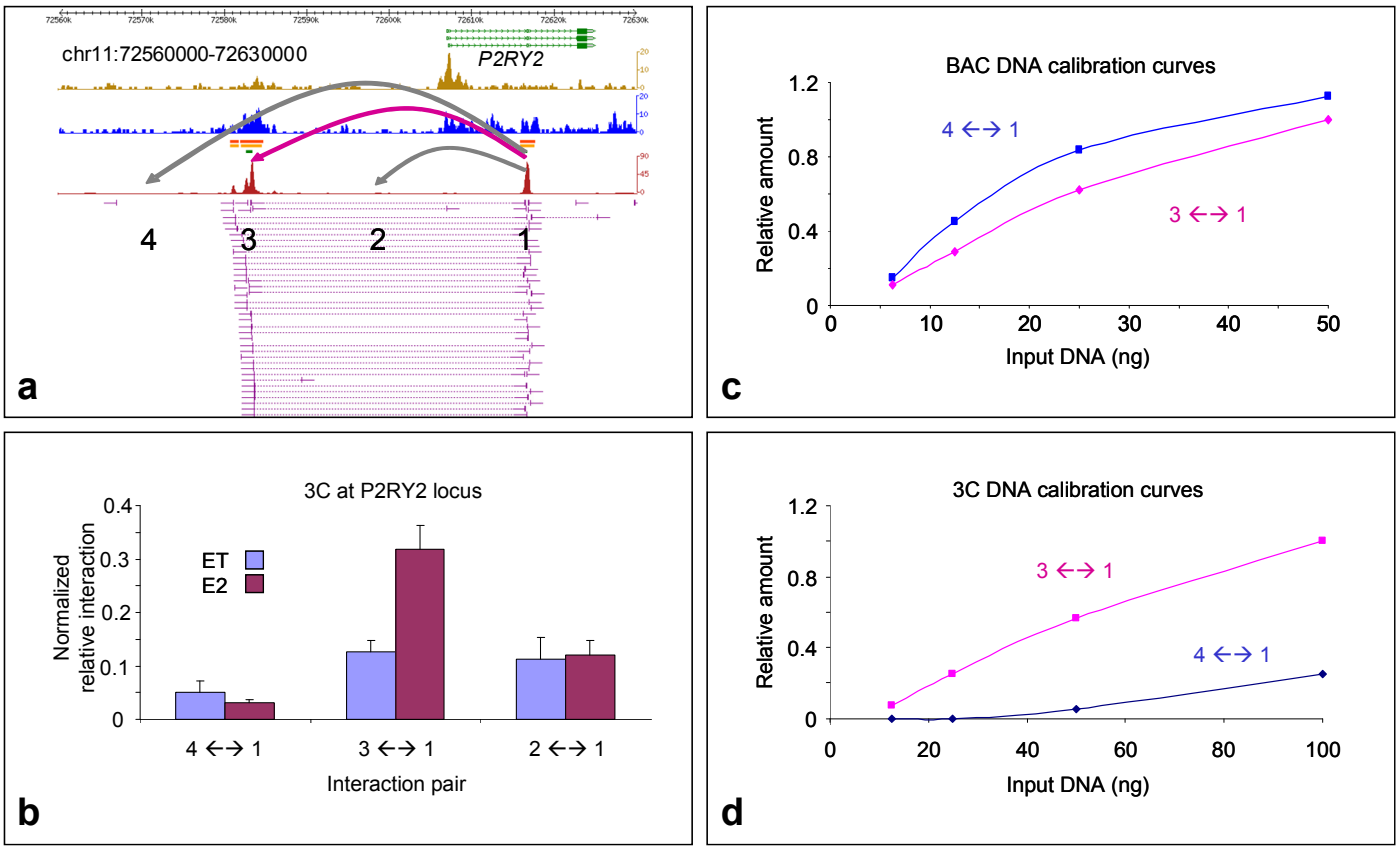
Supplementary Figure 8. ChIP-3C and RT-qPCR validations

h. Screenshot, ChIP-3C and RT-qPCR at *ELOVL2*. ChIP-3C experiments were done with oestrogen treated (E2+) and untreated control (E2-) MCF-7 cells. The results of RT-qPCR on mRNA over a time course of oestrogen induction show that the ChIA-PET associated genes are oestrogen-regulated.



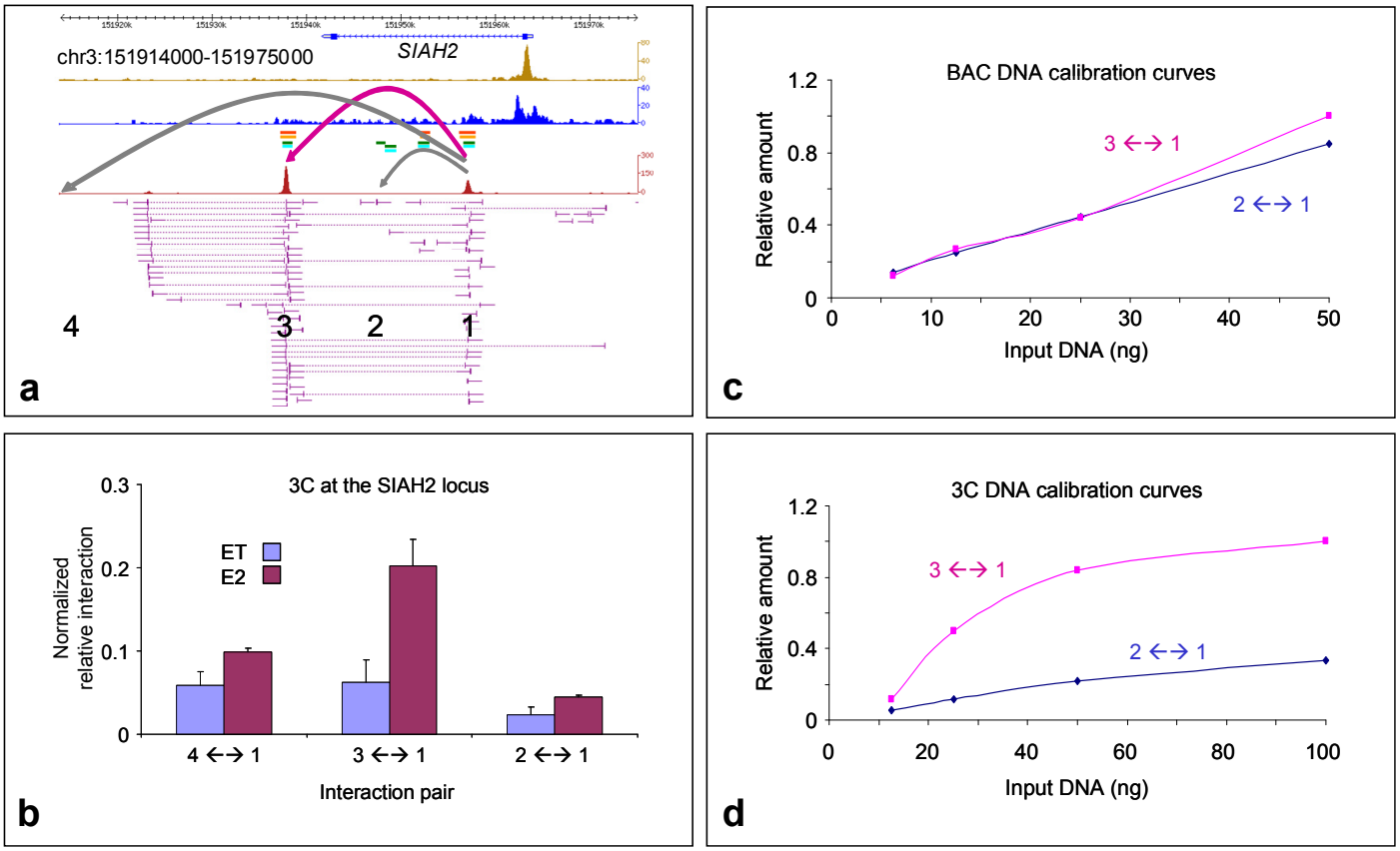
Supplementary Figure 8. ChIP-3C and RT-qPCR validations

i. Screenshot of a complex interaction region at the *KLF10, CR608006*, and *AZIN* loci. The results of RT-qPCR on mRNA over a time course of oestrogen induction show that the ChIA-PET associated genes are oestrogen-regulated.



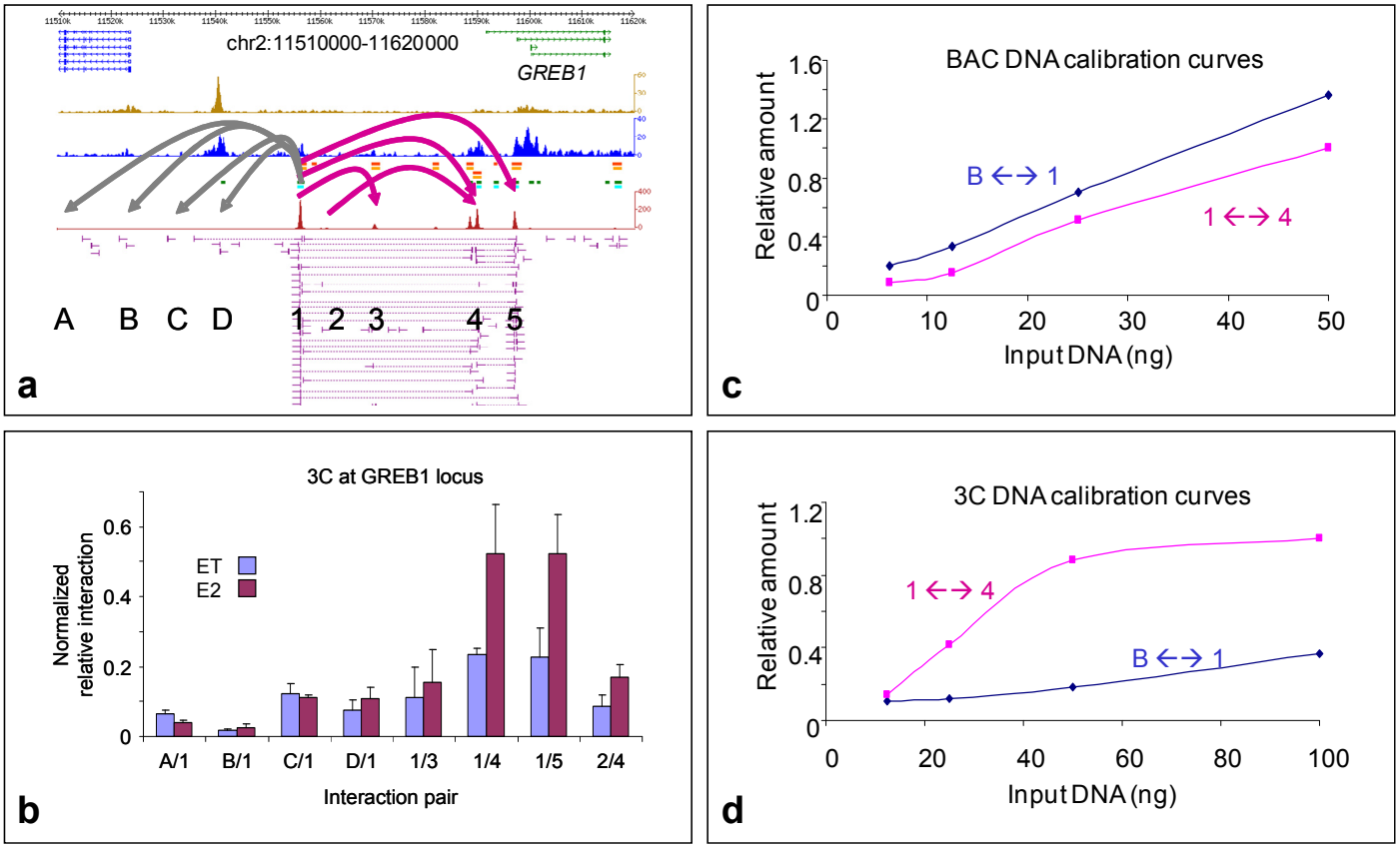
Supplementary Figure 9. 3C validations

a. 3C validation at the *P2RY2* locus. (a-a) ER-αBS and interactions identified by ER-α ChIA-PET data (concise screenshot of ChIA-PET genome browser). Four locations were chosen as 3C testing sites. Location #2 and #4 are negative control sites. (a-b) is 3C experiment result. (a-c) and (a-d) are calibrations to identify the linear range for PCR on (c) BAC clone DNA controls and (d) the 3C ligation DNA template. The 3C results confirm the interaction at *P2RY2* and show the interaction is oestrogen-dependent.



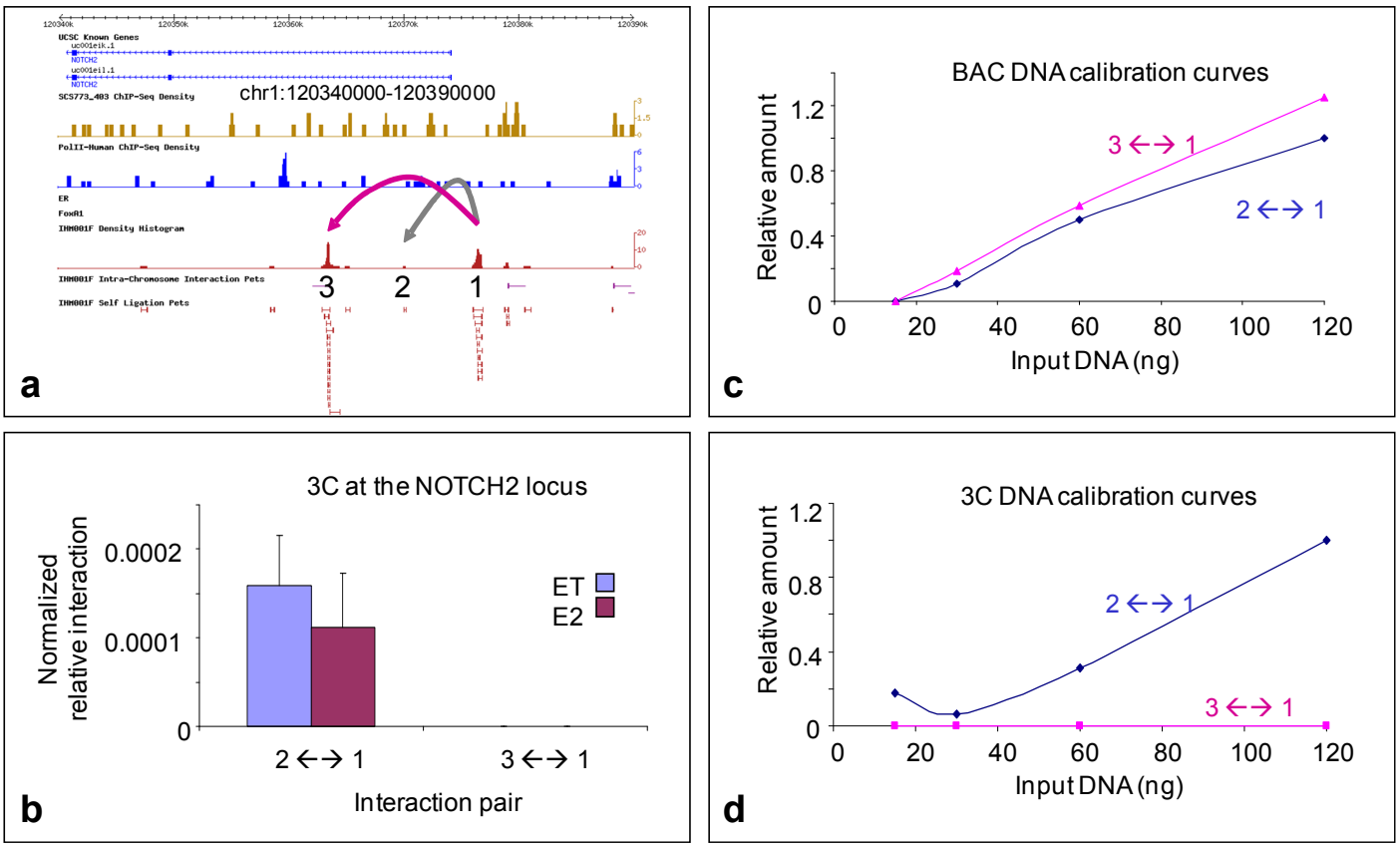
Supplementary Figure 9. 3C validations

b. 3C validation at the *SIAH2* locus. (b-a) ER-αBS and interactions identified by ER-α ChIA-PET data (concise screenshot of ChIA-PET genome browser). Four locations were chosen as 3C testing sites. Location #2 and #4 are negative control sites. (b-b) is 3C experiment result. (b-c) and (b-d) are calibrations to identify the linear range for PCR on (b-c) BAC clone DNA controls and (b-d) the 3C ligation DNA template. The 3C results confirm the interaction at *SIAH2* and show the interaction is oestrogen-dependent.



