The Chromatin-Looping Factor ZNF143 is Genetically Altered and Promotes the Oestrogen Response in Breast Cancer

by

Aislinn Treloar

A thesis submitted in conformity with the requirements for the degree of Master of Science
Department of Medical Biophysics
University of Toronto

© Copyright by Aislinn Treloar 2016
The Chromatin-Looping Factor ZNF143 is Genetically Altered and Promoters the Oestrogen Response in Breast Cancer

Aislinn Treloar

Master of Science

Department of Medical Biophysics
University of Toronto

2016

Abstract

Oestrogen signalling in breast cancer (BrCa) cells relies on chromatin-loops that connect distal regulatory elements bound by the oestrogen receptor 1 (ESR1) to target gene promoters. We show that chromatin-looping factor genes, including $CTCF$, $ZNF143$ and $RAD21$, are genetically altered in BrCa. Expanding on the function of CTCF and cohesin in BrCa, we demonstrate that ZNF143 binds promoters of most early-response oestrogen target genes connected to distal regulatory elements in ESR1-positive BrCa cells. Its chromatin occupancy is unaffected by oestrogen-stimulation, supporting a stable three-dimensional genomic architecture within the oestrogen response. Its loss abrogates the oestrogen-induced transcriptional response and growth of BrCa cells. Furthermore, we show that the overexpression of looping-factors within ESR-1 positive BrCa patients associates with a worse clinical outcome. Overall, our results suggest that ZNF143 is a new critical effector of the oestrogen response and highlights the contribution of the chromatin looping machinery to ESR1-positive BrCa development.
Acknowledgments

During the process of writing this thesis, I received help and support from many people. I would like to express my gratitude to them all in this acknowledgment.

Foremost, I would like to thank my supervisor Dr. Mathieu Lupien for giving me the opportunity to work on this project and who has been so generous with his time, knowledge and encouragement throughout the completion of this thesis.

My sincerest gratitude goes to the current and past members of the Lupien lab who contributed to my project. I thank Dr. Xue Wu and Dr. Xiaoyang Zhang for performing the ZNF143 ChIP-sequencing experiment, Dr. Nadia Penrod for performing RNA-sequencing analysis using the Cufflinks suite and both Dr. Swneke Bailey and Parisa Mazoorei for their guidance in learning several forms of data-analysis. Thanks to everyone in the lab for their encouragement, assistance and friendship during the pursuit of this degree.

I would like to thank the members of my supervisory committee, Dr. Brad Wouters and Dr. Senthil Muthuswamy for their helpful suggestions on this project.

Finally, I would like to acknowledge my friends and family for their encouragement and support along the way.
Table of Contents

ABSTRACT ........................................................................................................................................................................ II

ACKNOWLEDGMENTS .................................................................................................................................................................. III

TABLE OF CONTENTS .................................................................................................................................................................. IV

1 INTRODUCTION ..................................................................................................................................................................... 1

2 METHODS .................................................................................................................................................................................. 5

3 RESULTS ..................................................................................................................................................................................... 9

   3.1 ZNF143 BINDING OCCURS PRINCIPALLY AT PROMOTERS, INCLUDING AT THE MAJORITY OF THOSE OCCUPIED BY
       ESR1 9

   3.2 ZNF143 DIRECTLY REGULATES OESTROGEN-TARGET GENE TRANSCRIPTION IN ESR1-POSITIVE BREAST CANCER
       CELLS 12

   3.3 GENETIC ALTERATIONS IN CHROMATIN LOOPING FACTORS ARE FREQUENTLY OCCURRING AND TEND TOWARDS
       MUTUAL EXCLUSIVITY ............................................................................................................................................................ 15

   3.4 DIFFERENTIAL GENE EXPRESSION OF THE CHROMATIN LOOPING MACHINERY IS CLINICALLY RELEVANT AND RELATES
       TO GENETIC ALTERATIONS IN BREAST CANCER .................................................................................................................. 17

4 DISCUSSION & FUTURE DIRECTIONS ........................................................................................................................................ 20

5 SUPPLEMENTARY TABLES & FIGURES ....................................................................................................................................... 24

   5.1 SUPPLEMENTARY TABLES .................................................................................................................................................... 24

      Supplementary Table 1: E2-regulated genes in MCF-7 cells ................................................................................................. 24

      Supplementary Table 2: GSEA analysis of genes down-regulated upon ZNF143 depletion under E2
      stimulation .................................................................................................................................................................................... 26

      Supplementary Table 3: Genetic alterations tend towards mutual exclusivity (TCGA 2012) .............................................. 27

      Supplementary Table 4: Genetic alterations tend towards mutual exclusivity (TCGA Provisional
      Data Set) .................................................................................................................................................................................... 28

      Supplementary Table 5: Fold Change mRNA Expression upon Genetic Alteration of Chromatin
      Looping Factors (TCGA Provisional Data Set) .......................................................................................................................... 29

      Supplementary Table 6: Genetic alterations tend towards mutual exclusivity in lung squamous cell
      carcinoma (TCGA, Provisional) .................................................................................................................................................. 30
Supplementary Table 7: Genetic alterations tend towards mutual exclusivity in bladder urothelial carcinoma (TCGA, 2014) .............................................................................................................................................. 31

Supplementary Table 8: Genetic alterations tend towards mutual exclusivity in liver hepatocellular carcinoma (TCGA Provisional Data Set) .............................................................................................................................................. 32

Supplementary Table 9: Primers used in this study............................................................................................................................................................................................................................................................. 33

5.2 Supplementary Figures ................................................................................................................................................................................................................................................................. 34

5 REFERENCES ................................................................................................................................................................................................................................................................................................................. 40
1 Introduction

*Rhombin looping is involved in transcriptional regulation*

Gene expression is tightly controlled by spatially and temporally organized transcriptional programs to determine individual cell identity and function (Spitz and Eileen 2012). This is achieved through the coordinated interplay between transcription factor-bound promoters and distal regulatory elements, known as enhancers via long-range chromatin interactions that physically connect enhancers to their target promoter regions (Lieberman-Aiden 2009; Fullwood et al. 2009; Sanyal et al. 2012; Jin et al. 2013; Heidari et al. 2014; Rao et al. 2014; Plank & Dean 2014; Kron et al. 2014). The vast majority of long-range chromatin interactions are established during cellular differentiation and participate in the coordinated activation of cell-type specific transcriptional programs; however, a small subset can form upon response to external stimuli (Fraser et al. 2009; Sanyal et al. 2012; Jin et al. 2013; Rao et al. 2014).

*Enhancer-promoter loops are mediated by chromatin-looping factors*

The machinery regulating the formation of chromatin interactions consists of the DNA binding proteins CCCTC-binding factor (CTCF) and Zinc Finger Protein 143 (ZNF143), and the non-DNA binding cohesin complex, which must be recruited to the chromatin (Cubeñas-Potts & Corces 2015).

