DECIPHERING THE ENHANCEOSOME-MEDIATED

TRANSCRIPTIONAL ACTIVATION AND

CELL PROLIFERATION IN BREAST CANCER

KONG SAY LI
(M. Sc., NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF MEDICINE
YONG LOO LIN SCHOOL OF MEDICINE
NATIONAL UNIVERSITY OF SINGAPORE

2012

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

SUMMARY

Estrogen receptor α (ERα) is a ligand-inducible hormone nuclear receptor that has important physiology and pathology roles in various human systems. Our work focused on understanding the combinatorial complexity of ER regulatory control on a genomic scale. In our earlier studies (Joseph *et al*, 2010; Lin *et al*, 2007) we found that FOXA1 and GATA3 motifs were commonly enriched around ERα binding sites. We then pursued the question of binding site selection and found that though sequence was the most important determinant, the presence of FOXA1 binding and DNA Pol II binding were important secondary characteristics that are associated with ER binding site selection. Numerous microarray studies have documented the co-expression of ERα, FOXA1 and GATA3 in primary breast tumors (Badve *et al*, 2007; Wilson and Giguere, 2008). These evidences suggest that potentially these three transcription factors (TFs) function conjointly to contribute to the breast cancer phenotype. However, the nature of their coordinated interaction at the genome level or the biological consequences of their co-expression remains poorly understood.

To extend these observations, we mapped the genome-wide binding profiles of ERα, FOXA1, and GATA3. We observed that these three TFs co-localized in a coordinated fashion upon estrogen stimulation. Moreover, we found that the ERα+FOXA1+GATA3 conjoint sites were associated with highest p300 coactivator recruitment, RNA Pol II occupancy, and chromatin opening. Such results indicate that these three TFs form a functional enhanceosome and cooperatively modulate the transcriptional networks previously ascribed to ERα alone. In addition, such enhanceosome binding sites appear to regulate the genes driving core ERα function.

Though ERα is known to mediate the proliferative effects of estrogen (E$_2$) in breast cancer cells, the exogenous introduction of ERα into an ERα-negative line displayed inhibited growth (Garcia *et al*, 1992). We posited that the composition of enhanceosome is required to establish transcriptional regulatory cassettes favoring growth enhancement.

To test this, we stably transfected the MDA-MB-231 cells with individual ERα, FOXA1, GATA3 or in combinations. We demonstrated that the co-expression of ERα+FOXA1+GATA3 resulted in marked induction of estrogen-stimulated growth. This cellular reprogramming was recapitulated in another ERα-negative breast cancer cell line, BT-549 and observed similar E$_2$-responsive growth induction in the ERα+FOXA1+GATA3-expressing cells. This suggests that only with the full activation of conjoint binding sites by the three TFs will the proliferative phenotype associated with ligand induced ERα be manifest.

To assess the nature of this transcriptional reprogramming, we compared the expression profiles of the reprogrammed MDA-MB-231 and BT-549 cells with the ERα-positive breast cancer cell line, MCF-7. Strikingly, we found that the expression profiles of ERα+FOXA1+GATA3 expressing MDA-MB-231 and BT-549 cells display positive correlation with the E$_2$ induced expression profile of MCF-7. In contrast, negative correlation was found in the MDA-MB-231 and BT-549 transfected with ERα only. Furthermore, we observed that the enhanceosome component is competent to partially reprogramme the basal cells to resemble the luminal cells.

Taken together, we have uncovered the genomics impact as well as the functional importance of an enhanceosome comprising ERα, FOXA1 and GATA3 in the estrogen

responsiveness of ERα positive breast cancer cells. This enhanceosome exerts significant combinatorial control of the transcriptional network regulating growth and proliferation of ERα positive breast cancer cells.

LIST OF FIGURES

INTRODUCTION

MATERIALS AND METHODS

RESULTS

DISCUSSION

LIST OF TABLES

MATERIALS AND METHODS

RESULTS

LIST OF ABBRECIATIONS

AP1	activator protein 1
AR	androgen receptor
ChIP-seq	chromatin immunoprecipitation-sequencing
ChIP	chromatin immunoprecipitation
CDFBS	charcoal dextran-treated fetal bovine serum
ChIA-PET	Chromatin Interaction Analysis with Paired-End-Tag
DBA	differential binding analysis
DBD	DNA binding domain
E_2	estrogen
ERα	estrogen receptor α
ERβ	estrogen receptor β
ERE	estrogen responsive elements
FAIRE	formaldehyde assisted isolation of regulatory elements
FBS	fetal bovine serum
FOXA1	forkhead box A1
GATA3	GATA binding protein 3
GO	Gene Ontology
HAT	histone acetyltransferase
HNF3A	hepatocyte nuclear factor 3α
IHC	immunohistochemistry
IPA	Ingenuity Pathway Analysis
iPS	induced pluripotent stem
KO	knockout