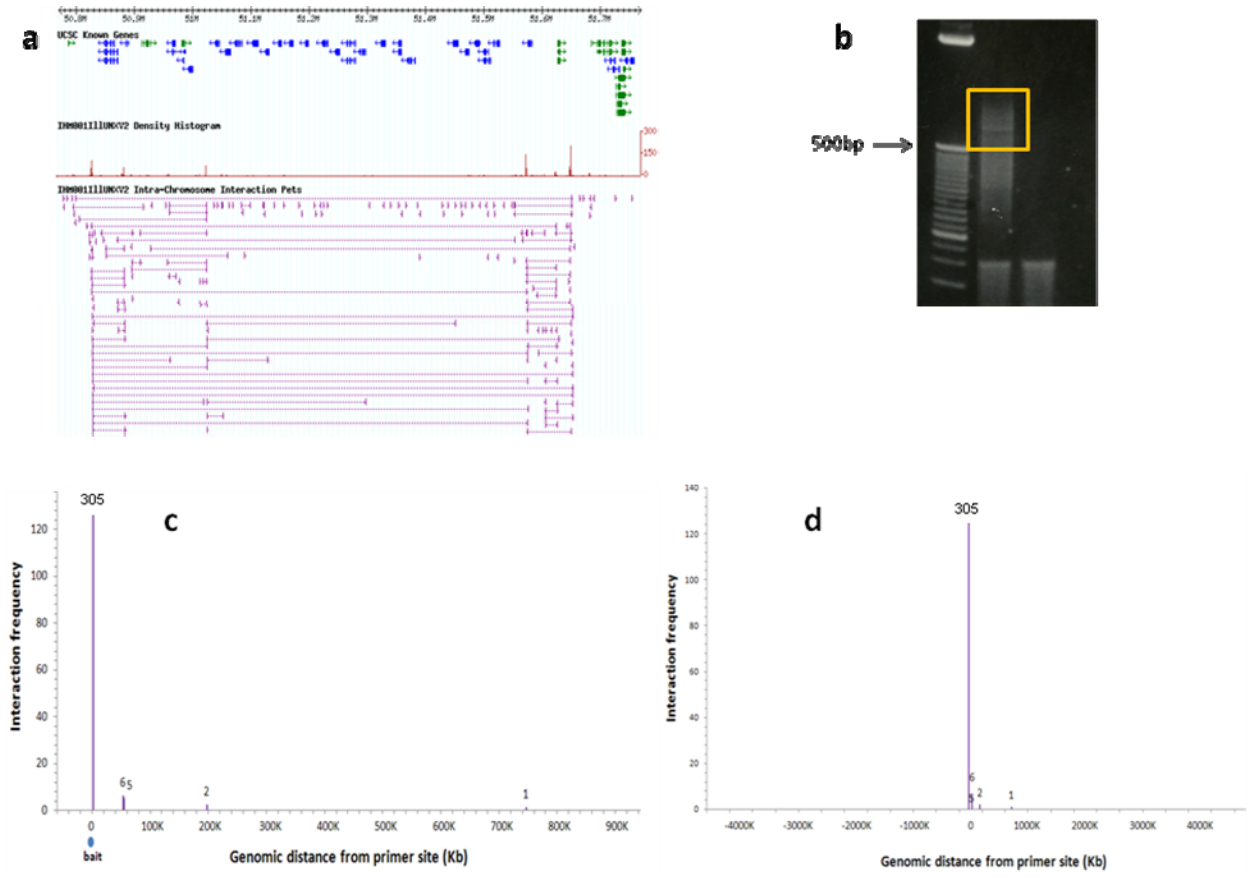
Supplementary Figure 9. 3C validations

c. 3C validation at the *GREB1* locus. (c-a) ER-αBS and interactions identified by ER-α ChIA-PET data (concise screenshot of ChIA-PET genome browser). Multiple locations were chosen as 3C testing sites in a scanning 3C experiment. Locations #1-#5 are experimental sites, and locations A-D are negative control sites. (c-b) is 3C experiment result. (c-c) and (c-d) are calibrations to identify the linear range for PCR on (c-c) BAC clone DNA controls and (c-d) the 3C ligation DNA template. The 3C results confirm the interaction at *GREB1* and show the interaction is oestrogen-dependent.



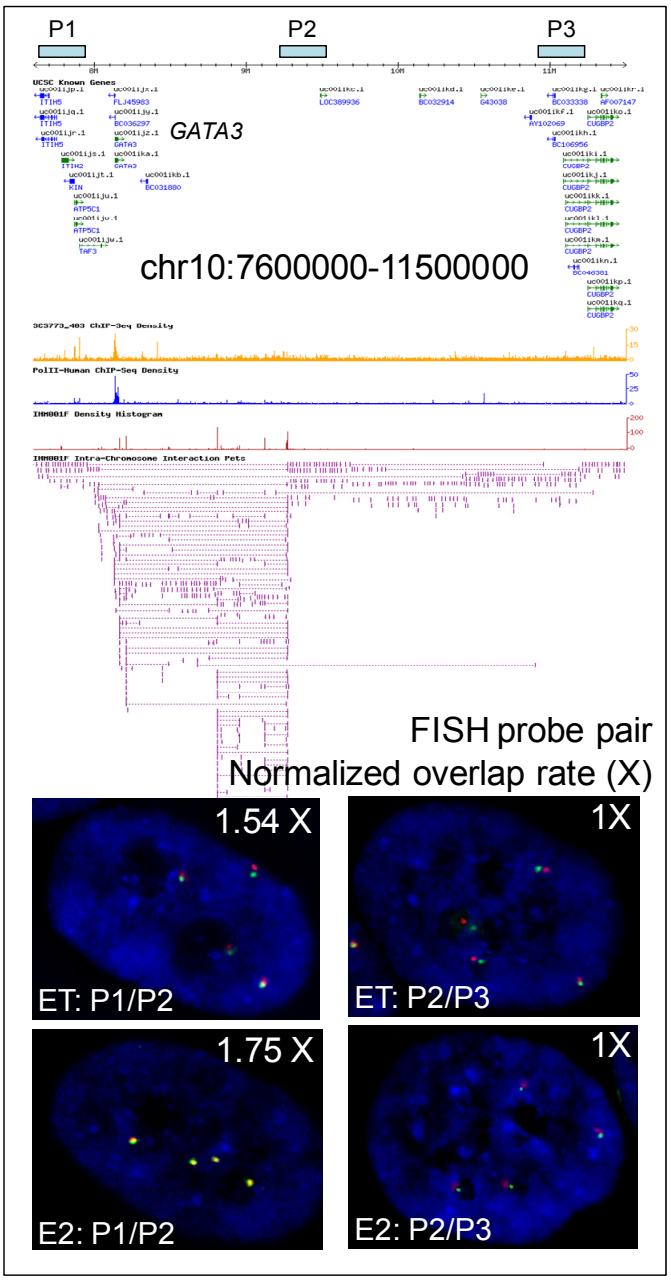
Supplementary Figure 9. 3C validations

d. 3C validation at the *NOTCH2* locus, a negative control site where we see ER-αBSs but not interactions in the ChIA-PET data. (d-a) ER-αBS but no interactions identified by ER-α ChIA-PET data. Three locations were chosen as 3C testing sites. Location #2 is a negative control site. (d-b) is 3C experiment result. (d-c) and (d-d) are calibrations to identify the linear range for PCR on (d-c) BAC clone DNA controls and (d-d) the 3C ligation DNA template. The 3C results confirm there is no interaction at *NOTCH2*. "3 ←→ 1" is very low as to be undetectable, and "2 ←→ 1" is also very low (0.00015 normalized relative interaction), indicating that there are no interactions at *NOTCH2*, confirming the lack of interactions seen in the ChIA-PET data.



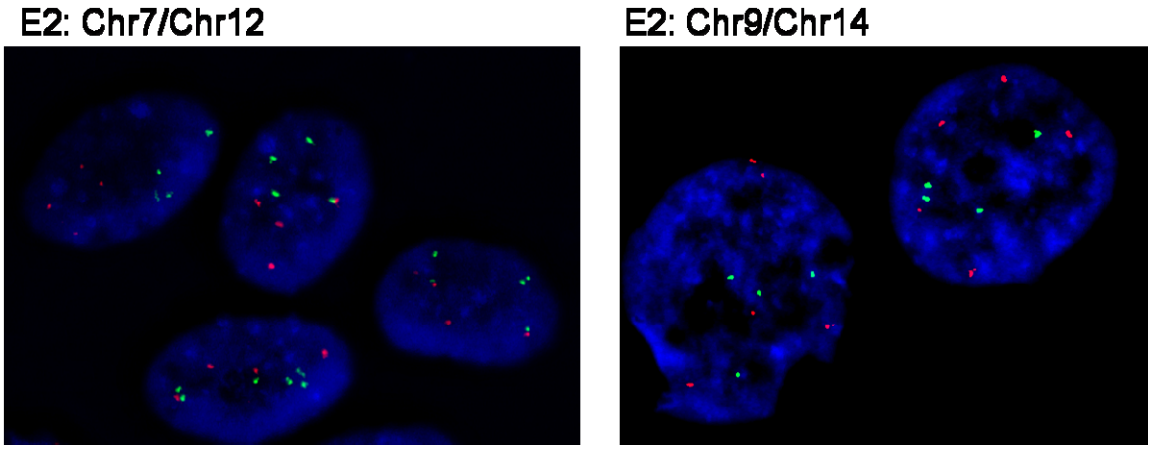
Supplementary Figure 10. 4C validations

a. Chromatin interactions at the keratin gene cluster identified by ER-α ChIA-PET analysis. **b.** The 4C PCR products using the "bait" primer pair based at the keratin chromatin interaction region. The boxed range of DNA amplicon was gel-excised for sequencing analysis. **c.** The 4C sequences mapped at the keratin gene cluster locus. The highest 4C sequence mapping peak is at the 4C "bait" site (indicated by a blue dot). The interaction anchors of this interaction complex were mapped 4C sequences. **d.** An enlarged view (10 Mb) of 4C sequence mapping centered at the keratin gene cluster shows that the 4C data is very clean.



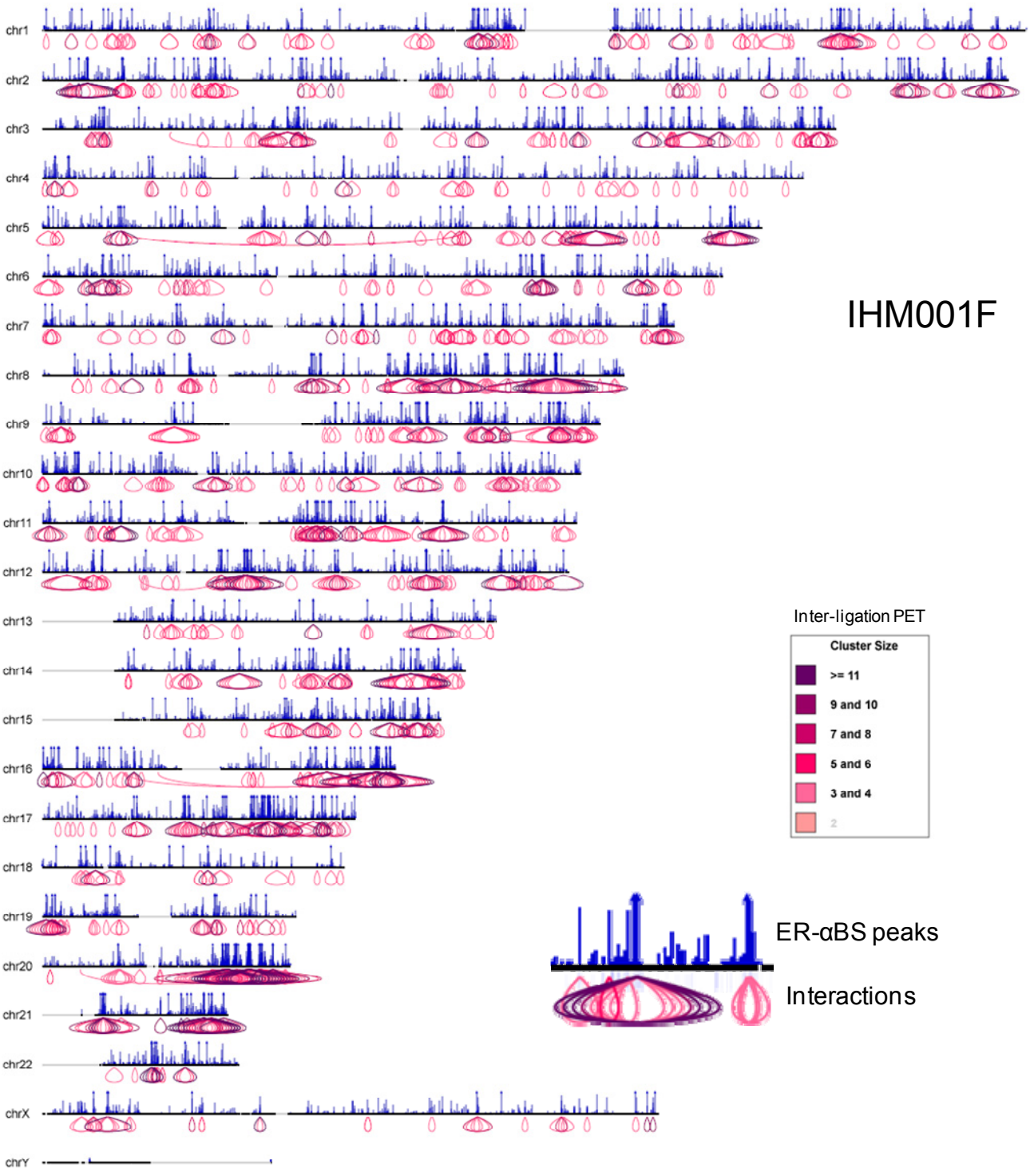
Supplementary Figure 11. FISH validation

a. The intrachromosomal chromatin interaction at the *GATA3* locus was validated by FISH experiments. P1 and P2 are test BAC probes near the two anchors of the interaction (>1 Mb), and P3 is a control probe. The FISH experiments using the combined probes of P2/P1 and P3/P2 were done in ethanol control (ET) and oestrogen treated (E2) MCF-7 cells. Each FISH analyzed 100-200 nuclei, and all spots in each nucleus were counted. The FISH images of the probe pairs show red and green spots when the probes are separated, and yellow sections between red and green spots when the probes overlap. The probe overlap rate is normalized using the control probe pair (P2/P3) as the base level of background noise.



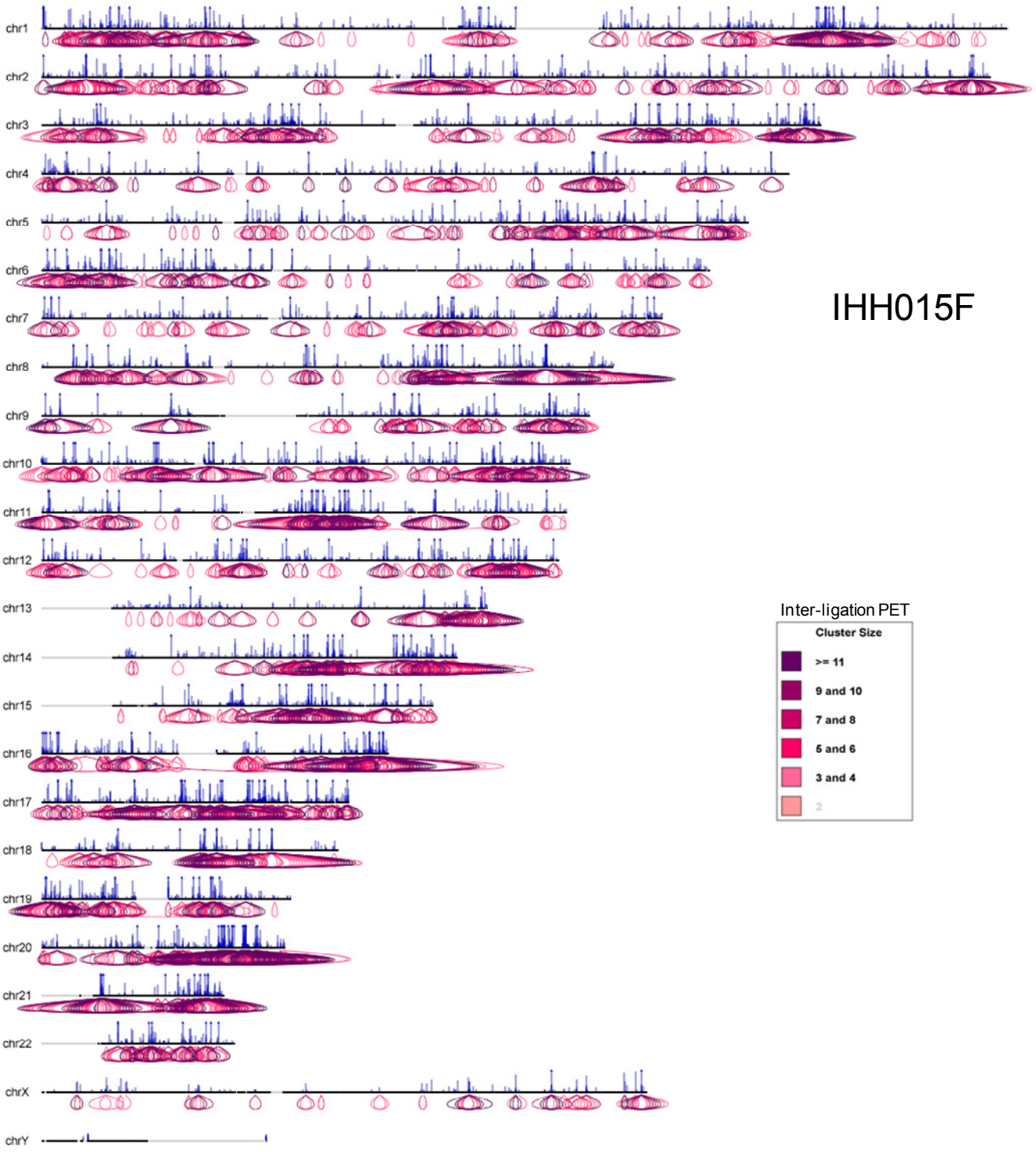
Supplementary Figure 11. FISH validation

b. The two interchromosomal chromatin interactions at Chr7/Chr12 and Chr9/Chr14 could not be validated by FISH experiments. The FISH images of the probes in oestrogen-treated cells do not show colocalization.