**CTCF:** CTCF is a ubiquitous and essential 11-zinc finger protein whose sequence is highly conserved (Klenova et al. 1993; Filippova et al. 1996; Burcin et al. 1997; Fedoriw et al. 2004). It binds tens of thousands of sites in the human genome (Chen et al. 2012), of which a small proportion are ultra-conserved between mammalian species (Schmidt et al. 2012) while 50-60% show cell-type specificity (Barski et al. 2007; Kim et al. 2007; Chen et al. 2008; Cuddapah et al. 2009). Although vertebrate CTCF has traditionally been implicated as an insulator protein, recent reports support CTCF’s role as a chromatin-looping factor. It regulates genome topology as a looping factor through two main mechanisms. First, CTCF contributes to the partitioning the genome into regulatory blocks, or topologically associated domains (TADs) (Nora et al. 2012; Phillips-Cremins et al. 2013; Rao et al. 2014; Lupianez et al. 2015; Guo et al. 2015; Barutcu et al. 2015; Ji et al. 2015). Chromatin interactions
enrich within these TADs but are relatively rare across TAD boundaries (Dixon et al. 2012; Hou et al. 2012; Sofueva et al. 2013). These sites may thus restrict enhancer-promoter interactions and establish functional domains of gene expression. In addition to this role in TAD boundaries, CTCF is also implicated in the cell-type specific interactions that occur within TADs. Enhancer elements are enriched for CTCF binding (Song et al. 2011; DeMare et al. 2013), indicating that a subset of CTCF sites may be important in shaping cell-type specific transcriptional programs. A significant overlap observed between cell-type specific CTCF binding sites and enhancer elements in the human genome supports this (Barski et al. 2007). Interaction analysis in 3 cell lines indicates that distal fragments that loop to promoter fragments are enriched for CTCF and active enhancer histone tail modifications (Sanyal et al. 2012). Furthermore, interaction analyses focused on CTCF binding sites in several cell types indicate that CTCF-bound fragments interact with gene promoters (Heidari et al. 2015; Guo et al. 2012) and that enhancer-promoter looping may require CTCF (Hirayama et al. 2012). These results indicate that CTCF function may be context-dependent, either defining TAD boundaries, or targeting cell-specific enhancers for looping events.

**The cohesin complex:** The cohesin complex is made up of several subunits, Rad21, SMC1A, SMC3 and either STAG1 or STAG2. Together, they form a ring-like structure approximately 40nm in diameter that encircles DNA fibers (Losada 2014). The complex contributes to DNA replication, as it promotes restart of replication forks that stall at regions that are difficult to replicate, has a role in facilitating doubles-strand break repair by homologous recombination (Remeseiro et al. 2012; Carretero et al. 2013) and is also involved in sister chromatid cohesion during mitosis (Uhlmann 2004). As the cohesin complex lacks a DNA binding domain, 60-80% of cohesin binding sites are dependent on CTCF for the complex’s recruitment to the chromatin (Wendt et al. 2008; Schmidt et al. 2010) where it is lost upon siRNA mediated depletion of CTCF expression (Parelho et al. 2008; Wendt et al. 2008; Hou et al. 2010). CTCF and cohesin occupancy strongly correlate with long-range interactions (Sanyal et al. 2012; Phillips-Cremins et al. 2013; Heidari et al. 2014) and the depletion of either of these factors destabilizes chromatin interactions (Splinter et al. 2006; Hadjur et al. 2009; Hou et al. 2010; Kagey et al. 2010; Sofueva et al. 2013; Li et al. 2013; Phillips-Cremins et al. 2013; Zuin et al. 2014). Cohesin may also have a CTCF-independent role in chromatin looping as it occupies a subset of active enhancer-promoter interactions that lack CTCF (Schmidt et al. 2010; Kagey et al. 2010; Phillips-Cremins et al. 2013). These sites often harbour the Mediator complex, a transcription co-activator, and
Nipb1, cohesin’s loading factor, in embryonic cells (Kagey et al. 2010); however the role for either of these factors in loop formation has neither been fully characterized in these cells, nor suggested in other cell types.

**ZNF143:** ZNF143 is a ubiquitously expressed 7 zinc-finger protein that has been described as a vertebrate transcriptional activator involved in RNA pol II-dependent transcription (Myslinski et al. 1998; Faresse et al. 2012). Although ZNF143’s requirement for growth and development is not well known in mammals, it is essential for normal development in zebrafish (Halbig et al. 2012). Analysis of ZNF143 binding across 4 mammalian genomes suggests that ZNF143 constitutes one of the most widespread transcription factor binding-sites in mammalian promoters (Myslinski et al. 2006), with binding in approximately 2000 mammalian protein-coding genes. Its enrichment in chromatin interaction anchors along with the cohesin complex and CTCF indicates that it may be involved in chromatin looping (Rao et al.; Bailey et al. 2015; Heidari et al. 2015). Indeed, ZNF143 binding at gene promoters relates directly to cell-specific interactions and the disruption of ZNF143 binding to the chromatin through genetic variation decreases chromatin interaction frequency and diminishes target-gene expression (Bailey et al. 2015).

**Oestrogen signalling relies on chromatin-looping**

Chromatin looping is a key mechanism in oestrogen receptor 1 (ESR1)-mediated transcriptional regulation in breast tumours that express oestrogen (classified as “luminal”) (Pan et al. 2008; Fullwood et al. 2009; Li et al. 2012; Zhang et al. 2010; Li et al. 2013). Accordingly, CTCF and the cohesin complex have been implicated in the ESR1-mediated response of breast cancer cells to oestrogen in a genome-wide manner. CTCF-bound sites involved in chromatin interactions demarcate the oestrogen response of the TFF1 and TFF3 genes, with CTCF silencing preventing E2-upregulation of gene expression. In addition, depletion of the cohesin complex subunit Rad21 abrogates the oestrogen response by interfering with the formation of chromatin interactions at oestrogen-regulated genes (Li et al. 2013) and by blocking growth of breast cancer cells (Schmidt et al. 2010). Both CTCF and the cohesin complex occupy genomic regions where ESR1 is recruited following oestrogen stimulation (Schmidt et al. 2010; Ross-Innes et al. 2011; Hah et al. 2013). These sites are primarily distal to promoters, and the factors that target promoters for ESR1-mediated long-range chromatin interactions
have yet to be elucidated. The characterization of ZNF143 in several cell types as a promoter-bound looping factor suggests that it may provide for this integral component of ESR1 signalling.

In the present study we identify ZNF143 as a key regulator of ESR1-signalling that enriches at oestrogen-responsive gene promoters involved in chromatin interactions. We show that it is required for oestrogen-induced gene transcription and for the luminal breast cancer growth response to oestrogen stimulation. We show that the chromatin interaction machinery, including CTCF, ZNF143 and the cohesin complex, is genetically altered in breast tumours and that the expression of these factors is clinically relevant in luminal breast cancer patients.
2 Methods

Cell Culture:
MCF-7 cells were cultured as previously described (Magnani et al. 2011). Briefly, cells were maintained in DMEM (Life Technologies) supplemented with 10%FBS and 1% Penicillin Streptomycin.

siRNA Transfection of MCF-7 breast cancer cells:
MCF-7 cells were maintained in phenol red free DMEM medium (Life Technologies) supplemented with 10% heat-inactivated CDT-FBS, 1% Penicillin Streptomycin and 1mM Sodium Pyruvate prior to transfection, as described previously (Lupien et al 2008). Following two days of oestrogen starvation, cells were transfected with siZNF143 #1 (Ambion siRNA ID: s15192), siZNF143 #2 (Ambion siRNA ID: s15194), or a negative scrambled control siRNA (Ambion cat: 4404020). Transfection was performed using the RNAiMAX reagent (Invitrogen) according to the manufacturer’s instructions. For cell proliferation assays, cell number or O.D. (562nm) was determined every 24hrs post E2-stimulation (10µM 17β oestradiol). For expression and protein assays, RNA was extracted 3hrs following 10µM, 17β oestadiol stimulation.

RNA preparation/collection and real-time PCR:
Total RNA extraction was performed using the RNeasy Plus Mini Kit (Qiagen) according to the manufacturer’s instructions. cDNA was synthesized using the iScript cDNA synthesis kit (Bio-Rad) and real-time PCR was performed using the SensiFAST SYBR No-ROX master mix (BioLine) and primers listed in Supplementary Table 9.

RNA-seq:
RNA-seq reads were mapped to the human reference genome (hg19) using TopHat2 with default parameters (Kim et al. 2013). The Cufflinks suite was used to assemble transcripts, estimate transcript abundance, and test for differential expression (Trapnell et al. 2010). Specifically, differential expression was called using CuffDiff2 (Trapnell et al. 2013) with the following parameters: -u to
correct for reads mapping to more than one location and –b to correct for fragment bias (Roberts et al. 2011).

**TCGA and BCCRC Xenograft data for breast cancer genetic alterations:**

The DNA copy number, RNA-sequencing and genetic alteration datasets of 482 TCGA published (TCGA 2012), 962 TCGA provisional (2015) and 29 BCCRC Xenograft (Eirew et al. Nature 2014) breast cancer samples used in this work were obtained from the CBio Cancer Genomics Portal (Cerami et al. 2012; Gao et al 2013). Copy number of each gene was generated from the algorithm GISTIC (Genomic Identification of Significant Targets in Cancer) (Beroukhim et al. 2007). Somatic mutation data were obtained from whole-genome shotgun sequencing for the BCCRC Xenograft dataset (Eirew et al. 2014) and exome-sequencing for the TCGA Published and Provisional datasets (TCGA 2012). Figures were recreated using GraphPad Prism Version 5.00 for Mac OS X, (GraphPad Software, San Diego California USA, www.graphpad.com). Datasets are available through CBioPortal (http://www.cbioportal.org).

**ChromHMM/Genomic distribution of binding:**

The ChromHMM algorithm (Ernst et al. 2011) was used to train a multivariate Hidden Markov Model. ChIP-Seq reads for eight histone modifications in E2-stimulated MCF-7 cells (H3K27ac, H3K9ac, H3K4me3, H3K14ac, H3K4me1, H3K27me3, H3K9me3 and H3K36me3) (http://genome.ucsc.edu/ENCODE/downloads.html) were transformed into binary values using a default 200-bp bin size, and the LearnModel function was used to learn models and to assign each genomic position to one of twelve chromatin states. Biological relevance for each state was assigned based on previous studies (Ernst et al. 2011).
**ChIP-seq**

ChIP assays were performed as previously described (Lupien et al. 2010). Briefly, five million cells were crosslinked with 1% Formaldehyde and lysed. Chromatin was sonicated by biorupter and immunoprecipitated with 4µg anti-ZNF143 (Novus Biologicals H00007702-M01), followed by reverse crosslinking. DNA was extracted using MinElute PCR purification kit (Qiagen 28004), processed with library construction using NEB ChIP-seq library prep reagent set (E6200S), and then sequenced by HiSeq 2000. Sequencing reads were aligned to hg19 by Bowtie and ZNF143 binding peaks were called by MACS2.0 (Zhang et al. 2008). All assays were conducted in duplicates. These were pooled to improve peak detection as duplicates for each condition were highly correlated with one another (R=0.844 vehicle; R=0.928 E2-stimulation) (Supplementary Figure 1).

**Statistical Analysis:**

P-value was determined using Student’s t-test with two-tail distribution. Kaplan-Meier analyses (http://www.kmplot.com) were used to assay differences in overall survival (Györffy 2010). Prism and Omnigraffle were used for figures.

**Overlap Analysis and genome structure correction (GSC):**

The intersections between the binding sites for ZNF143, ESR1 and ChIA-PET regions were performed using the BEDTools software package (Quinlan et al. 2010). Overlapping binding sites were defined by having at least one base pair in common. Genome structure correction (GSC) statistic (Birney et al. 2007; Bickel et al. 2011) was run to establish the significance of the overlap of ZNF143 with E2-regulated promoters involved in loops. The software was run using region fraction = 0.2, sub-region fraction –S = 0.4 and –bm as statistic test. All DNAse I Hypersensitive sites in E2-stimulated (10nM E2, 45min) MCF-7 cells were used as the null list (He et al. 2012).

**Gene Set Enrichment Analysis**

Biological pathways disrupted by ZNF143 depletion in MCF-7 cells were investigated by performing a gene set enrichment analysis (GSEA), software that is available from the Broad Institute (Subramanian
et al. 2005; Mootha et al. 2003). The hallmark gene sets were used for enrichment analysis (www.broadinstitute.org/gsea).
3 Results

3.1 ZNF143 binding occurs principally at promoters, including at the majority of those occupied by ESR1

The role for CTCF and the cohesin complex in ESR1 signalling has been established in breast cancer cells (Liu and Cheung 2014). However, ZNF143’s contribution to the progression of this cancer type is unknown. We therefore determined the genome-wide chromatin binding profile of ZNF143 in ESR1-positive breast cancer cells through chromatin immunoprecipitation (ChIP) followed by massively parallel sequencing (ChIP-seq) in MCF-7 cells. This was performed before (vehicle control) and after stimulation with 17β-oestradiol (E2), a potent form of oestrogen (Supplementary Figure 1). We identified a total of 76,802 and 73,798 high-confidence (p≤1.0x10^-5) ZNF143 chromatin-binding sites in vehicle and E2-stimulated conditions, respectively. We observe a strong correlation (R = 0.809) in ZNF143’s binding affinity for the chromatin under these two conditions (Figure 1A, upper panel), suggesting that ZNF143 is bound to the chromatin prior to E2 stimulation in ESR1-positive breast cancer cells and that ZNF143 remains stably bound at these sites following hormonal stimulation. The ZNF143 binding regions in MCF-7 breast cancer cells employed for all subsequent analysis were those identified using the merged sequencing files.

To determine the genomic distribution of ZNF143 binding events, we first defined chromatin states in MCF-7 cells based on the ChromHMM algorithm (Ernst and Kellis 2012) using the ChIP-Seq data for Histone 3 Lysine 27 acetylation (H3K27ac), H3K9ac, H3K14ac, Histone 3 Lysine 4 trimethylation (H3K4me3), H3K4me1, H3K27me3, H3K9me3 and H3K36me3 generated in E2-stimulated MCF-7 cells (Joseph et al. 2010; Magnani et al. 2013; Li et al. 2013). We applied a 12-state chromatin model to segment the genome and then grouped predicted functional elements as promoters, enhancers, transcribed, repressed, CTCF or no signal regions (Supplementary Figure 2). In agreement with reports revealing a strong enrichment for ZNF143 binding at promoters in other cell types (Bailey et al. 2015), we find 38,320 (43%) ZNF143 bound sites at promoters (Figure 1A, lower panel). An additional 16,232 sites (18%) map to enhancers, 3,173 sites (3.6%) to transcribed regions, 2,073 sites (2.3%) to
repressed regions, 7,410 (8.3%) to CTCF regions, and the remaining 22,486 sites (26%) fall in regions with no signal in our ChromHMM segmentation model (Figure 1A, lower panel). The bias of ZNF143 binding at promoters is further highlighted by its significantly increased binding intensity at promoters over other genomic regions (Figure 1B).

ESR1 recruitment to the chromatin following E2-stimulation in MCF-7 cells has previously been reported to occur primarily away from promoters (Carroll et al. 2006; Lin et al. 2007; Welboren et al. 2009). These assessments were performed based on the position of annotated genes across the reference human genome as opposed to chromatin states. Indeed, using this approach, we map less than 5% (1,168) ESR1 binding sites to promoters (Figure 1C, upper panel). However, the proportion of ESR1-bound promoters increases to 27.4% (6,406 sites) when using chromatin state to define genomic elements in MCF-7 cells (Figure 1C, lower panel). ESR1 binding is still predominantly (33.8%; 7,907 sites) found at enhancers (Figure 1C) and further assessment of ESR1 called peaks indicates that only 10% of binding sites are within ±2.5kb from the TSS of coding genes; this is in line with reports of a subset of ESR1 binding sites that are called in proximity of promoters when co-localized with CTCF (Ross-Innes et al. 2011). The majority of ESR1 sites that are called in promoter regions are so far un-annotated and do not fall within ±2.5kb of TSS of IncRNAs or coding genes.