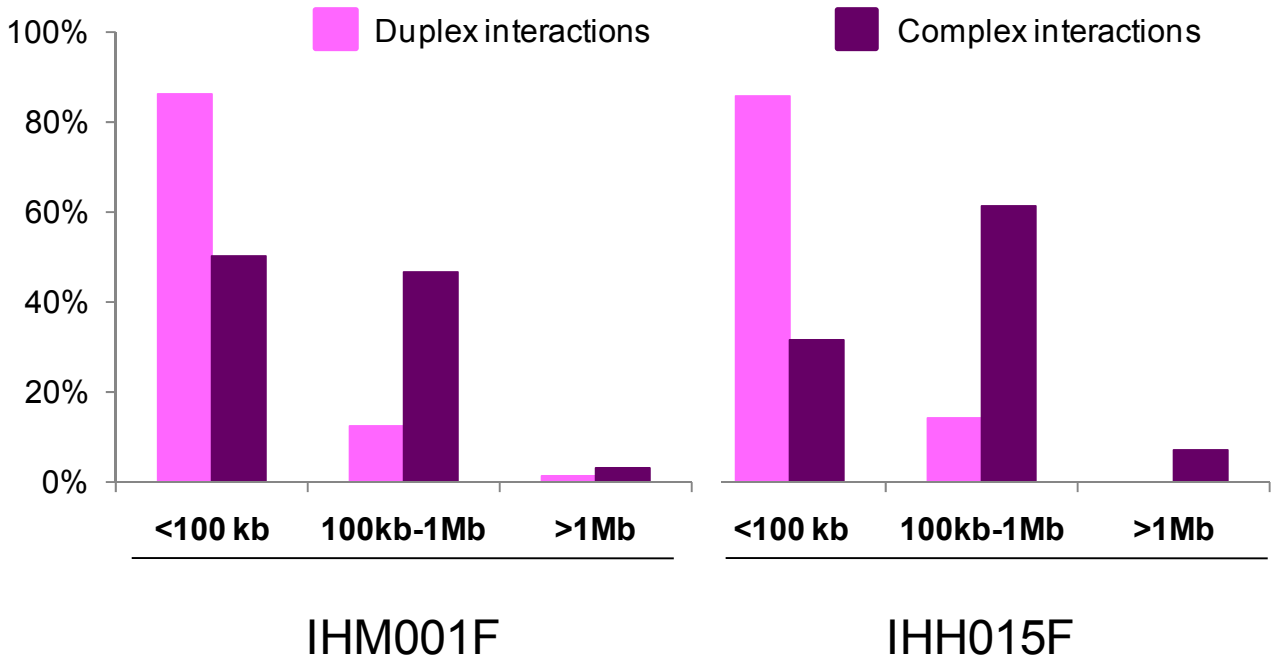
Supplementary Figure 12. Whole genome view of ER-α-bound human chromatin interactome

a. Whole genome view of ChIA-PET library IHM001F



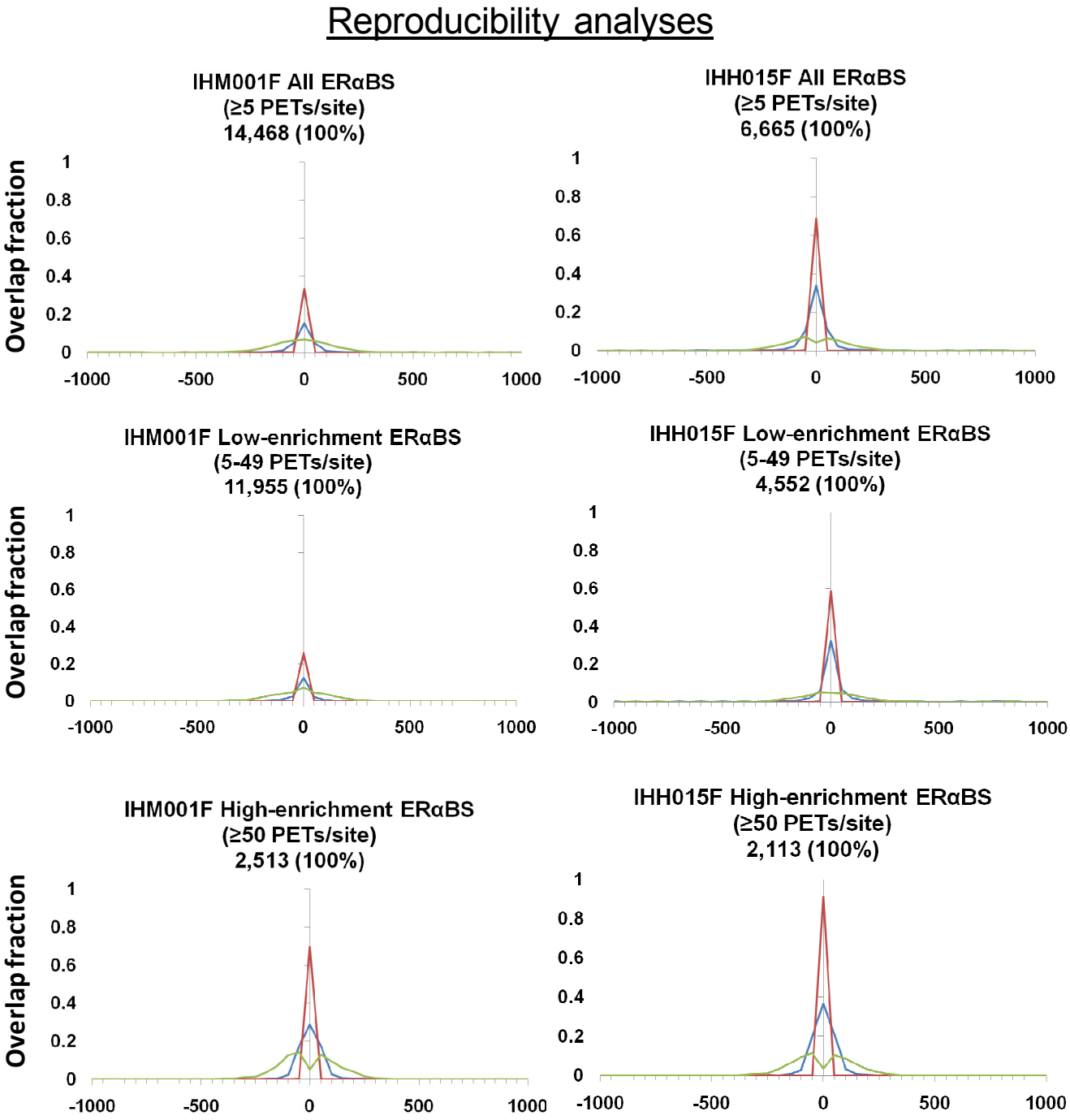
Supplementary Figure 12. Whole genome views of ER-α-bound human chromatin interactome

b. Whole genome view of ChIA-PET library IHH015F



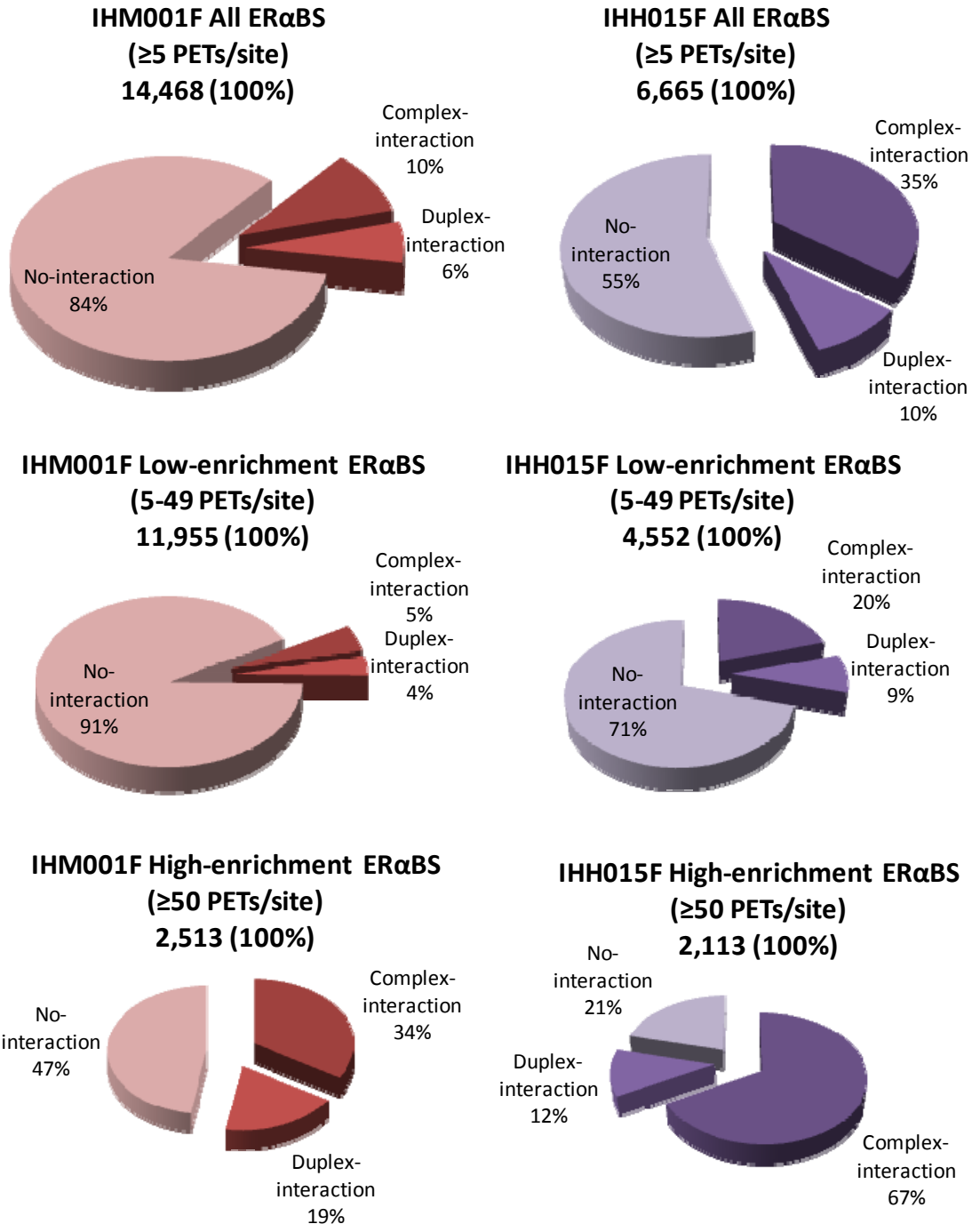
Supplementary Figure 12. Whole genome views of ER-α-bound human chromatin interactome

c. Genomic span plots showing the percentages of intrachromosomal interaction regions in different bin sizes of the genomic lengths covered by the interaction regions in IHM001F and IHH015. The majority of interaction regions are < 1 Mb in span, indicating ER-α mainly employs local chromatin interactions as a mechanism.



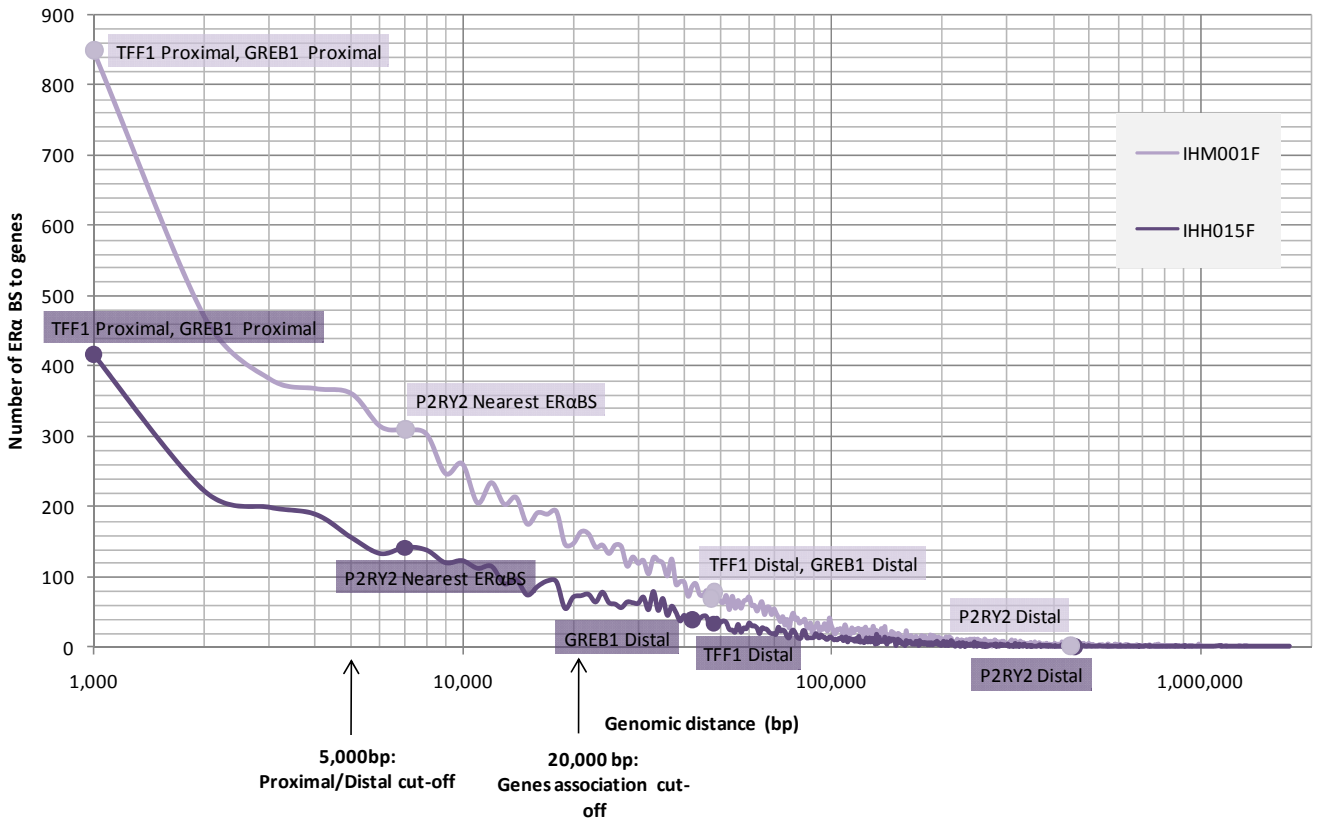
Supplementary Figure 13. ER-αBS involvement in chromatin interactions

a. Reproducibility analyses showing the overlaps between different classes of ER-αBS with ChIP-chip (green), ChIP-Seq (red) and replicate ChIA-PET libraries (blue) for IHM001F and IHH015F.



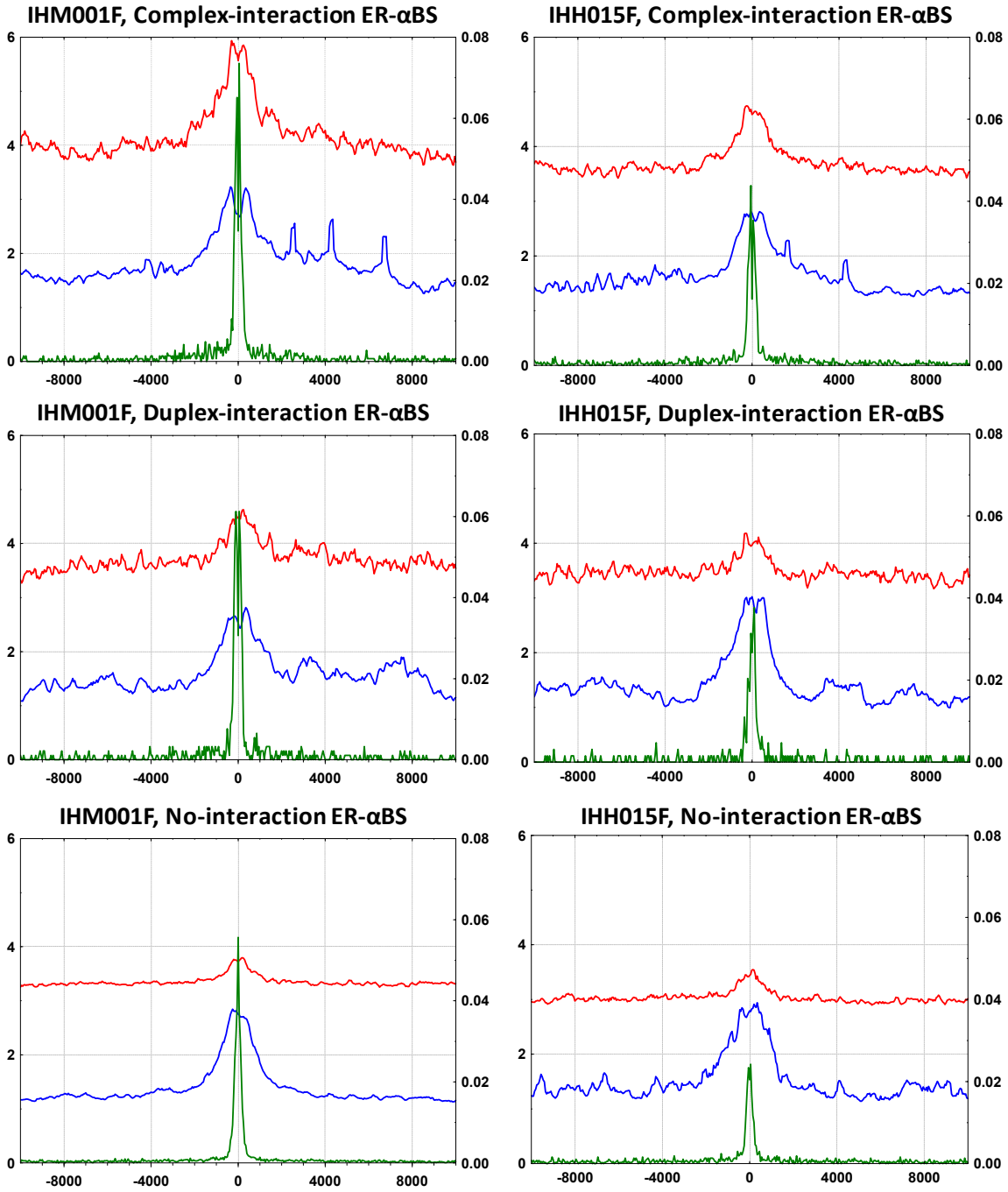
Supplementary Figure 13. ER-αBS involvement in chromatin interactions

b. Numbers of ER-αBS identified with different ChIP enrichment cutoffs and percentages of ER-αBS distribution indifferent categories of involvement with complex and duplex chromatin interaction regions, and no-interactions. The majority of high-confidence high-enrichment ER-αBSs are involved in chromatin interactions (complex and duplex chromatin interaction regions) as opposed to no-interactions.