Comparing ZNF143 and ESR1 binding profiles reveals over 8,671 shared binding sites (Supplementary Figure 3). This translates into 72% of ESR1-bound promoter regions occupied by ZNF143 prior to and after E2-stimulation (4,782 of 6,647 ESR1 promoter bound regions)(Figure 1D). Over 36% of ESR1-bound enhancers are also occupied by ZNF143 binding in MCF-7 cells but transcribed, CTCF and repressed ESR1-bound regions show minimal overlap with ZNF143 binding sites (15%, 4% and 18% of ESR1 sites, respectively) (Figure 1D). We find that ESR1 binding is significantly stronger at enhancer regions than promoters occupied by ZNF143 (p=1x10^-3). In contrast, ZNF143 binds the ESR1–bound promoters with greater affinity (p<1x10^-3) compared to enhancers (Figure 1E). These results are in line with primary ESR1 binding occurring at enhancers, while primary ZNF143 binding occurs at promoters.
Figure 1: Genome-Wide Binding of ZNF143 Suggests a Role in E2-Induced ESR1 Recruitment to Promoter Elements

Scatterplot comparing read count for ZNF143 binding in MCF-7 breast cancer cells treated with vehicle or E2 (10µM) for 45 minutes. Scatterplot was generated with SeqMonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) from ZNF143 ChIP-sequencing data. R, Pearson’s correlation, was calculated from all data points (Upper Panel). Genomic distribution of ZNF143 binding events in MCF-7 cells. Chromatin states were annotated by the ChromHMM algorithm (Supplemental Figure S2) (Ernst et al. 2012) (Lower Panel) (A). Boxplots representing the signal intensity (MACS score) for ZNF143 at promoter, enhancer, transcribed, CTCF ad no-signal regions of the genome. P-value was calculated using the Mann-Whitney test (B). Global genomic distribution of ESR1 binding events called in MCF-7 cells under E2 stimulation. Genomic elements were annotated by the Cis-Regulatory Element Annotation System (CEAS) web application (cistrome.org/ap) or by the ChromHMM algorithm (Supplemental Figure S2) (C). Histogram illustrating the proportion of ESR1-bound sites that are occupied by ZNF143 at promoters, enhancers, transcribed, repressed and CTCF regions (D). Boxplots showing the signal intensity of ZNF143 and ESR1 at shared binding sites called in promoters or enhancers, as defined by ChromHMM. P-values were calculated using the Mann-Whitney test (E). (*p<0.05; ** p<0.01; *** p<0.001)
3.2 ZNF143 directly regulates oestrogen-target gene transcription in ESR1-positive breast cancer cells

To determine how ZNF143 occupancy relates to chromatin interactions at E2-regulated promoters, we first defined E2-responsive genes as genes whose expression, measured by RNA-seq, is significantly altered following E2-stimulation for three hours (FC > 1.5; p=0.05). This identified 194 E2-regulated genes (Supplementary Table 1). We then mined the RNA Polymerase II (Pol II) Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) dataset (Li et al. 2012) to identify E2-upregulated gene promoters that form chromatin interactions. A total of 223 promoters ascribed to the 194 E2-regulated genes were assessed in this analysis. We identified 137 (61%) E2-upregulated genes with at least one chromatin interaction anchored at the promoter (±2.5 kilobases (kb) from the Transcription Start Site (TSS)) (Figure 2A, upper panel). ZNF143 binds to 110 (80%) of the 137 E2-regulated gene promoters, a significantly higher proportion (P<1.0×10^{-4}) than expected by chance (Figure 2A, lower panel). These results indicate ZNF143 binding is enriched at E2-regulated gene promoters involved in chromatin interactions, supporting a direct contribution of ZNF143 to chromatin interactions associated with the E2-response in breast cancer.

We then further delineated the requirement for ZNF143 in the E2 response, specifically in the regulation of E2 target gene expression. We performed RNA-seq following E2-stimulation in MCF-7 cells depleted of ZNF143 using siRNAs. We found that 91(47%) of the 194 E2-regulated genes were no longer responsive to E2-stimulation following ZNF143 depletion (Figure 2B-C). Among the genes whose expression remained responsive to E2-stimulation, 32 genes (16%) showed significantly reduced expression (p<0.05) following ZNF143 depletion compared to control cells (Figure 2B-C). The impact of ZNF143 depletion on E2-target gene regulation relates to ZNF143 binding intensity at promoters; we found that genes unaffected by the loss of ZNF143 (No Change gene category) showed weak ZNF143 binding at their promoters whereas genes that showed reduced expression or loss of induction to E2-stimulation following ZNF143 depletion harbourered significantly stronger ZNF143 binding at their promoters in control conditions (p= 4.0×10^{-3}; p=7.1×10^{-3}) (Figure 2D). These findings suggest that the genes that are unaffected by the loss of ZNF143 do not require ZNF143 at their promoters to
regulate expression whereas the genes that exhibit either loss of induction or reduced expression when ZNF143 is lost do rely on ZNF143 occupancy at their promoters. We further interrogated patterns of mRNA expression in MCF-7 cells depleted of ZNF143 by performing Gene Set Enrichment Analysis (GSEA) for genes whose expression was significantly down-regulated upon ZNF143 depletion (FC >1.5, p <0.05) under E2-stimulated conditions. This revealed enrichment for early and late oestrogen stimulation gene sets (p= 7.21x10^{-15} & p= 3.92x10^{-9}) (Supplementary Table 2).

MCF-7 cells rely on E2-stimulation to activate the ESR1-regulated gene expression program, which mediates their re-entry into cell cycle and drives proliferation. We therefore assessed the requirement for ZNF143 in the E2-induced growth of MCF-7 luminal breast cancer cells following its siRNA-based depletion (Figure 2E). ZNF143 depletion significantly impaired growth of MCF-7 cells under E2-stimulation (Figure 2F), supporting its central role in the oestrogen response. Overall, these results show that ZNF143 is required for the ESR1-mediated transcriptional response and for the E2-induced growth response of luminal breast cancer cells.
Figure 2

Proportion of E2-responsive gene promoters (+/-2.5kb from TSS) that are involved in chromatin interactions associated with RNA polymerase II in E2-stimulated MCF-7 cells (upper panel). Scatterplot illustrating the proportion of these promoters that exhibit ZNF143 binding (lower panel) (A). Volcano plot showing changes in gene expression of E2-responsive genes upon E2 stimulation in cells depleted of ZNF143 (pink) or transfected with scrambled siRNA (green). Each circle represents one gene. The log fold change in gene expression in E2-stimulated versus vehicle treated conditions is represented on the x-axis. The y-axis shows the -log10 of the p-value. A p-value of 0.05 and a fold change of 1.5 are indicated by grey lines. Quadrants 1-6 are indicated (B). Case examples of genes that fall within quadrants 1 and 5 that illustrate gene expression changes under vehicle and E2-stimulated conditions. P-value calculated using student’s t-test (C). Boxplots illustrating signal intensity of ZNF143 at promoters of genes that show no change, reduced expression or loss of E2-induction upon ZNF143 depletion. P-value calculated using Kruskal-Wallis test (D). ZNF143 depletion via siRNA significantly reduces its mRNA levels 48hrs post-transfection. P-value calculated using student’s unpaired t-test (E). MCF-7 breast cancer cells depleted of ZNF143 fail to grow in response to E2 stimulation compared to control cells. P-value calculated using student’s unpaired t-test (F) (*p<0.05; **p<0.01; *** p<0.001; NS = Not Significant.)
3.3 Genetic alterations in chromatin looping factors are frequently occurring and tend towards mutual exclusivity

Pan-cancer analyses led by The Cancer Genome Atlas (TCGA) have revealed that CTCF is significantly mutated in several cancers, including breast (Lawrence et al. 2014). Having established a role for ZNF143 in ESR1 signalling in luminal breast cancer cells, these reports prompted us to assess the extent of genetic alterations to the known chromatin looping factors in breast cancer. We investigated the frequency of genetic alterations in CTCF, ZNF143 and the cohesin complex subunits (Rad21, STAG1, STAG2, SMCA1 and SMC3) in the published and provisional breast cancer datasets from TCGA as well as in the Primary Derived Xenograft (PDX) populations from the British Columbia Cancer Research Centre (BCCRC) (TCGA 2012; Eirew et al. 2014). Collectively, the chromatin-looping factors harbour genetic alterations (amplification, deletion or somatic mutation) in at least 22% of primary breast tumours and up to 65% of PDXs derived from primary and metastatic breast tumours (Figure 3A-B). Although we observe a trend towards mutual exclusivity between genetic alterations in the chromatin-looping factors across samples from the TCGA provisional and published datasets, 3 gene pairs do show significant co-occurrence in the TCGA provisional dataset (Figure 1C, Supplementary Table 3-4).

Finally, we assessed the relevance for these genetic alterations to breast cancer subtype as defined by the PAM50 gene expression classifier (Parker et al. 2009) in the TCGA published and provisional data sets. Although genetic alterations to ZNF143 and CTCF seem to enrich in luminal cancers (3/3 and 15/18, respectively) in the published data set, this observation is not reproduced in the provisional dataset; genetic alterations in each of the chromatin looping factors are found in the other breast cancer subtypes. These results suggest that in addition to the chromatin looping machinery being altered in luminal cancers, where it contributes to ESR1 signalling, the dysregulation of these factors may also contribute to disease in the other breast cancer subtypes.
Copy number alterations and somatic mutations in *CTCF*, *ZNF143* and five members of the cohesin complex were analyzed in three separate breast cancer datasets. DNA amplification, deletion, mutation or multiple alterations are indicated (A). Overall frequency of genetic alterations in the chromatin looping factors specified in A, in three breast cancer data sets (B). For each breast cancer data set, the proportion of samples with genetic alterations in a single factor versus in multiple factors is indicated by the strength of colour (C). Breast cancer subtype of tumours harbouring a genetic alteration in a chromatin looping factor; subtype determined by gene expression of 50 genes (PAM50) (D). (*p<0.05; ** p<0.01; *** p<0.001)
3.4 Differential gene expression of the chromatin looping machinery is clinically relevant and relates to genetic alterations in breast cancer

We determined the relevance of genetic alterations in chromatin looping factors to their expression by segregating breast tumour samples based on the GISTIC-based copy-number alteration score (Mermel et al. 2011). We observe a significant decrease in ZNF143 expression in tumours harbouring heterozygous and homozygous deletions compared to diploid cases (p=2.0x10^{-3} and p<1.0x10^{-3} respectively) (Figure 4A), while copy number gains correlate with significantly increased ZNF143 expression (p<1.0x10^{-3}). Similar results were obtained for all other chromatin-looping factors with the exception of STAG2 (Figure 4A; Supplementary Table 5). These results suggest that copy number variations in the chromatin looping factors directly impacts their expression.

To address the clinical relevance of differential expression of looping factors in luminal breast cancer we performed Kaplan-Meier analysis using the METABRIC dataset consisting of close to 2,000 expression profiles from independent, clinically annotated breast cancer samples (Curtis et al. 2012). Using the KMplot tool (http://kmplot.com/private/) (Győrffy et al. 2010) we segregated samples based on low versus high levels for each chromatin looping factor. Kaplan-Meier curves focused on overall survival reveal that ESR1-positive breast cancer patients with elevated ZNF143 expression do worse than those with low expression levels (p=1.1x10^{-3}) (Figure 4B). This observation is valid across all ESR1-positive breast cancer patients as well as within luminal A or B subtypes (p=4.3x10^{-2} & p=7.0x10^{-4}, respectively)(Figure 4B). This suggests that ZNF143 expression does not simply discriminate luminal A from Luminal B cancers, but that high expression correlates with more aggressive disease within each subtype. The association of elevated gene expression with more aggressive breast cancers is also observed for other chromatin-looping factors. For instance, the overall survival of breast cancer patients whose tumours expressed high levels of CTCF or the cohesin subunit Rad21 is worse than for patients whose tumours express these factors at lower levels (p=6.3x10^{-6} & p=1.3x10^{-3}, respectively across all ESR1-positive breast cancer patients). CTCF expression discriminates poor outcome within both Luminal A and Luminal B subtypes (p=4.3x10^{-4} & p=7.7x10^{-4}, respectively) (Figure 4B), while RAD21’s expression is only predictive in the more aggressive Luminal B subtype (p=0.027) (Supplementary Figure 5). These results suggest that the expression levels of
these chromatin-looping factors are relevant to the clinical outcome of ESR1-positive breast cancers patients.
Figure 4: Chromatin Looping Factor Expression is Affected by Genetic Alterations and is Clinically Relevant in Luminal Breast Cancer

Box-and-whisker plots showing mRNA expression for the chromatin looping factors assessed in A-C that have altered copy number status, as determined from GISTIC. Gene mutation status and mRNA expression were analyzed using publically available data obtained through the cBioPortal for Cancer Genomics. P-value is calculated using Mann-Whitney test (*p<0.05; ** p <0.01; *** p <0.001) (A). Kaplan-Meier plots derived from the METABRIC data set (Curtis et al. 2012) evaluating overall survival in ESR1-positive breast cancer patients (n=1486), Luminal A patients (n=825) and Luminal B (n=668) patients, stratified by ZNF143 or CTCF expression. Data were obtained from the Kaplan-Meier plotter breast cancer survival analysis database (Györffy 2010). Hazard ratios (HR) and logrank P-values are displayed (B).
4 Discussion & Future Directions

Chromatin interactions regulate transcriptional networks that drive differentiation and cell-specific responses to stimuli (Fraser et al. 2009; Sanyal et al. 2012; Jin et al. 2013; Rao et al. 2014). Dysregulation of transcriptional regulation mechanisms and consequent changes to gene expression networks are central to tumourigenesis and disease progression (Kolch et al. 2015). Using three different studies characterizing genetic alterations in breast tumours, we show that the chromatin interaction factors that are known to regulate chromatin loops (ZNF143, CTCF and the subunits of the cohesin complex) are frequently genetically altered. Furthermore, using an independent dataset, we show that elevated expression of these factors typifies aggressive ESR1-positive breast tumours. These results expand on the reported significant mutational load in CTCF and the subunits of the cohesin complex in some solid and haematological cancers (Lawrence et al. 2014). We find that genetic alterations in chromatin-looping factors tend towards mutual exclusivity in breast cancer, an observation shared with lung squamous cell carcinoma (TCGA, Provisional), bladder urothelial carcinoma (TCGA 2014) and liver hepatocellular carcinoma (TCGA, Provisional) (Supplementary Tables 6-8) The tendency towards mutual exclusivity in mutations targeting chromatin-looping factors supports previous reports showing that distinct mutations identified in tumours can converge on proteins involved in a central oncogenic pathway (Leiserson et al. 2015). In addition to mutations that directly target genes, genetic variance in the sequences of regulatory elements can alter looping machinery binding to the chromatin and predispose the cell to improper gene expression. For instance, CTCF/cohesin binding sites in or adjacent to the CTCF motif are frequently mutated in colorectal cancers (Katainen et al. 2015) and deletion of CTCF/cohesin co-bound sites in ESCs alter interaction frequencies of enhancers with proximal genes and can dysregulate their expression (Dowen et al. 2014). These findings indicate that looking beyond the genetic alterations in looping factors to changes that occur in regulatory elements harbouring ZNF143 and CTCF binding motifs in breast cancer may further elucidate the extent to which chromatin looping factor activity is affected in tumours.