Supplementary Figure 14. Distances between ER-αBS and target genes TSS

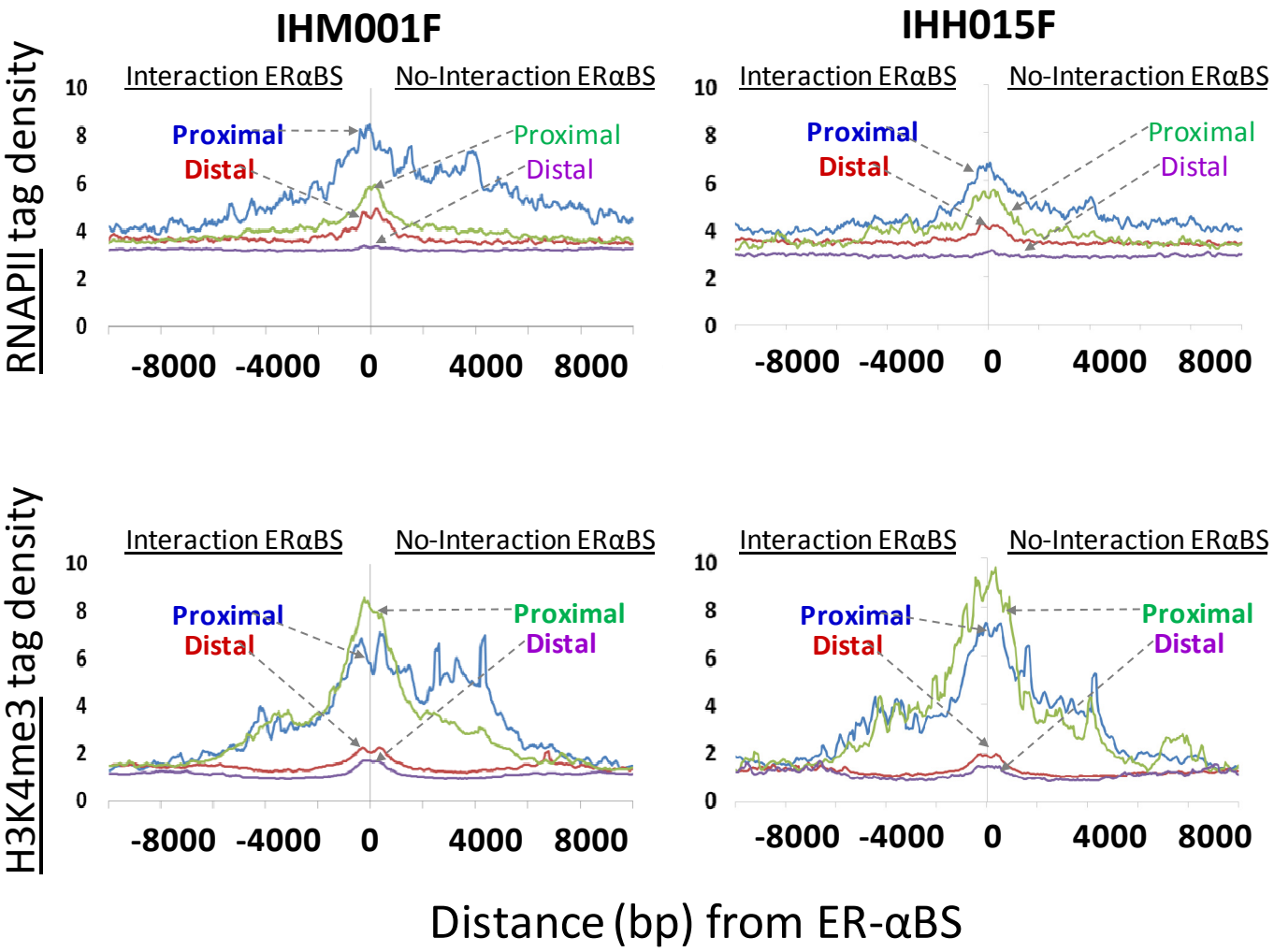
The genomic distance between ER-αBS and UCSC Known Gene TSS is shown in light purple for IHM001F and dark purple for IHH015F. Locations of a few well-characterized ER-α target gene binding sites are displayed on the chart. Many ER-αBS are very close to genes, but there is also a long tail with many ER-αBS far away from gene TSS. We used 5 Kb as a cut-off to differentiate promoter-proximal and distal ER-αBS, and 20 Kb as a cut-off in performing gene associations.



Supplementary Figure 15. H3K4me3, RNAPII and FoxA1 analyses at ER-αBS

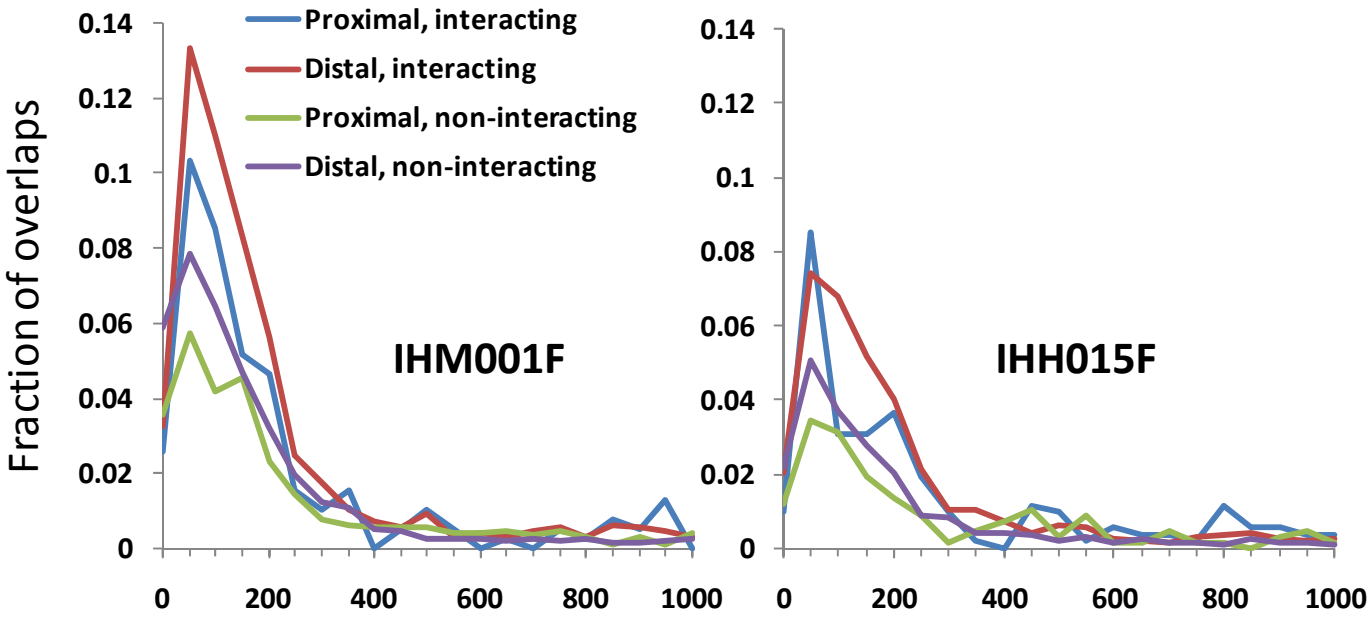
a. H3K4me3 (blue), RNAPII (red), and FoxA1 (green) profiles at ER-αBS associated with complex-interactions, duplex-interactions, and no-interactions in IHM001F and IHH015F datasets. The left Y-axis shows the average number of ChIP-Seq tags for H3K4me3 and RNAPII profiles. The right Y-axis shows the

frequency of FoxA1 binding sites (ChIP-chip[11]) in binned 50 bp intervals around ER-αBS. The X-axis shows the genomic distance (in base pairs) from the middle of the ER-αBS (denoted as position 0).



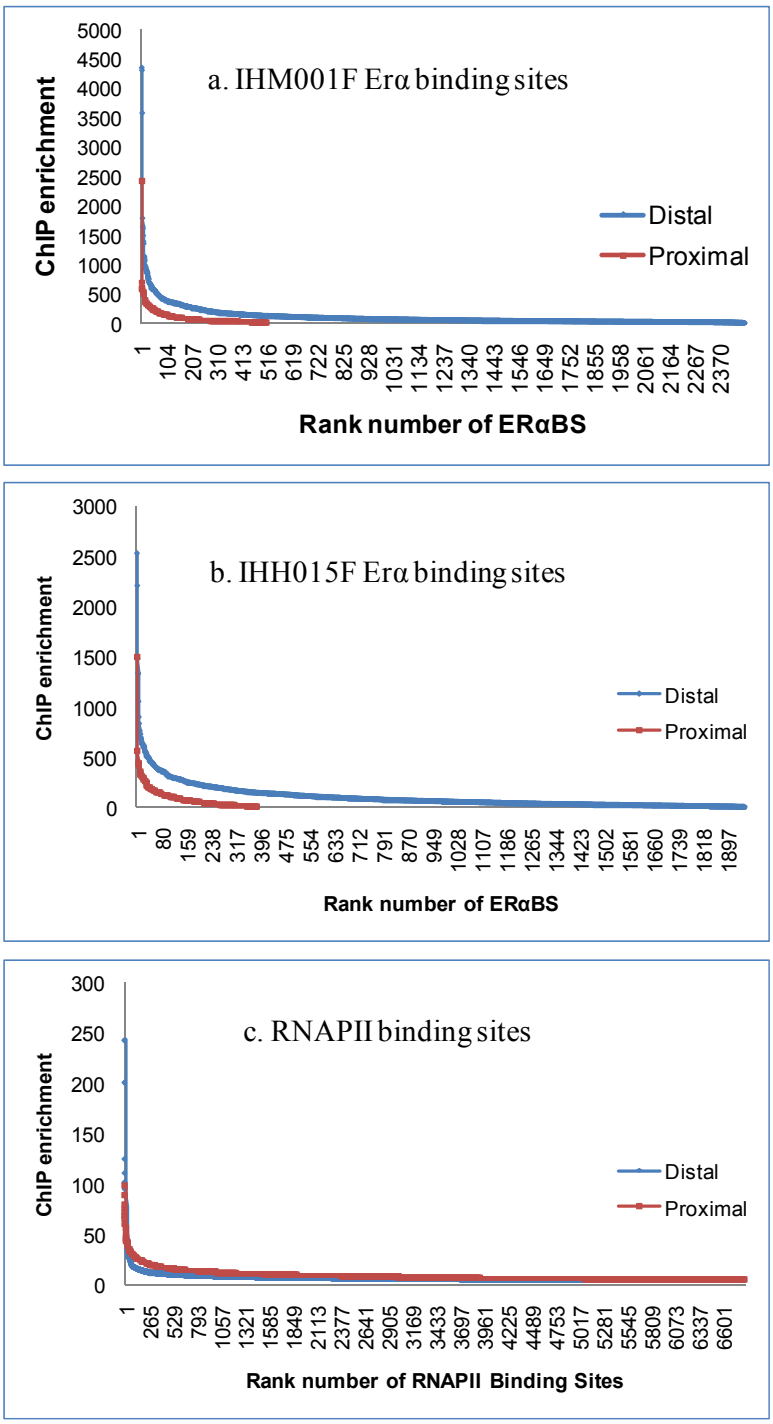
Supplementary Figure 15. H3K4me3, RNAPII and FoxA1 analyses at ER-αBS

b. H3K4me3, and RNAPII profiles at proximal and distal, interacting and non-interacting ER-αBS in IHM001F and IHH015F datasets.

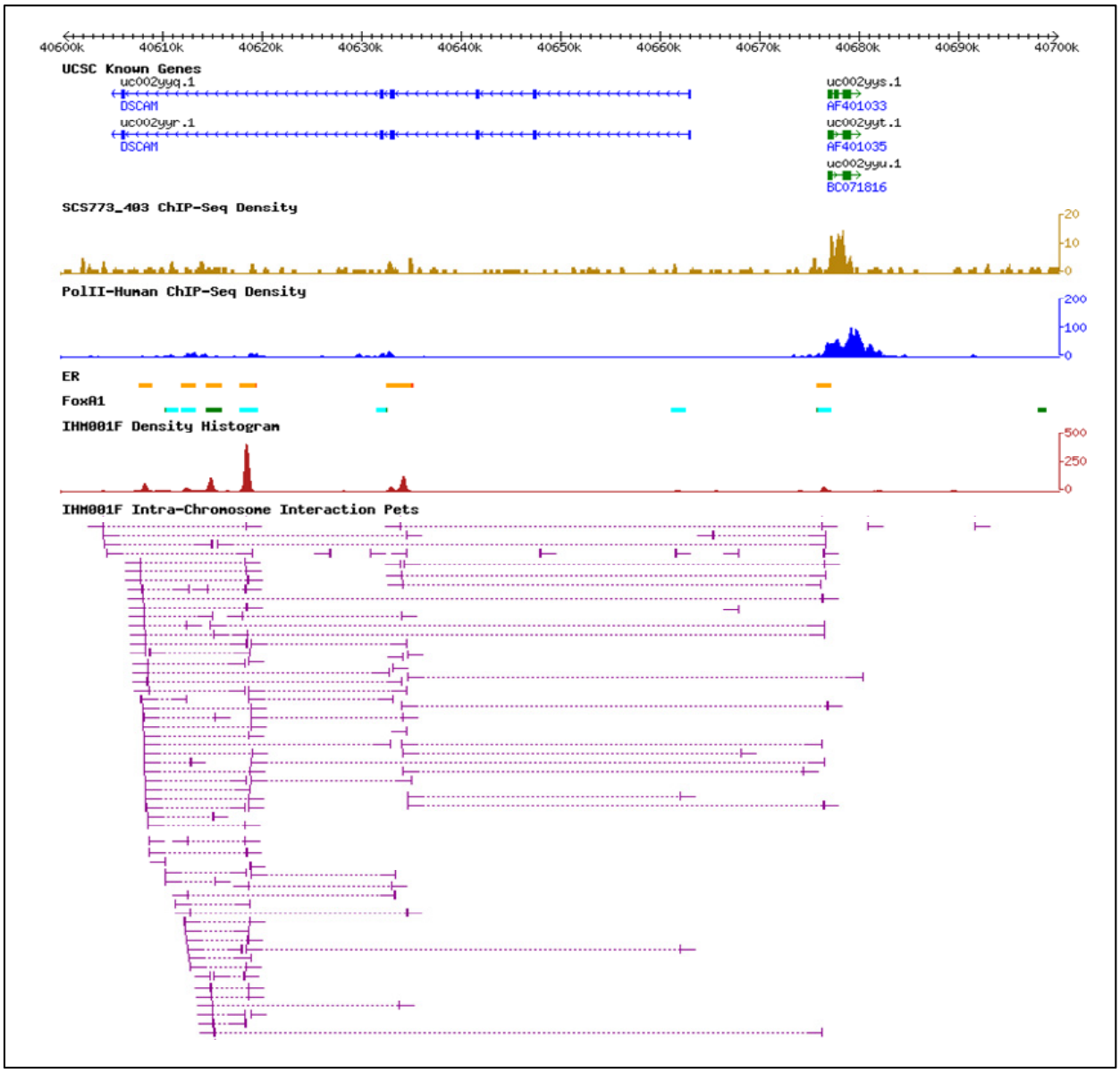


Supplementary Figure 15. H3K4me3, RNAPII and FoxA1analyses at ER-αBS

c. FoxA1 binding sites (ChIP-chip[11]) distribution at proximal and distal, interacting and non-interacting ER-αBS in IHM001F and IHH015F datasets by distance (bp).



Supplementary Figure 16. ChIP enrichment at distal and proximal ER-α and RNAPII binding sites
ChIP enrichment at distal (blue) and proximal (red) binding sites to UCSC known gene transcription start sites (TSS). The ChIP enrichment is higher at distal ER-αBS than proximal ER-αBS, and by contrast, is higher at proximal RNAPII binding sites than distal RNAPII binding sites.

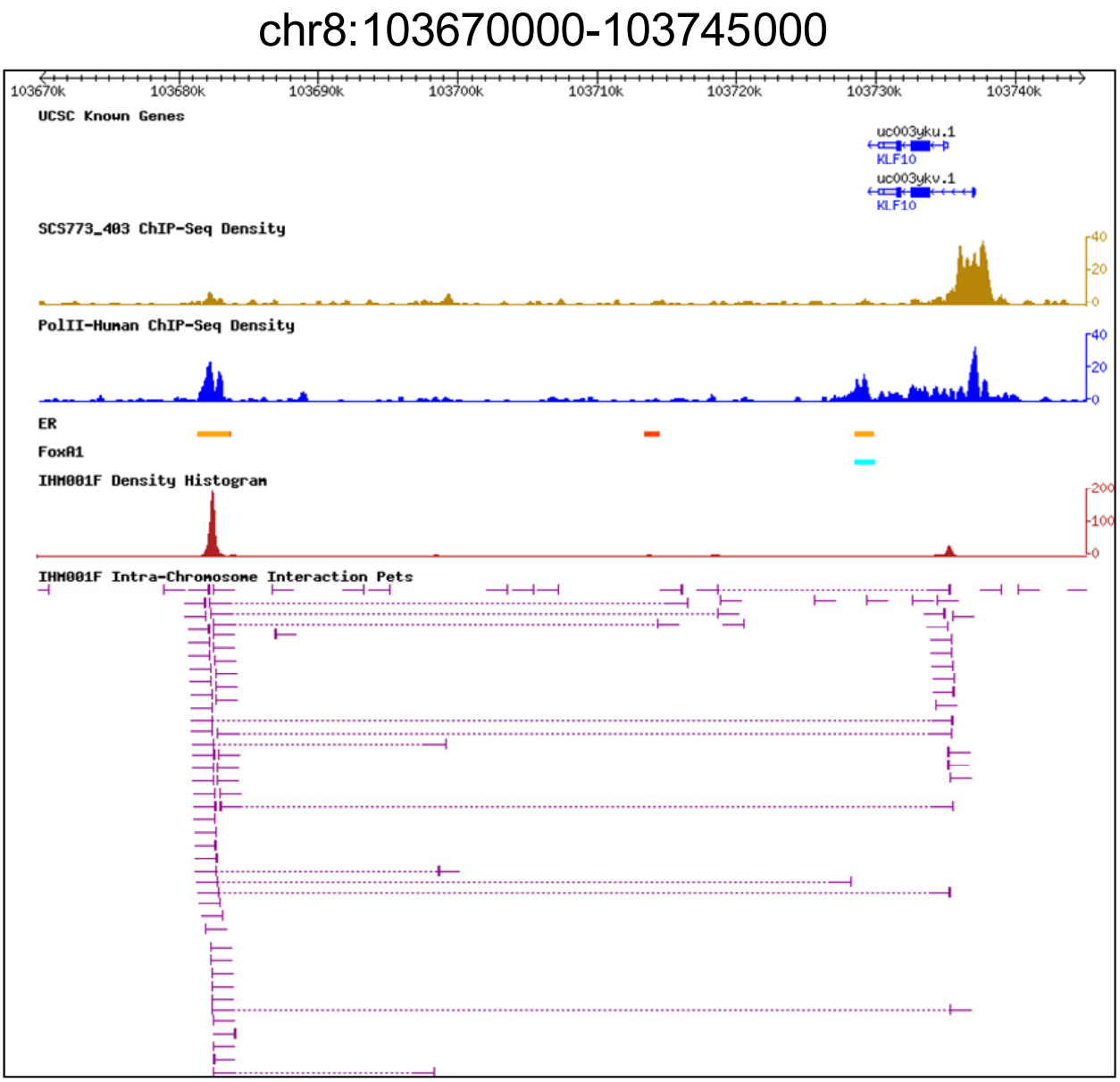


Supplementary Figure 17. Examples of genes and chromatin interactions with stronger anchors at distal sites and weaker anchors at promoter sites

a. An example of complex chromatin interaction with stronger ER-α binding and anchoring at the distal `5' site than at the promoter of *BC071816*.

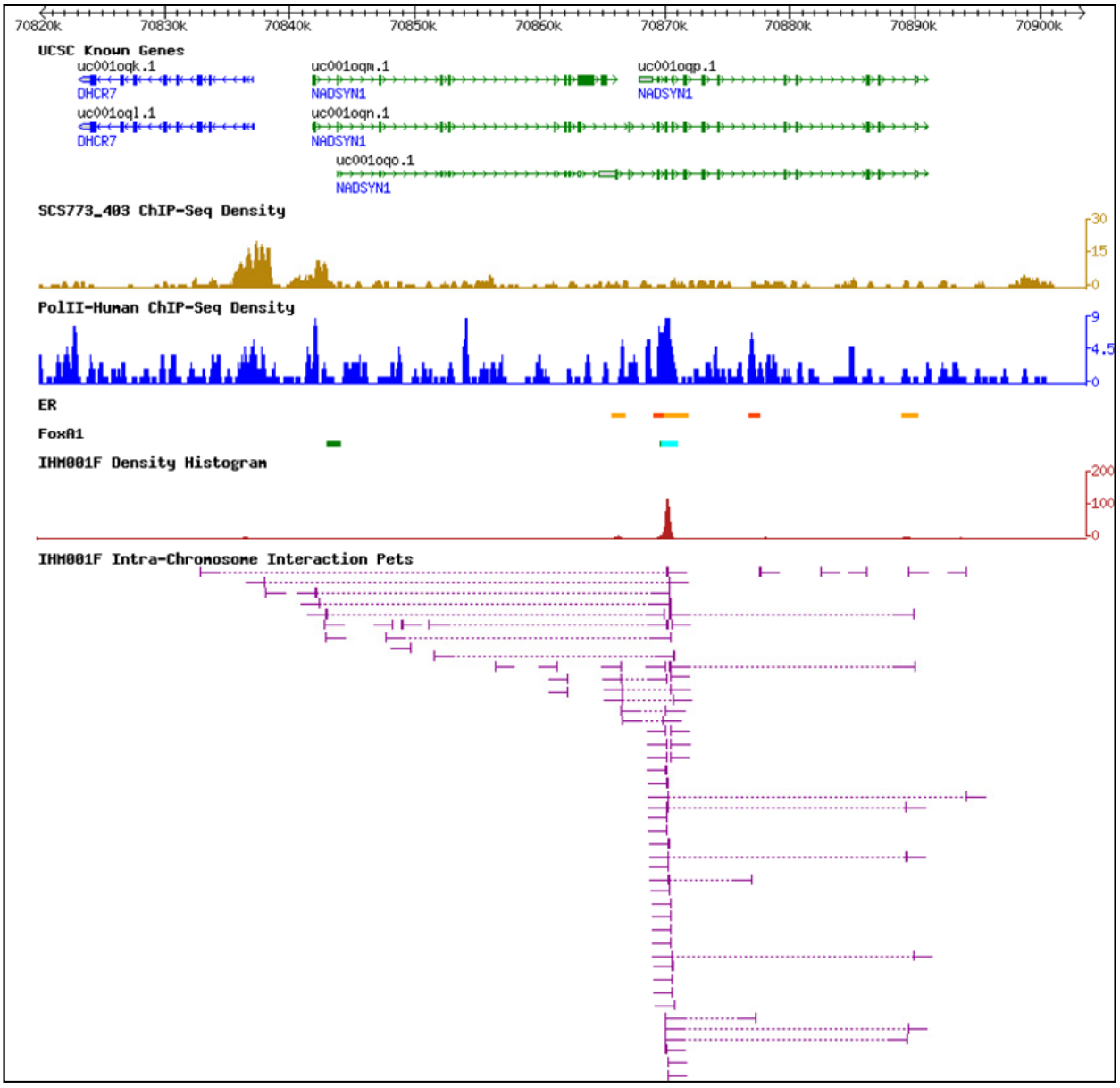
Supplementary Figure 17. Examples of genes and chromatin interactions with stronger anchors at distal sites and weaker anchors at promoter sites

b. An example of complex chromatin interaction with stronger ER-α binding and anchoring at distal 5' sites than at the promoter of *STC2*.



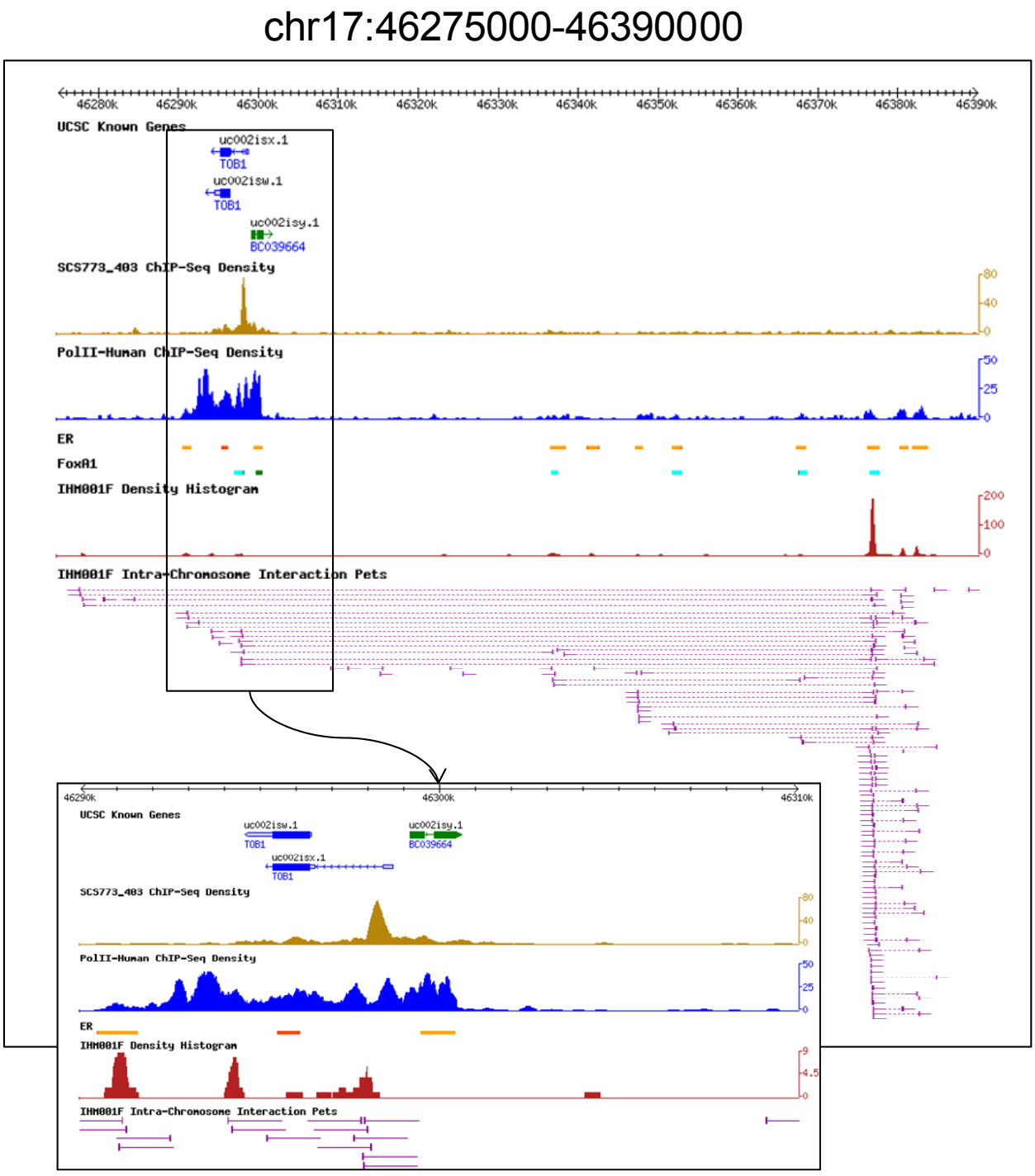
Supplementary Figure 17. Examples of genes and chromatin interactions with stronger anchors at distal sites and weaker anchors at promoter sites

c. An example of complex chromatin interaction with stronger ER-α binding and anchoring at a distal 3' site than at the promoter of *KLF10*.



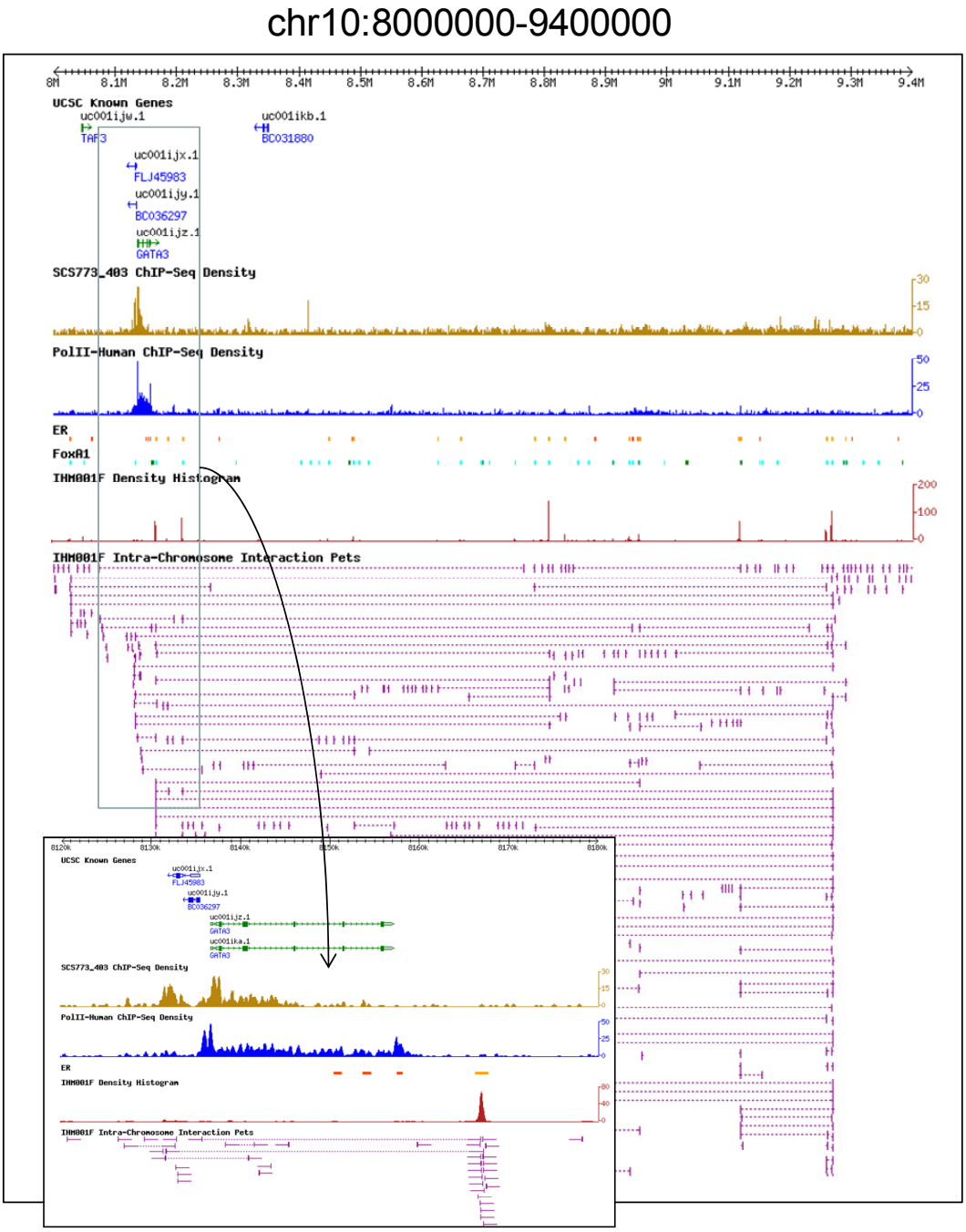
Supplementary Figure 17. Examples of genes and chromatin interactions with stronger anchors at distal sites and weaker anchors at promoter sites

d. An example of complex chromatin interaction with stronger ER-α binding and anchoring at distal sites than at a bi-directional promoter at *DHCR7* and *NADSYN1*.



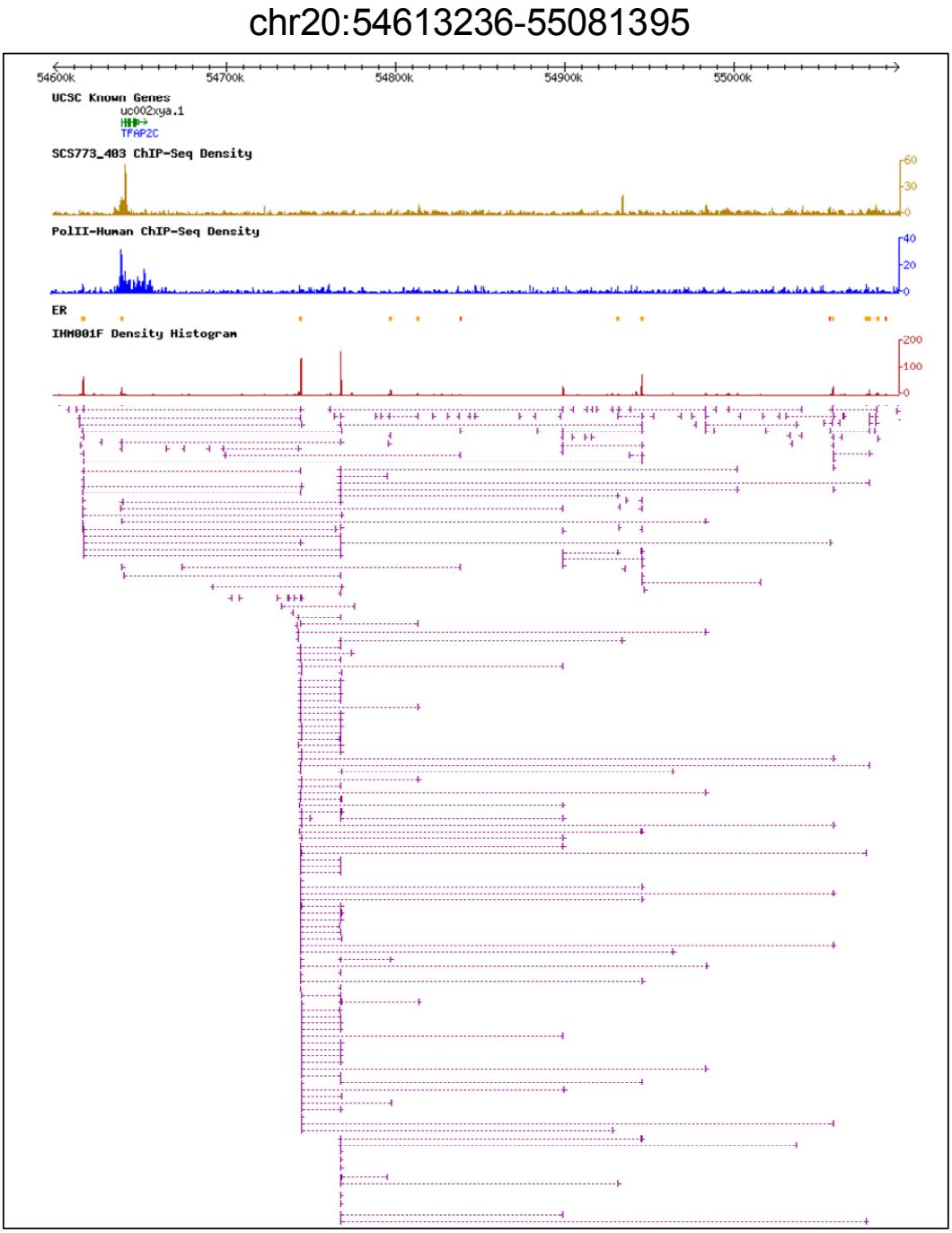
Supplementary Figure 17. Examples of genes and chromatin interactions with stronger anchors at distal sites and weaker anchors at promoter sites

e. An example of complex chromatin interaction with stronger ER-α binding and anchoring at distal sites than at a bi-directional promoter at *TOB1* and *BC039664*.