Recent studies suggest that looping structures are heterogeneous and are mediated by different combinations of chromatin interaction proteins (De Laat and Duboule 2013). Chromatin interactions
between two CTCF-bound regions are associated with long (>1Mb) stable interactions conserved across cell types that partition the genome into regulatory blocks (Zhang et al. 2012; Phillips-Cremins et al. 2013; Heidari et al. 2014, Rao et al. 2014; Dowen et al. 2014). Interactions that connect enhancer and promoter elements are associated with shorter, more cell-type specific interactions and generally occur within CTCF-CTCF interaction boundaries (Schmidt et al. 2010; Kagey et al. 2010; DeMare et al. 2013; Heidari et al. 2014; Dowen et al. 2014; Bailey et al. 2015). They are believed to be mediated by CTCF at distal sites and ZNF143 at promoters (Cubeñas-Potts and Corces 2015). The dynamics of these enhancer-promoter interactions is subject to debate, particularly those that mediate the transcriptional response to external stimuli. Chromatin Conformation Capture (3C) assays in MCF-7 cells depleted of ESR1 indicates that this receptor is required at specific chromatin interaction anchors to activate transcription of certain E2-regulated genes (Pan et al. 2008; Fullwood et al. 2009). HiC analysis in stimulated MCF-7 cells also indicates that E2-stimulation increases the frequency of interactions (Mourad et al. 2014), with the largest changes in interaction frequency occurring at ESR1 binding sites. Finally, locus-specific studies have indicated that E2-stimulation can increase the frequency of specific promoter-enhancer interactions (Li et al. 2013). Taken together, these results would suggest that chromatin interactions involving the ESR1 are transient and are induced by E2-stimulation. These studies however, stand in contrast to the reports that indicate enhancer-promoter interactions are established during cellular differentiation, do not require transcription to remain stable and are mostly pre-formed prior to external stimulation (Fraser et al. 2009; Sanyal et al. 2012; Jin et al. 2013; Rao et al. 2014). Our work shows that ZNF143 is present at promoters interacting with distal ESR1-bound sites and is required for the ESR1-mediated response to E2-stimulation. ZNF143 is present at E2-regulated gene promoters prior to stimulation and remains bound upon hormone treatment. A similar observation was reported for CTCF in MCF-7 breast cancer cells (Ross-Innes et al. 2011). Together, these results are consistent with chromatin looping factors stably bookmarking the chromatin interaction anchors to predispose the cells to a transcriptional response guided by a three-dimensional chromatin configuration. Interestingly, the ZNF143 motif is found at the majority of mammalian promoters (Myslinski et al. 2006) and CTCF binding is highly conserved across tissues (Schmidt et al. 2012). Furthermore, the chromatin looping machinery is genetically altered in each breast cancer subtype (Figure 3C) as well as in many different cancers-types (Supplementary Figure 6), indicating that the looping machinery contributes to cell-specific transcriptional regulation. Therefore, how the looping machinery establishes cell-type specific interactions in a normal cell and how the
alterations to these interactions affects gene expression regulation in the context of distinct cancers remain important questions to be addressed.

The ESR1-mediated transcriptional response is a major driver of growth and proliferation in a majority of diagnosed breast cancer cases (Tyson et al. 2011) and is accordingly an important focus for targeted therapy (Renoir et al. 2013). Although the treatment of breast cancer has benefitted tremendously from the generation of therapies against ESR1 activity, disease relapse continues to pose a challenge due to intrinsic or acquired drug resistance. Better understanding of the mechanism of ESR1 signalling may provide alternative treatment avenues. Our work identifies the chromatin-looping machinery, inclusive of ZNF143, as a central player in ESR1 signalling in breast cancer cells. As targeting zinc-finger proteins, either ZNF143 or CTCF, is technically challenging, a comprehensive identification of the chromatin-looping machinery is warranted. In Drosophila, accessory proteins (CP190, Rad21, Mdg4, CapH2, condensin factor, Fsn1h-L and L3mbt) that interact with and assist DNA binding chromatin-looping factors have been identified (Moshkovich et al. 2011; Gurudatta et al. 2013; Kellner et al. 2013; Van Bortle et al. 2014; Vogelmann et al. 2014). However, the characterization of chromatin-looping accessory proteins has been less extensive in mammals. In humans, the cohesin complex proteins serve as accessory protein to the chromatin-looping factors (Cubenas-Potts and Corces 2015) and recent reports suggest that the Mediator complex may assist in the formation of chromatin interactions. The Mediator enriches at enhancer-promoter contact sites and MED1 or MED12 subunit depletion alters chromatin interactions in embryonic stem cells (Kagey et al. 2010; Phillips-Cremins et al. 2013). Additional factors such as Nipbl, which can load cohesin on the chromatin, and CTCF-interacting proteins YYA, Kaiso, CHD8, PARP1, MAZ, JUND, nuceophosmin, PRDM5 and TF-II might also prove critical in targeting the chromatin-looping machinery in cancer (Kagey et al. 2010; Cubenas-Potts & Corces 2015). Finally, enhancer RNAs (eRNAs) have been suggested to participate in the formation and stability of chromatin interactions (Li et al. 2013; Hah et al. 2013). However, chromatin loops can still exist in the absence of eRNA production (Hah et al. 2013). Although our work focuses on enhancer-promoter loops, the delineation of the full complement of factors involved in the different types of interactions may provide us with a better understanding of the mechanism through which chromatin interactions regulate gene transcription.
Overall, we establish the chromatin-looping factor ZNF143 as a key regulator of ESR1 signalling in luminal breast cancer cells. We show that the chromatin-looping machinery, inclusive of ZNF143, is altered in over 20% of breast tumours and that the expression of these looping factors is of clinical relevance to breast cancer patients. This work expands our understanding of the mechanism that drives two-thirds of breast cancer cases. The prevalence of alterations to the chromatin-interaction machinery across multiple cancer types is indicative that these factors likely contribute to transcriptional regulation beyond ESR1 signalling. Thus, in addition to established genetic and epigenetic contributions, the three-dimensional architecture of the chromatin may prove to be another important regulator of the hallmark altered gene transcription regulation seen in cancer.
## 5 Supplementary Tables & Figures