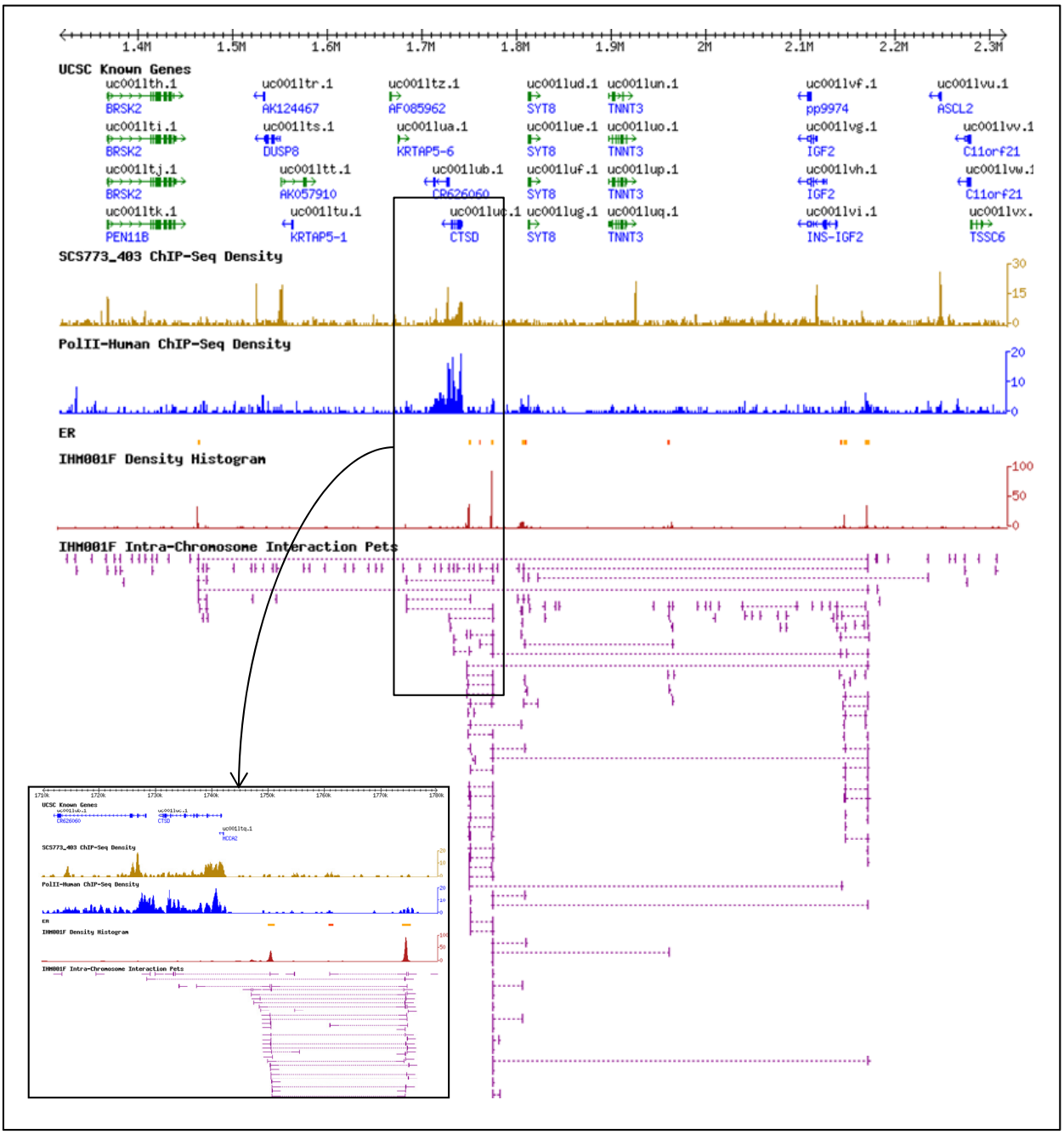
Supplementary Figure 17. Examples of genes and chromatin interactions with stronger anchors at distal sites and weaker anchors at promoter sites

f. An example of complex chromatin interaction with stronger ER-α binding and anchoring at distal 3' sites than at a promoter for *GATA3*.



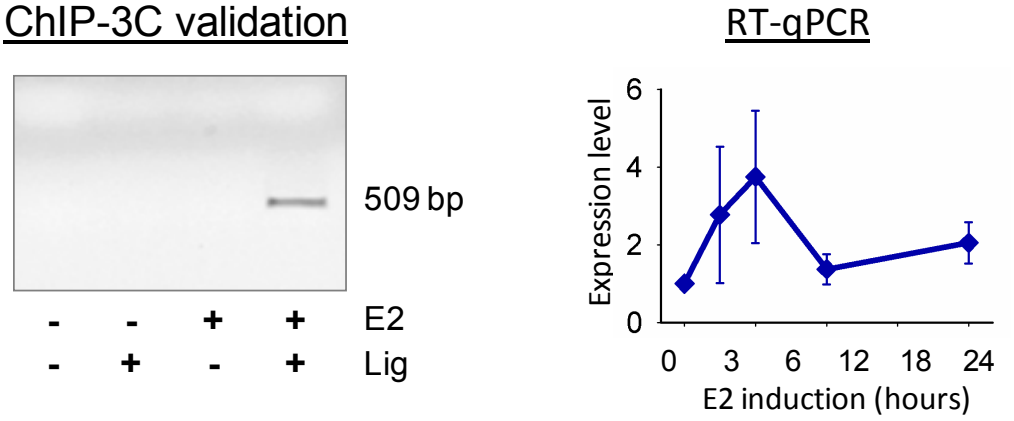
Supplementary Figure 17. Examples of gene and chromatin interactions with stronger anchors at distal sites and weaker anchors at promoter sites

g. An example of complex chromatin interaction with stronger ER-α binding and anchoring at distal 3' sites than at the promoter for *TFAP2C*.



Supplementary Figure 17. Examples of gene and chromatin interactions with stronger anchors at distal sites and weaker anchors at promoter sites

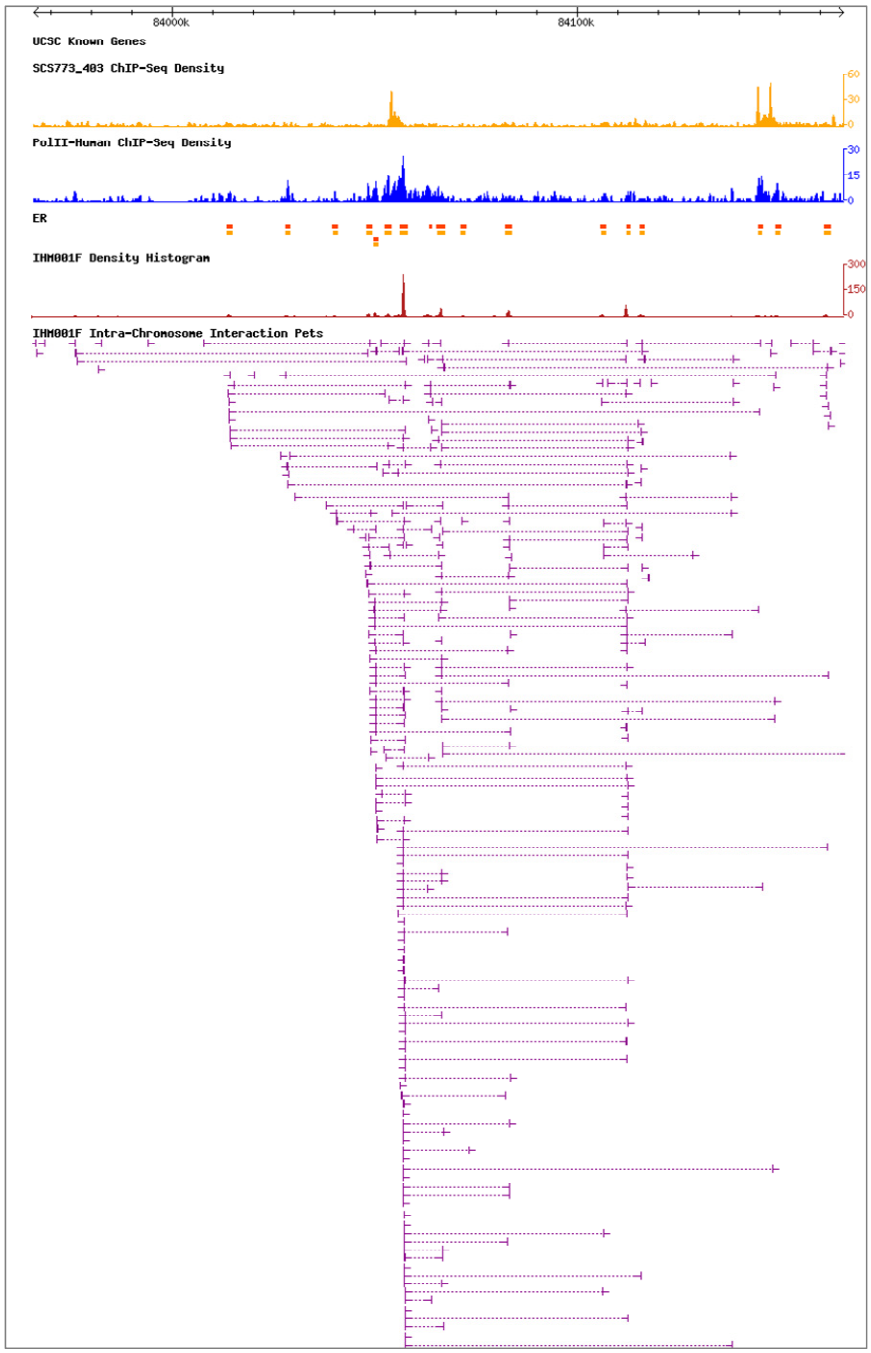
h. An example of complex chromatin interaction with stronger ER-α binding and anchoring at the distal 3' site than at the bi-promoter of *CTSD*.



Supplementary Figure 18. Examples of ER-α-bound chromatin interaction regions with no associated anchor genes

a. An example of ChIP-3C validated chromatin interaction with no well annotated genes nearby but some computational predications. The H3K4me3 and RNAPII marks and RT-qPCR data suggest possible transcription activity at the anchor region of this interaction.

chr16:83965723-84165722

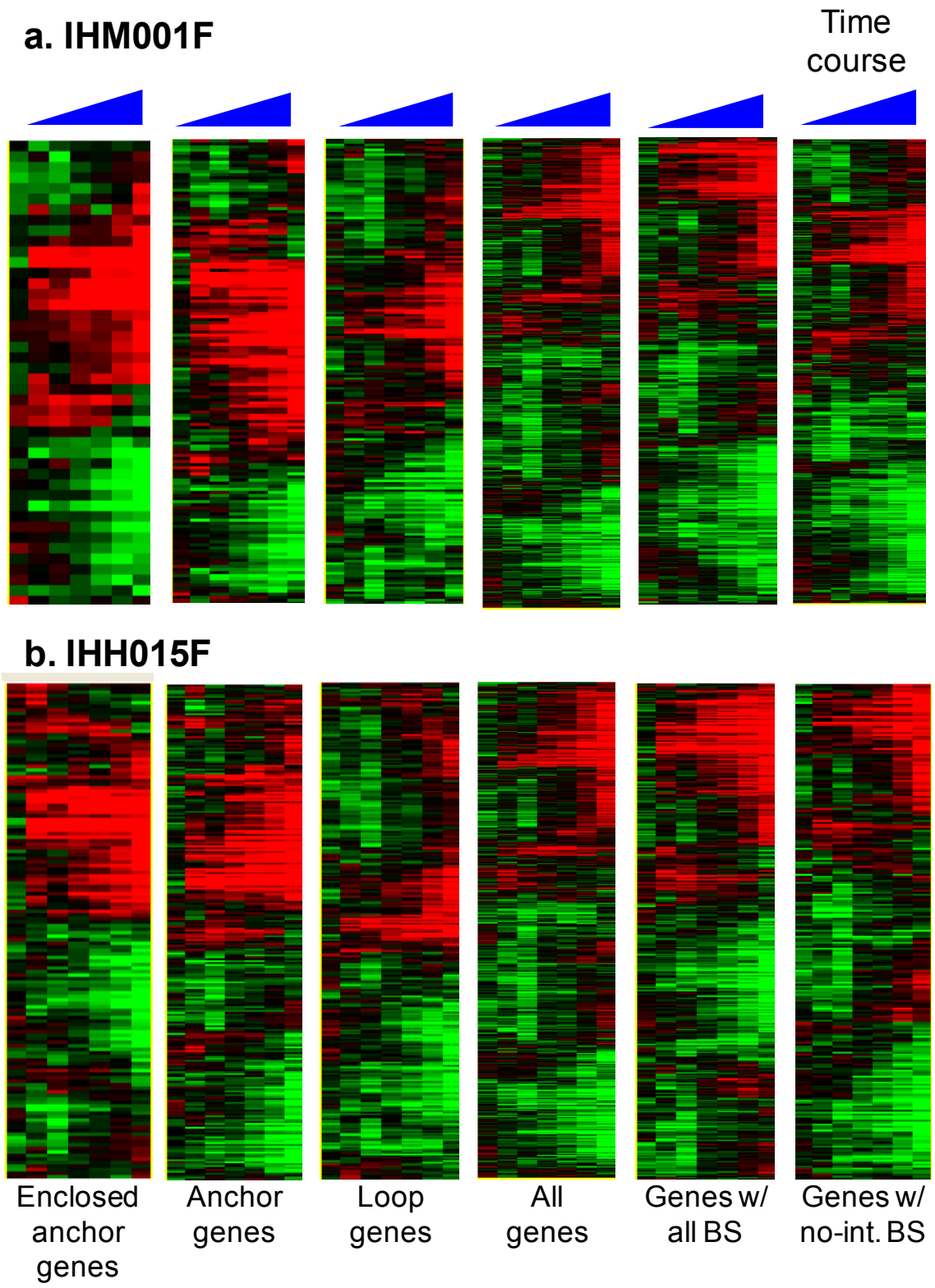


Supplementary Figure 18. Examples of ER-α-bound chromatin interaction regions with no associated anchor genes

b. An example of chromatin interaction with no well annotated genes nearby, but which involves binding sites marked by H3K4me3 and RNAPII marks. Possibly, there are unknown genes in this region.

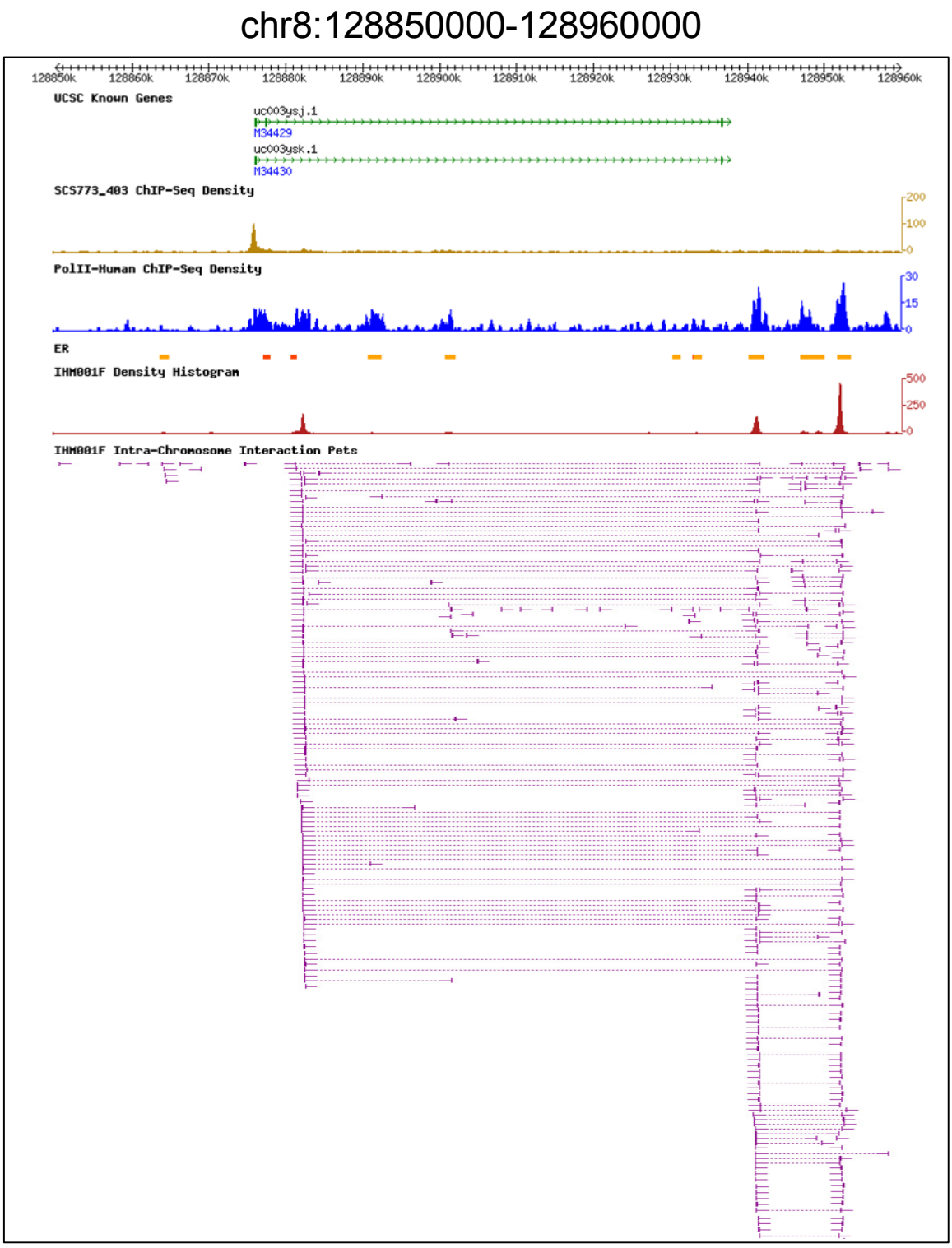
Supplementary Figure 18. Examples of ER-α-bound chromatin interaction regions with no associated anchor genes

c. An example of chromatin interaction with no well annotated genes nearby, but which involves binding sites marked by H3K4me3 and RNAPII marks. There is a weak singleton inter-ligation PET reaching out from the interaction anchor to the promoter region of a gene, suggesting that the interaction region actually includes the active gene *LAPTM4A*.