### 5.1 Supplementary Tables

**Supplementary Table 1**: E2-regulated genes in MCF-7 cells

<table>
<thead>
<tr>
<th>E2 Up-Regulated Genes</th>
<th>E2 Down-Regulated Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN2</td>
<td>KAZN</td>
</tr>
<tr>
<td>FAT4</td>
<td>NAT8L</td>
</tr>
<tr>
<td>PFKFB3</td>
<td>SYNE3</td>
</tr>
<tr>
<td>DPM2,FAM102A</td>
<td>CELSR2</td>
</tr>
<tr>
<td>CBFA2T3</td>
<td>SYBU</td>
</tr>
<tr>
<td>FUT4</td>
<td>MYEOV</td>
</tr>
<tr>
<td>KBTBD8</td>
<td>FO5</td>
</tr>
<tr>
<td>TMEM201</td>
<td>SIAH2</td>
</tr>
<tr>
<td>P2RY2</td>
<td>GAB2</td>
</tr>
<tr>
<td>EPS15L1</td>
<td>ZNF703</td>
</tr>
<tr>
<td>GEMIN5</td>
<td>PITX1</td>
</tr>
<tr>
<td>AMER1</td>
<td>SCAND1</td>
</tr>
<tr>
<td>ADAT2</td>
<td>ASB13</td>
</tr>
<tr>
<td>LAMA3</td>
<td>FAM178B</td>
</tr>
<tr>
<td>TET2</td>
<td>PADI3</td>
</tr>
<tr>
<td>ITPK1</td>
<td>CALC5</td>
</tr>
<tr>
<td>FOSL2</td>
<td>CALCR</td>
</tr>
<tr>
<td>GADD45B</td>
<td>C5orf4</td>
</tr>
<tr>
<td>SOX9</td>
<td>KCNK6</td>
</tr>
<tr>
<td>CA12</td>
<td>ARID5A</td>
</tr>
<tr>
<td>HPDL</td>
<td>STX1B</td>
</tr>
<tr>
<td>WDR35</td>
<td>HSPB8</td>
</tr>
<tr>
<td>RTN4RL1</td>
<td>ARTN</td>
</tr>
<tr>
<td>C10orf2</td>
<td>PPM1K</td>
</tr>
<tr>
<td>PRSS23</td>
<td>SLC22A5</td>
</tr>
<tr>
<td>RAB3A</td>
<td>XBPI</td>
</tr>
<tr>
<td>NCOR2</td>
<td>SGK223</td>
</tr>
<tr>
<td>PLD6</td>
<td>HS3ST3B1</td>
</tr>
<tr>
<td>TBKBP1</td>
<td>RASGRP1</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>POLR1B</td>
<td>MPPED2</td>
</tr>
<tr>
<td>ONECUT2</td>
<td>RNF223</td>
</tr>
<tr>
<td>ADAMTSL5</td>
<td>PKIB</td>
</tr>
<tr>
<td>C1QTNF6</td>
<td>SEMA6A</td>
</tr>
<tr>
<td>COL12A1</td>
<td>ZNF488</td>
</tr>
<tr>
<td>CAMKK1</td>
<td>PEX11A</td>
</tr>
<tr>
<td>SLC7A5</td>
<td>TSKU</td>
</tr>
<tr>
<td>MED13L</td>
<td>ISM1</td>
</tr>
<tr>
<td>SLC9A3R1</td>
<td>STC2</td>
</tr>
<tr>
<td>TMED8</td>
<td>NAV2</td>
</tr>
<tr>
<td>FOXC1</td>
<td>MBOAT1</td>
</tr>
<tr>
<td>AMOTL1</td>
<td>TFF1</td>
</tr>
<tr>
<td>TGM2</td>
<td>PABPN1L</td>
</tr>
<tr>
<td>KDM4B</td>
<td>AGR3</td>
</tr>
<tr>
<td>SLC2A1</td>
<td>CCDC88C</td>
</tr>
<tr>
<td>NRIP1</td>
<td>SLC2A4</td>
</tr>
<tr>
<td>USP35</td>
<td>SIGLEC15</td>
</tr>
<tr>
<td>WIPF1</td>
<td>STC1</td>
</tr>
<tr>
<td>TFAP4</td>
<td>SMOX</td>
</tr>
<tr>
<td>ANKRD1</td>
<td>KCNQ4</td>
</tr>
<tr>
<td>OTUB2</td>
<td>TIFA</td>
</tr>
<tr>
<td>CARD10,MF</td>
<td>FAM86B3P</td>
</tr>
<tr>
<td>NG</td>
<td>TRIM31</td>
</tr>
<tr>
<td>ARHGEF26</td>
<td>HEG1</td>
</tr>
<tr>
<td>PPARGC1B</td>
<td>C8orf44,C8orf44-</td>
</tr>
<tr>
<td>RERG</td>
<td>SGK3,SGK3</td>
</tr>
<tr>
<td>JAK2</td>
<td></td>
</tr>
<tr>
<td>SNORD15A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SALL4</td>
</tr>
<tr>
<td></td>
<td>DIO2</td>
</tr>
<tr>
<td></td>
<td>BMF</td>
</tr>
<tr>
<td></td>
<td>FAM46B</td>
</tr>
<tr>
<td></td>
<td>EPHB3</td>
</tr>
<tr>
<td></td>
<td>CCNG2</td>
</tr>
<tr>
<td></td>
<td>EDARADD</td>
</tr>
</tbody>
</table>
**Supplementary Table 2:** GSEA analysis of genes down-regulated upon ZNF143 depletion under E2 stimulation

<table>
<thead>
<tr>
<th>Gene Set Name</th>
<th>P-Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Oestrogen Response</td>
<td>7.21e-15</td>
</tr>
<tr>
<td>Hypoxia</td>
<td>7.21e-15</td>
</tr>
<tr>
<td>Late Oestrogen Response</td>
<td>3.92e-9</td>
</tr>
<tr>
<td>MTORC1 signalling</td>
<td>3.92e-9</td>
</tr>
<tr>
<td>Fatty Acid Metabolism</td>
<td>9.99e-9</td>
</tr>
</tbody>
</table>
**Supplementary Table 3**: Genetic alterations tend towards mutual exclusivity (TCGA 2012)

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor 2</th>
<th>P-value</th>
<th>Log Odds Ratio</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCF</td>
<td>RAD21</td>
<td>0.243</td>
<td>-1.071</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>STAG2</td>
<td>0.394</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>SMC1A</td>
<td>0.538</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>RAD21</td>
<td>0.628</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>STAG2</td>
<td>0.795</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>SMC1A</td>
<td>0.858</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>CTCF</td>
<td>0.892</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG1</td>
<td>STAG2</td>
<td>0.904</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>SMC3</td>
<td>0.927</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG1</td>
<td>SMC1A</td>
<td>0.935</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG1</td>
<td>0.951</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG2</td>
<td>SMC1A</td>
<td>0.951</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG2</td>
<td>0.963</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG1</td>
<td>SMC3</td>
<td>0.967</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>SMC1A</td>
<td>0.975</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG2</td>
<td>SMC3</td>
<td>0.975</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC1A</td>
<td>SMC3</td>
<td>0.983</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>SMC3</td>
<td>0.988</td>
<td>&lt;=3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
</tbody>
</table>
**Supplementary Table 4**: Genetic alterations tend towards mutual exclusivity (TCGA Provisional Data Set)

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor 2</th>
<th>P-value</th>
<th>Log Odds Ratio</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF143</td>
<td>RAD21</td>
<td>0.222</td>
<td>-1.131</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>RAD21</td>
<td>0.443</td>
<td>-0.210</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>SMC3</td>
<td>0.545</td>
<td>-0.248</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>SMC3</td>
<td>0.691</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG1</td>
<td>SMC3</td>
<td>0.737</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG2</td>
<td>0.781</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
</tbody>
</table>
**Supplementary Table 5:** Fold Change mRNA Expression upon Genetic Alteration of Chromatin Looping Factors (TCGA Provisional Data Set)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Homozygous Deletion</th>
<th>Heterozygous Loss</th>
<th>Gain</th>
<th>Amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF143</td>
<td>FC = 0.69 P-val = 0.002</td>
<td>FC = 0.91 P-val &lt; 0.0001</td>
<td>FC = 1.2 P-val &lt; 0.0001</td>
<td>FC = 1.2 P-val = NS</td>
</tr>
<tr>
<td>CTCF</td>
<td>FC = 0.69 P-val = 0.004</td>
<td>FC = 0.91 P-val &lt; 0.0001</td>
<td>FC = 1.2 P-val = 0.0004</td>
<td>FC = 1.2 P-val = 0.0038</td>
</tr>
<tr>
<td>Rad21</td>
<td>NA</td>
<td>FC = 0.73 P-val = 0.002</td>
<td>FC = 1.6 P-val &lt; 0.0001</td>
<td>FC = 2.7 P-val &lt; 0.0001</td>
</tr>
<tr>
<td>STAG1</td>
<td>NA</td>
<td>FC = 0.82 P-val = 0.0001</td>
<td>FC = 1.1 P-val = 0.0004</td>
<td>FC = 1.4 P-val = NS</td>
</tr>
<tr>
<td>STAG2</td>
<td>FC = 1.3 P-val = NS</td>
<td>FC = 1.0 P-val = NS</td>
<td>FC = 1.0 P-val = NS</td>
<td>FC = 1.4 P-val = 0.02</td>
</tr>
<tr>
<td>SMC1A</td>
<td>FC = 0.62 P-val = 0.01</td>
<td>FC = 1.1 P-val = 0.015</td>
<td>FC = 1.2 P-val &lt; 0.0001</td>
<td>FC = 2.4 P-val &lt; 0.0001</td>
</tr>
<tr>
<td>SMC3</td>
<td>FC = 0.44 P-val &lt; 0.0001</td>
<td>FC = 0.8 P-val &lt; 0.0001</td>
<td>FC = 1.2 P-val &lt; 0.0001</td>
<td>FC = 0.7 P-val = NA</td>
</tr>
</tbody>
</table>
**Supplementary Table 6:** Genetic alterations tend towards mutual exclusivity in lung squamous cell carcinoma (TCGA, Provisional)

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor 2</th>
<th>P-value</th>
<th>Log Odds Ratio</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMC3</td>
<td>STAG1</td>
<td>0.268</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG1</td>
<td>STAG2</td>
<td>0.325</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>STAG1</td>
<td>0.691</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>SMC1A</td>
<td>0.691</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC1A</td>
<td>STAG2</td>
<td>0.729</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>SMC3</td>
<td>0.751</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>RAD21</td>
<td>0.783</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>STAG2</td>
<td>0.783</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC3</td>
<td>STAG2</td>
<td>0.783</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG2</td>
<td>0.812</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>RAD21</td>
<td>0.923</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>CTCF</td>
<td>0.934</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>STAG2</td>
<td>0.934</td>
<td>&lt;3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG1</td>
<td>0.733</td>
<td>-0.014</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>STAG1</td>
<td>0.664</td>
<td>-0.203</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
</tbody>
</table>
**Supplementary Table 7**: Genetic alterations tend towards mutual exclusivity in bladder urothelial carcinoma (TCGA, 2014)

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor 2</th>
<th>P-value</th>
<th>Log Odds Ratio</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD21</td>
<td>STAG2</td>
<td>0.297</td>
<td>-0.998</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>RAD21</td>
<td>0.460</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>STAG2</td>
<td>0.504</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC1A</td>
<td>STAG2</td>
<td>0.580</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG1</td>
<td>STAG2</td>
<td>0.665</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>SMC3</td>
<td>0.736</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC3</td>
<td>STAG2</td>
<td>0.763</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>RAD21</td>
<td>0.858</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG2</td>
<td>0.874</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>STAG1</td>
<td>0.886</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC1A</td>
<td>STAG1</td>
<td>0.908</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>SMC3</td>
<td>0.923</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC1A</td>
<td>SMC3</td>
<td>0.938</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC3</td>
<td>STAG1</td>
<td>0.953</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>CTCF</td>
<td>0.961</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>SMC1A</td>
<td>0.969</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG1</td>
<td>0.976</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>SMC3</td>
<td>0.984</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
</tbody>
</table>
### Supplementary Table 8: Genetic alterations tend towards mutual exclusivity in liver hepatocellular carcinoma (TCGA Provisional Data Set)

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor 2</th>
<th>P-value</th>
<th>Log Odds Ratio</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD21</td>
<td>STAG1</td>
<td>0.319</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>RAD21</td>
<td>SMC3</td>
<td>0.506</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>RAD21</td>
<td>0.569</td>
<td>-0.432</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>SMC1A</td>
<td>0.830</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>STAG1</td>
<td>0.830</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC1A</td>
<td>STAG1</td>
<td>0.876</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>CTCF</td>
<td>0.895</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>SMC3</td>
<td>0.895</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG1</td>
<td>0.924</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC1A</td>
<td>SMC3</td>
<td>0.924</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC3</td>
<td>STAG1</td>
<td>0.924</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>CTCF</td>
<td>STAG2</td>
<td>0.929</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMCA1</td>
<td>STAG2</td>
<td>0.949</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>STAG1</td>
<td>STAG2</td>
<td>0.949</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>SMC3</td>
<td>0.954</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>ZNF143</td>
<td>STAG2</td>
<td>0.969</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
<tr>
<td>SMC3</td>
<td>STAG2</td>
<td>0.969</td>
<td>&lt;-3</td>
<td>Tendency towards mutual exclusivity</td>
</tr>
</tbody>
</table>
### Supplementary Table 9: Primers used in this study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actin</td>
<td>5’ GGACTTCGAGCAAGAGATGG 3’</td>
<td>5’ AGCAGCTGTCTTGGCGTACAG 3’</td>
</tr>
<tr>
<td>ZNF143</td>
<td>5’ CGCAGTCTGACACCATCTTG 3’</td>
<td>5’ CCAATCATTCCAGTACCTGCT 3’</td>
</tr>
<tr>
<td>RARA</td>
<td>5’ AGGCTACCACATGAGG1GCTCA 3’</td>
<td>5’ CGGTCACCTTTGATGATG 3’</td>
</tr>
<tr>
<td>XBP1</td>
<td>5’ CAGGCCCGAGGAGAAG 3’</td>
<td>5’ TGTTCCAGCTCATTCCG 3’</td>
</tr>
<tr>
<td>CA12</td>
<td>5’ GTGTTGTCCATTTGGCGTIC 3’</td>
<td>5’ CAAGTGCTTCCTGGATGTG 3’</td>
</tr>
<tr>
<td>ERBB3</td>
<td>5’ TCTTCCTGAGTGGATTCG 3’</td>
<td>5’ CATCTCGTCCCTCAGATG 3’</td>
</tr>
<tr>
<td>CCND1?</td>
<td>5’ TTGTTCAAGCAGCGAGTCCC 3’</td>
<td>5’ CTGTTCCCTCGAGACCGAAG 3’</td>
</tr>
</tbody>
</table>
Supplementary Figure 1: ZNF143 ChIP-sequencing replicates show a high degree of reproducibility

Scatterplots depicting read count for ZNF143 binding in MCF-7 breast cancer cells treated with vehicle or E2 (10µM) for 45 minutes in two separate experiments. Plot was generated with SeqMonk from ZNF143 ChIP-sequencing data. R, Pearson’s correlations, were calculated from all data points.
Supplementary Figure 2: ChromHMM annotation of genomic states in E2-treated MCF-7 cells

Emission states of chromatin structure from ChromHMM. Darker blues correspond to higher percentage of representation in a specific state (A). ChromHMM transition parameters (B). Overlap enrichment (C).
Supplementary Figure 3:

Venn diagram illustrating the extent of overlap between ZNF143 and ESR1 binding sites.

Supplementary Figure 3: ZNF143 and ESR1 share a subset of binding sites.
Supplementary Figure 4:
Depletion of ZNF143 mRNA does not affect ESR1 or FOXA1 expression under vehicle or E2-stimulated conditions

Gene expression changes in cells transfected with siZNF143 or scrambled siren under vehicle and E2-stimulated conditions (RPKM: reads per kilobase per million). (*p<0.05; ** p<0.01; *** p<0.001; NS = Not Significant)
Supplementary Figure 5: Rad21 expression is relevant to clinical outcome of luminal breast cancer patients

Kaplan-Meier plots derived from the METABRIC data set (Curtis et al. 2012) evaluating overall survival in ESR1-positive breast cancer patients (n=1486), Luminal A patients (n=825) and Luminal B (n=668) patients, stratified by RAD21 expression. Data were obtained from the Kaplan-Meier plotter breast cancer survival analysis database (Györffy 2010). Hazard ratios (HR) and logrank P-values are displayed.
Supplementary Figure 6

Supplementary Figure 3: Chromatin Looping Factors are Genetically Altered in Many Cancers

Overall frequency of genetic alterations in the chromatin looping factors (ZNF143, CTCF and the cohesion subunits) in datasets available through the CBioPortal database (http://www.cbioportal.org).
5 References