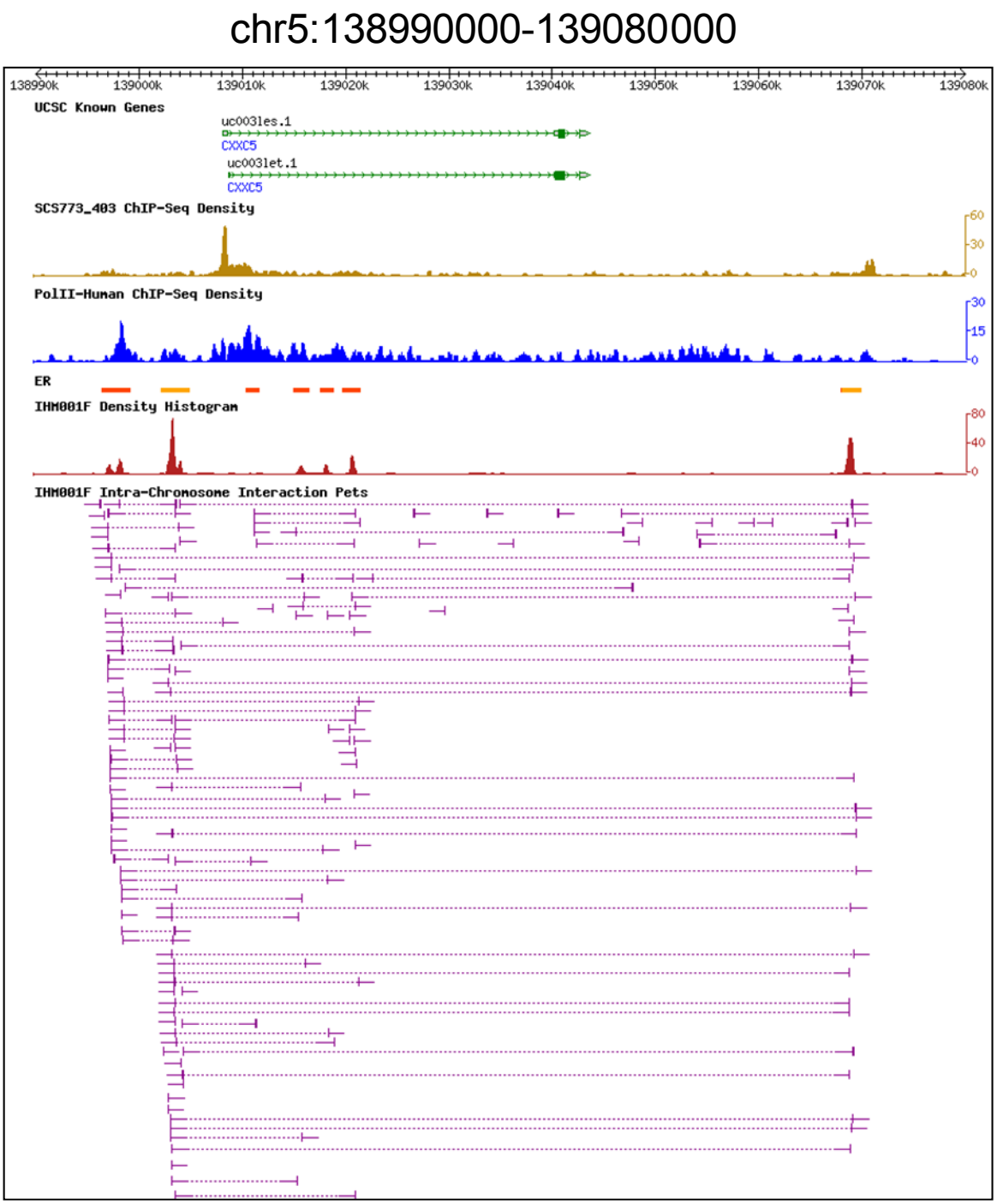
Supplementary Figure 19. Microarray treeviews for different classes of genes

a. IHM001F data; **b.** IHH015F data.



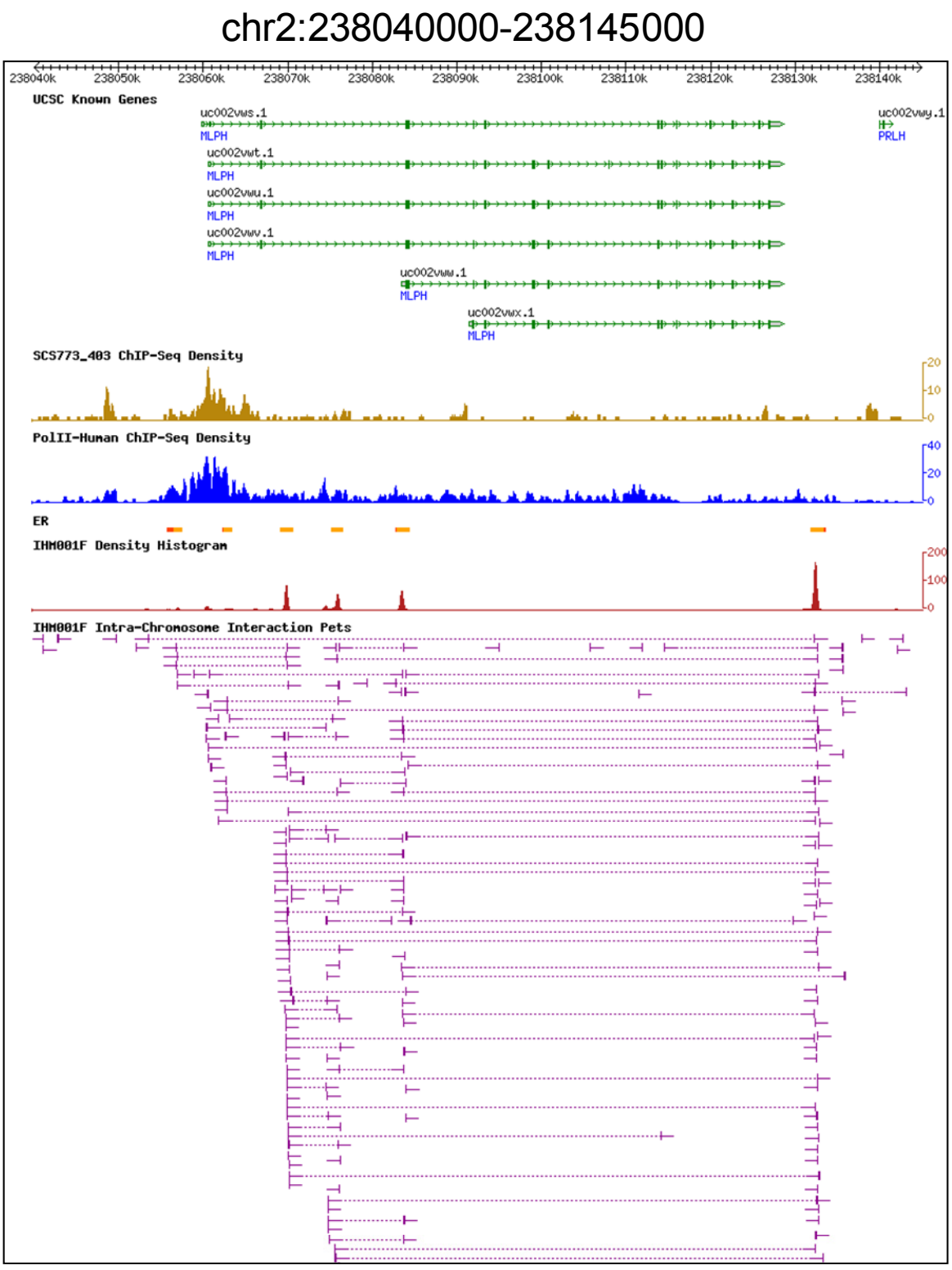
Supplementary Figure 20. Examples of enclosed anchor genes

a. Enclosed anchor gene *M34429* in chr8:128850000-128960000



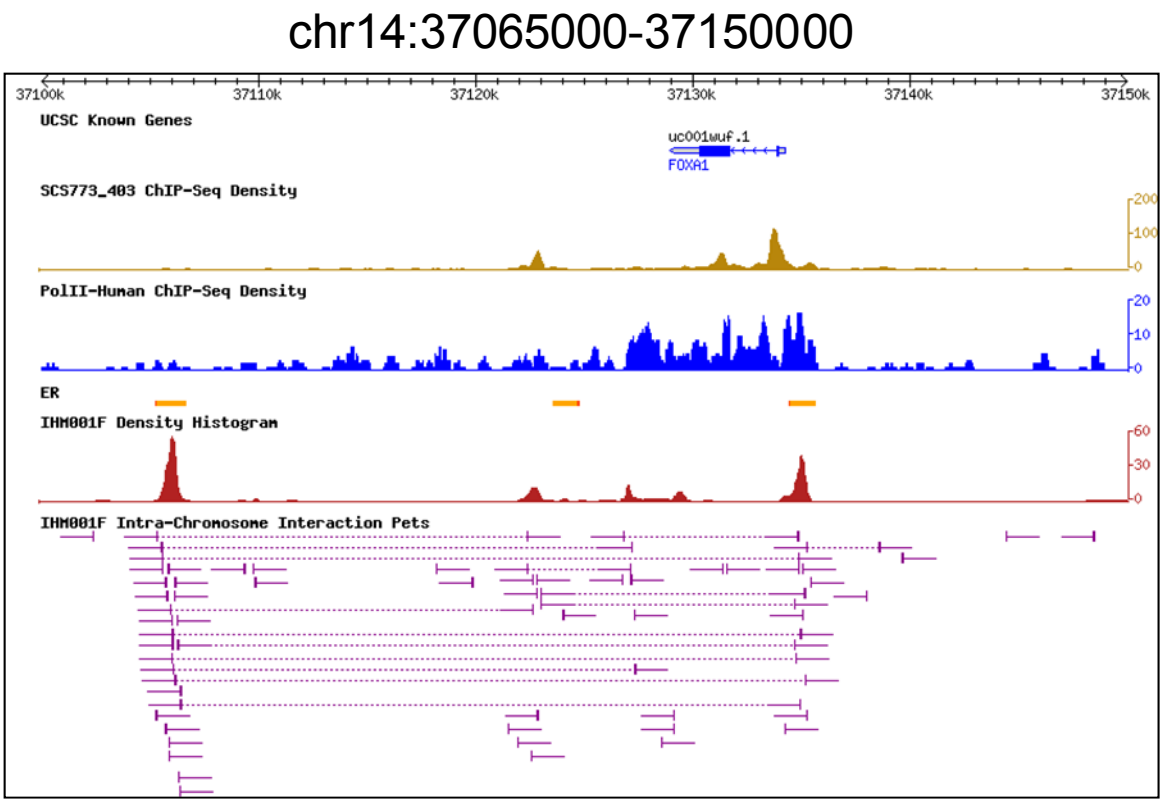
Supplementary Figure 20. Examples of enclosed anchor genes

b. Enclosed anchor gene *CXXC5* in chr5:138990000-139080000



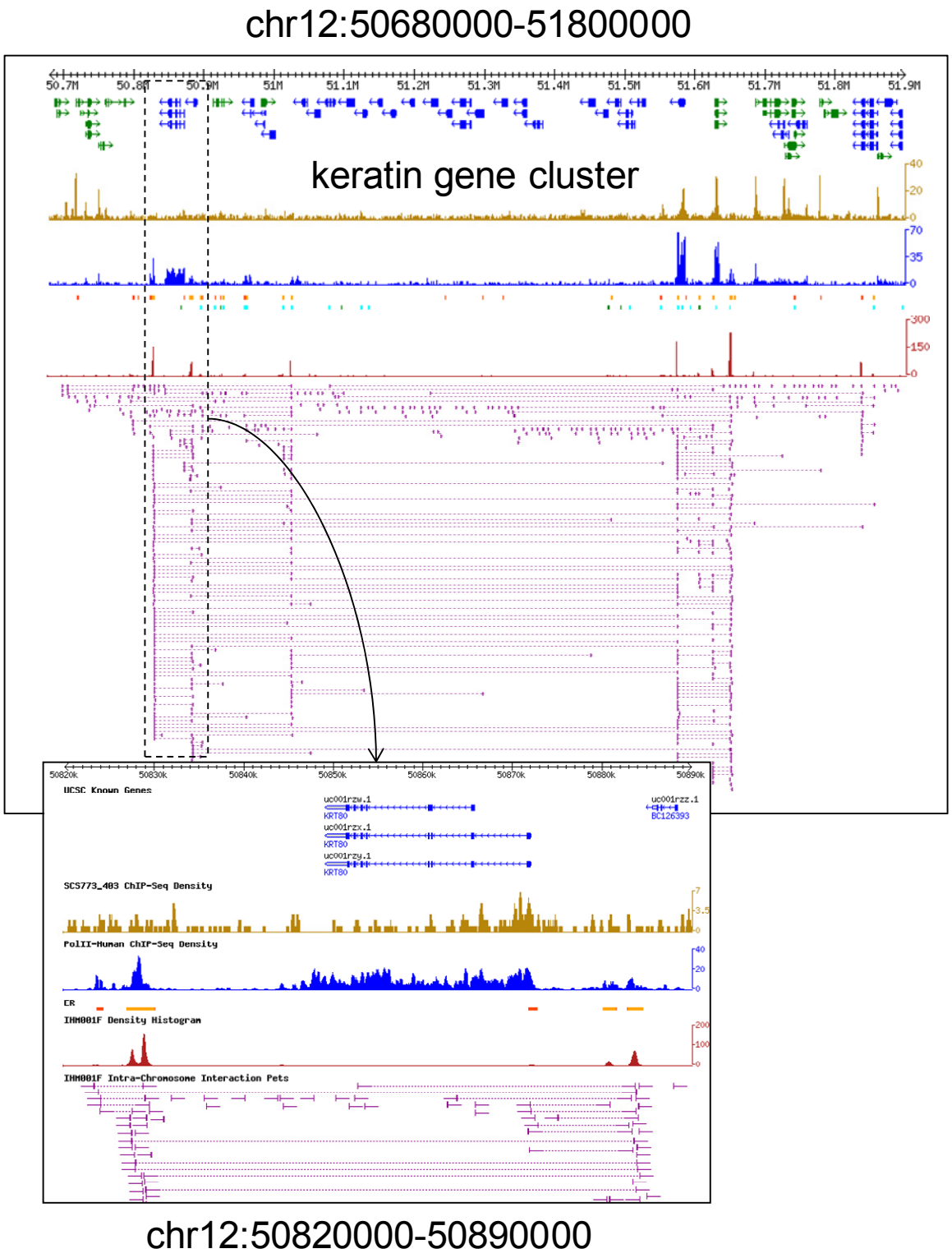
Supplementary Figure 20. Examples of enclosed anchor genes

c. Enclosed anchor gene *MLPH* in chr2:238040000-238145000



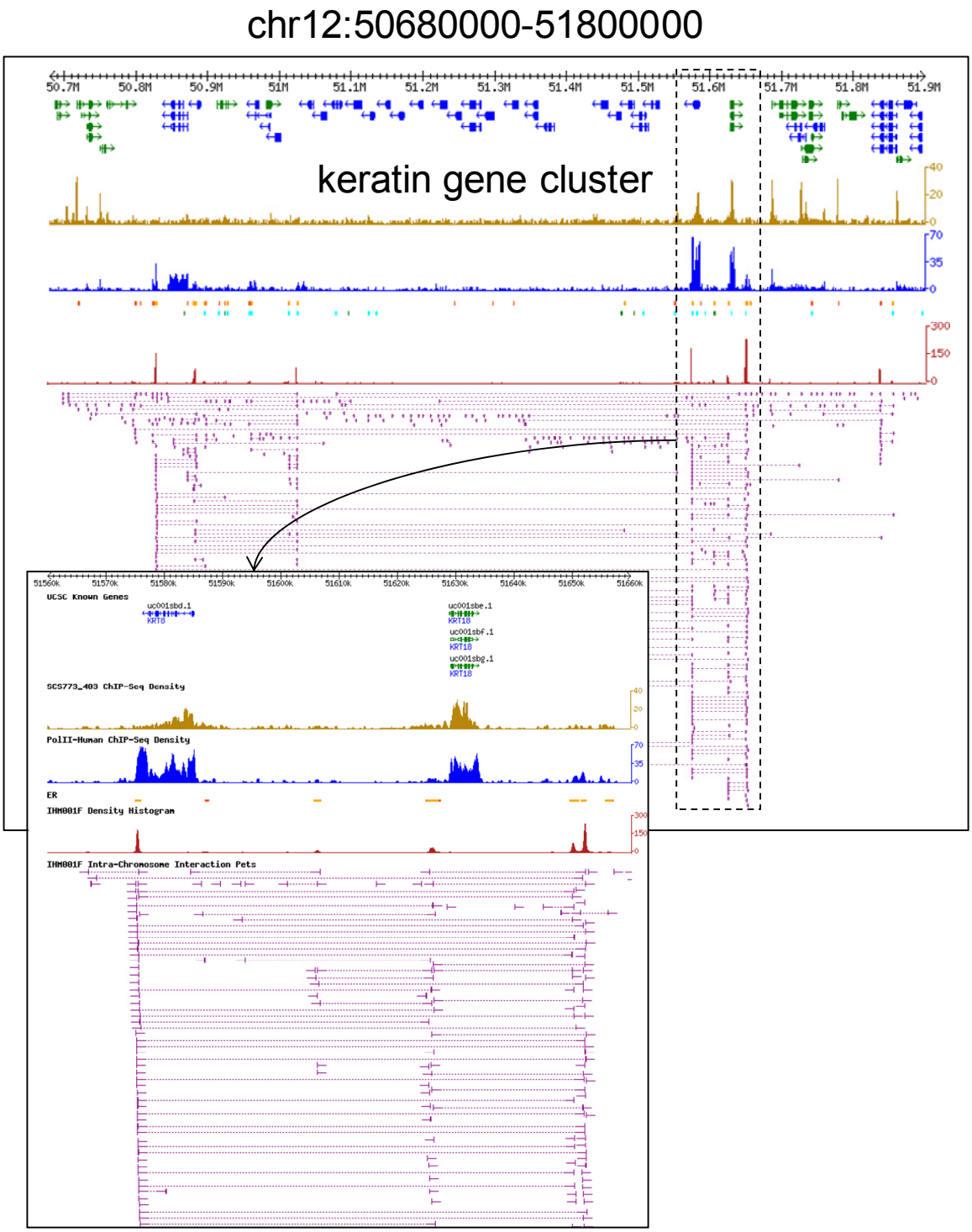
Supplementary Figure 20. Examples of enclosed anchor genes

d. Enclosed anchor gene *FOXA1* in chr14:37065000-37150000



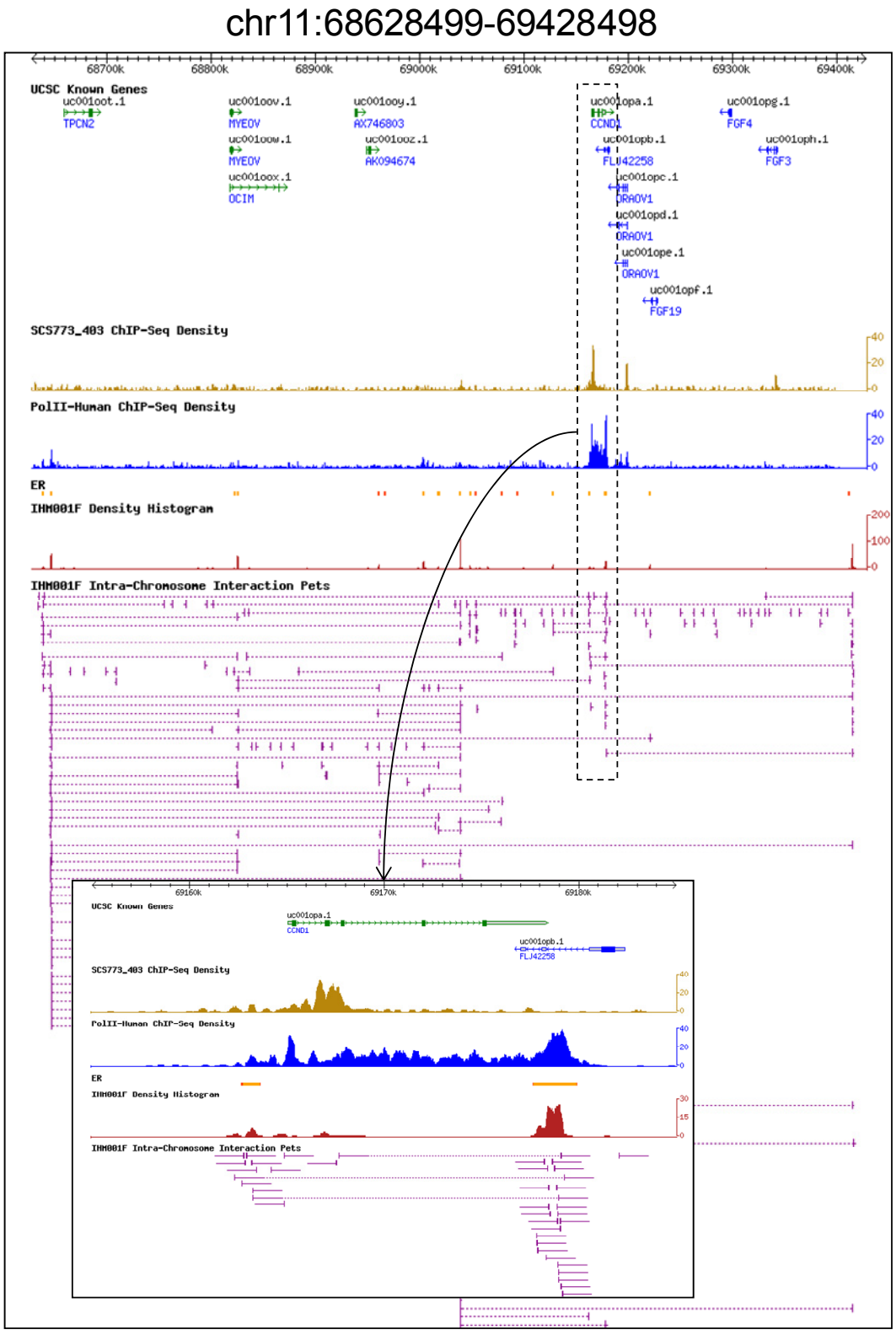
Supplementary Figure 20. Examples of enclosed anchor genes

e. Enclosed anchor gene *KRT80* in chr12:50820000-50890000



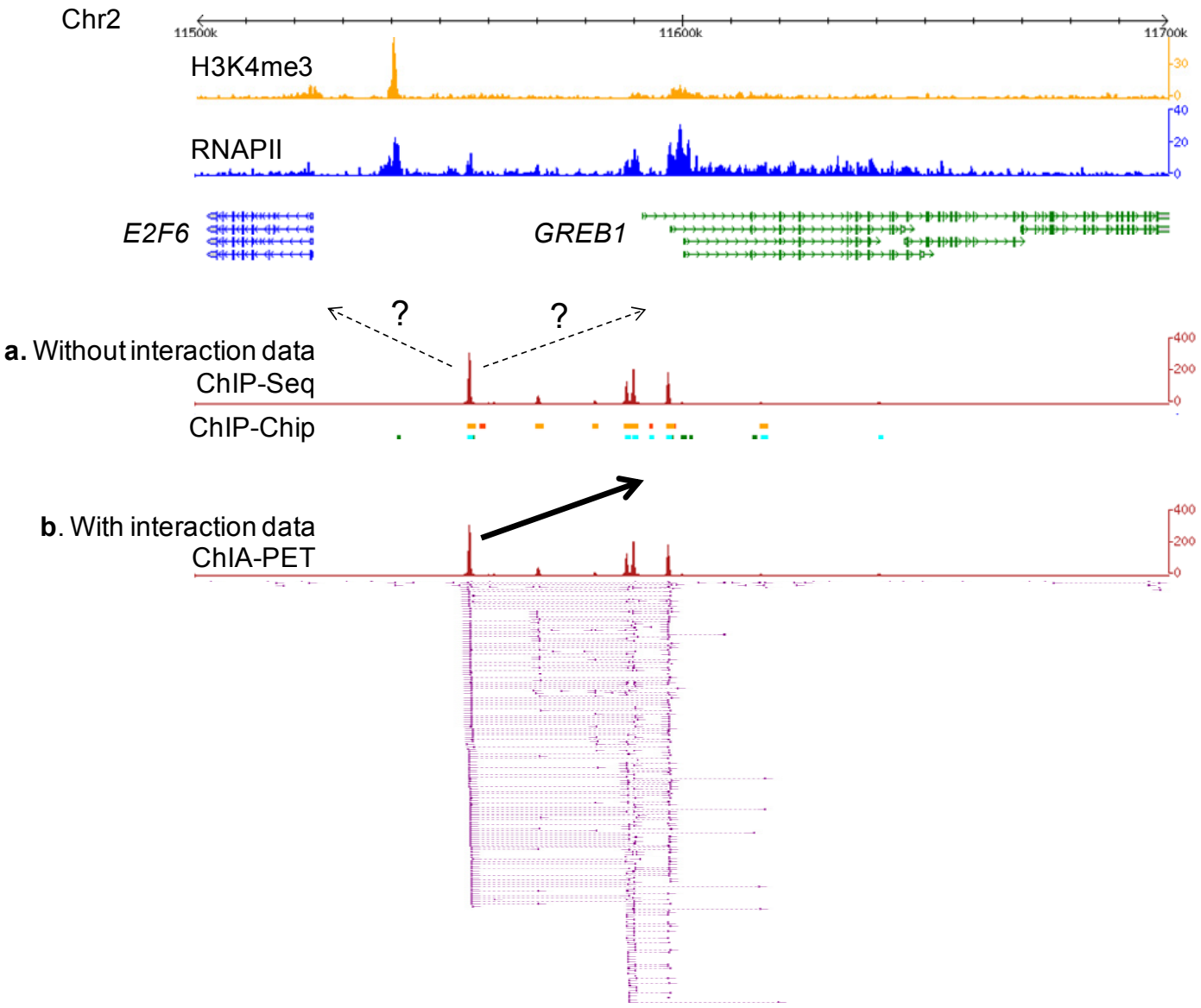
Supplementary Figure 20. Examples of enclosed anchor genes

f. Enclosed anchor gene *KRT18*, *KRT8* in chr12:51560000-51660000



Supplementary Figure 20. Examples of enclosed anchor genes

g. Enclosed anchor gene *CCND1* in chr11:68628499-69428498



Supplementary Figure 21. ChIA-PET for less ambiguous gene assignments to binding sites

Screenshots of ER-αBS and genes, **a.** without chromatin interaction information and **b.** with chromatin interaction information. Often, it is difficult to know which gene a distal ER-αBS should be assigned to in the absence of chromatin interaction information. In previous studies that used 50 Kb as a cutoff (for example Carroll et al., 2006[22]; Welboren et al., 2009[2]) both *E2F6* and *GREB1* would have been assigned. If 20 Kb were to be used, neither *E2F6* nor *GREB1* would have been assigned to the distal ER-αBS in question. However, with the ChIA-PET data, because we use 20 Kb to assign genes to the ER-αBS, the proximal ER-αBS would have been assigned *GREB1*, and because we consider distal ER-αBS to have been brought into close spatial proximity to the proximal ER-αBS if they have chromatin interactions, so *GREB1* would have been assigned to the distal ER-αBS.

Supplementary references

1. Lin, C.Y. et al. Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet* **3**, e87 (2007).
2. Welboren, W.J. et al. ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J* **28**, 1418-28 (2009).
3. Wei, C.L. et al. A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207-19 (2006).
4. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-17 (2008).
5. Volik, S. et al. Decoding the fine-scale structure of a breast cancer genome and transcriptome. *Genome Res* **16**, 394-404 (2006).
6. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-8 (1998).
7. Chiu, K.P. et al. PET-Tool: a software suite for comprehensive processing and managing of Paired-End diTag (PET) sequence data. *BMC Bioinformatics* **7**, 390 (2006).
8. Dunn, J.J. et al. Genomic signature tags (GSTs): a system for profiling genomic DNA. *Genome Res* **12**, 1756-65 (2002).
9. Smit, A.F.A., Hubley, R. & Green, P. RepeatMasker Open-3.0. (1996-2004).
10. Shadeo, A. & Lam, W.L. Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res* **8**, R9 (2006).
11. Lupien, M. et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958-70 (2008).
12. Stein, L.D. et al. The generic genome browser: a building block for a model organism system database. *Genome Res* **12**, 1599-610 (2002).
13. Hsu, F. et al. The UCSC Known Genes. *Bioinformatics* **22**, 1036-46 (2006).
14. Zhao, X.D. et al. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286-98 (2007).
15. Hagege, H. et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2**, 1722-33 (2007).
16. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. in *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds. Krawetz, S. & Misener, S.) 365-386 (Humana Press, Totowa, NJ, 2000).
17. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-65 (2002).
18. Perkins, K.J., Lusic, M., Mitar, I., Giacca, M. & Proudfoot, N.J. Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. *Mol Cell* **29**, 56-68 (2008).
19. Zhao, Z. et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* **38**, 1341-7 (2006).
20. Karolchik, D. et al. The UCSC Genome Browser Database. *Nucleic Acids Res* **31**, 51-4 (2003).
21. Kwon, Y.S. et al. Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. *Proc Natl Acad Sci U S A* **104**, 4852-7 (2007).
22. Carroll, J.S. et al. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**, 1289-97 (2006).
23. Fuchs, E. & Weber, K. Intermediate filaments: structure, dynamics, function, and disease. *Annu Rev Biochem* **63**, 345-82 (1994).

24. Rogers, M.A. et al. Characterization of new members of the human type II keratin gene family and a general evaluation of the keratin gene domain on chromosome 12q13.13. *J Invest Dermatol* **124**, 536-44 (2005).
25. Steinert, P.M. & Roop, D.R. Molecular and cellular biology of intermediate filaments. *Annu Rev Biochem* **57**, 593-625 (1988).
26. Moll, R., Divo, M. & Langbein, L. The human keratins: biology and pathology. *Histochem Cell Biol* **129**, 705-33 (2008).
27. Lu, X. & Lane, E.B. Retrovirus-mediated transgenic keratin expression in cultured fibroblasts: specific domain functions in keratin stabilization and filament formation. *Cell* **62**, 681-96 (1990).
28. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-37 (2007).
29. Johnson, D.S., Mortazavi, A., Myers, R.M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-502 (2007).
30. Fullwood, M.J. & Ruan, Y. ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem* (2009).
31. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-80 (2005).
32. Fullwood, M.J., Wei, C.L., Liu, E.T. & Ruan, Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res* **19**, 521-32 (2009).
33. Shendure, J. et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728-32 (2005).
34. Harris, T.D. et al. Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106-9 (2008).
35. Bourque, G. et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**, 1752-62 (2008).
36. 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).
37. Metivier, R., Reid, G. & Gannon, F. Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep* **7**, 161-7 (2006).
38. Pan, Y.F. et al. Regulation of estrogen receptor-mediated long-range transcription via evolutionarily conserved distal response elements. *J Biol Chem* (2008).
39. Carroll, J.S. et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33-43 (2005).
40. Deschenes, J., Bourdeau, V., White, J.H. & Mader, S. Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences. *J Biol Chem* **282**, 17335-9 (2007).
41. Barnett, D.H. et al. Estrogen receptor regulation of carbonic anhydrase XII through a distal enhancer in breast cancer. *Cancer Res* **68**, 3505-15 (2008).
42. Bretschneider, N. et al. E2-mediated cathepsin D (CTSD) activation involves looping of distal enhancer elements. *Mol Oncol* **2**, 182-90 (2008).

Glossary

1. **Tag**: A short fragment of DNA sequence, here about 20-21 bp, derived from a ChIP DNA fragment. The tag sequence is mapped to the human reference genome, thus giving the genomic location of the tag.
2. **Tag-defined ChIP fragment**: To represent the original ChIP fragments that the tags were derived from, a tag is extended in the 3' direction by an extra 1500 bp.
3. **PET**: Abbreviation of "paired-end-tag". Sequenced construct consisting of two covalently linked tags, such that the relationship between the two tags is known. The tags come from two DNA ends.
4. **Self-ligation PET**: A PET arising from a ligation between the two ends of the same ChIP DNA fragment.
5. **Inter-ligation PET (or iPET):** A PET arising from a ligation between the ends of two different ChIP DNA fragments as represented by two **Tag-defined ChIP fragments**.
6. **Inter-ligation PET cluster:** A group of inter-ligation PETs whose genomic locations of **Tag-defined ChIP fragments** are directly overlapping. They are considered as derived from the same interacting pair of genomic regions.
7. **Singleton inter-ligation PET**: Inter-ligation PETs that do not overlap with other inter-ligation PETs. These are generally noise or very weak interactions at best, and hence are also called "**weak-interactions**".
8. **Interaction**: Generally refers to an interconnection between two or multiple loci (anchors). Conceptually speaking, this is the genomic region defined by three or more overlapping tag-defined ChIP fragments. In data analysis, we defined interactions in the following way: as each inter-ligation PET was derived from two ChIP DNA fragments, the majority of which were less than 1,500 bp in size, we extended the mapped 20 bp tags to 1,500 bp along the reference genome to represent the virtual DNA fragments in pairs, and then we defined peaks (anchors) from the profile. We counted the number of PETs that connect two peaks (anchors) to each other. The number of PETs therefore, is the measurement of the frequency of an interaction between two regions, and 3 or more inter-ligation PETs was used as a cut-off to define an interaction.
9. **Anchor**: In the genome browser visualization, and conceptually speaking, this is the genomic region covered by three or more overlapping tag-defined ChIP fragments from inter-ligation PETs. In the data analysis, as each inter-ligation PET was derived from two ChIP DNA fragments, the majority of which were less than 1,500 bp in size, we extended the mapped 20 bp tags to 1,500 bp along the reference genome to represent the virtual DNA fragments in pairs, and then we defined peaks from the profile. These peaks are considered equivalent to "anchors", and are called as such.
10. **Duplex interaction:** The basic unit of chromatin interaction, involving two **Anchors** and one intermediate **Loop** region.
11. **Stand-alone duplex interaction**: Duplex interactions with anchors that do not overlap with anchor regions of other interactions. A stand-alone duplex is made up of only 1 interaction, and must have at least three overlapping inter-ligation PETs. As they are, on average, weaker than complex interactions, stand-alone duplex interactions are also called "**intermediate-interactions**".
12. **Complex interaction**: A complex interaction that has three or more anchors, indicating the interactions were further clustered upon further clustering of duplex interactions. As a complex interaction is made up of at least 2 interactions, each of which has at least 3 inter-ligation PETs, a complex interaction must have at least six overlapping inter-ligation PETs. Complex interactions are thus also called "**strong-interactions**". The 5' most and 3' most anchors on either end constitute the boundaries of the complex interaction.

13. **Interaction region**: Refer to a distinct interaction region after further clustering of duplex interactions. Includes stand-alone duplex interactions and complex interactions.
14. **Loop**: This is the genomic region between two adjacent anchors within an interaction region.
15. **Transcriptional unit**: A particular gene models which has a unique gene ID as given by the UCSC Known Gene database[13].
16. **TSS**: Transcription Start Site.
17. **Interaction-associated gene**: A gene with a TSS of a transcriptional unit that falls anywhere within the interaction boundaries of a complex interaction or duplex interaction plus 20 Kb (20 Kb upstream of the middle of the 5'-most anchor to 20 Kb downstream of the middle of the 3'-most interaction anchor).
18. **Anchor gene**: A gene with a TSS of a transcriptional unit that falls within $\pm$ 20 kb of the middle of any anchor in a complex interaction or stand-alone duplex interaction.
19. **Loop gene**: A gene with a TSS of a transcriptional unit that does not fall within $\pm$ 20 kb of the middle of any anchor in a complex interaction or stand-alone duplex interaction, but is within the interaction boundaries (and hence falls within the "loop" regions).
20. **Enclosed anchor gene**: An anchor gene where the TSS and the transcription end site of the same transcriptional unit are both entirely within the interaction boundaries.