LBD	ligand binding domain
LB	lysogeny broth
LDL	low-density lipoprotein
NF-κB	nuclear factor-kappa B
N.S.	not significant
PCA	principal component analysis
PBS	phosphate buffered saline
PPAR-γ	Peroxisome-Proliferator-Activated Receptor-γ
PR	progesterone receptor
qPCR	quantitative Polymerase Chain Reaction
RNA Pol II	RNA polymerase II
SERM	selective estrogen receptor modulators
TF	transcription factor
TIC	tumor initiating cell
TES	transcription end site
TSS	transcription start site

CHAPTER 1: INTRODUCTION

1.1 The physiological and pathological roles of estrogen

Estrogen has widespread biological functions in numerous human tissues and diseases. It stimulates the growth of reproductive systems, maintains the bone density by mediating the function of osteoclasts and exerts cardiovascular protection effects through its vasodilation properties in the vascular smooth muscle (Grodstein *et al*, 2000) (Figure 1). The estrogen has also encompassed the neuroprotective roles in the brain tissues by inducing the synaptic and dendritic activation (Naftolin *et al*, 1999). Moreover, estrogen can enhance the lipoprotein receptors leading to the reduction of low-density lipoprotein (LDL) cholesterol in the serum (Paganini-Hill *et al*, 1996).

Though estrogen exerts several beneficial roles in various human tissues, increased exposure to estrogen possessed harmful effects in raising the breast and endometrial cancer risks by stimulating the growth of tumor cells (Bernstein, 2002).

The complex mechanisms by which these effects are mediated remain incompletely understood. The major question is how a single nuclear hormone receptor can provide such disparate phenotypic and molecular outputs dependent on the tissue or origin. Hence, the molecular actions of estrogen in various tissues are the subject of enormous research efforts.
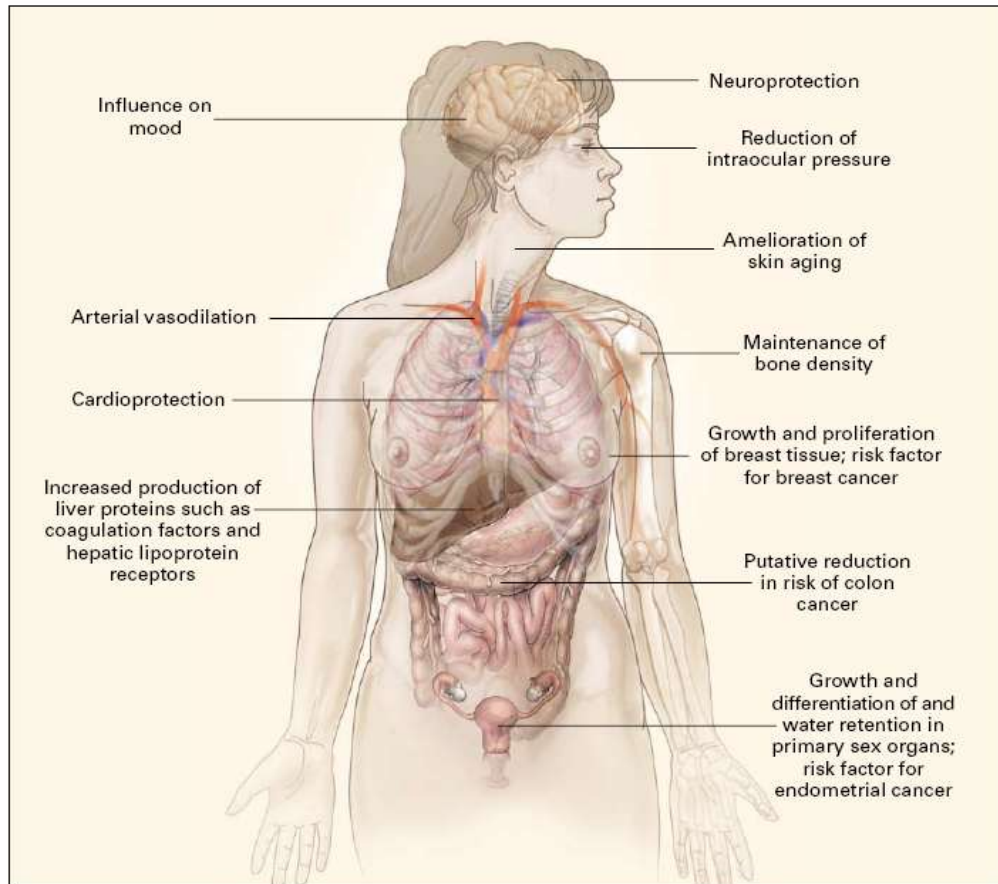
Figure 1. The physiological and pathological roles of estrogen in various human organ systems (Gruber *et al*, 2002).

1.2 The structure of estrogen receptor

The identification of estrogen receptor (ER) by Jensen and Jacobsen in 1960 has shifted the paradigm of steroid hormone action from the enzymatic mode to the current model of estrogen regulation by a receptor protein (Jensen and Jacobson, 1960). The pleiotropic effect of estrogen is largely mediated by its receptor, ER, which is a member of the nuclear receptor superfamily.

The ERα gene is located at 6q25.1 that extends more than 140kb. It contains eight exons and encodes a protein of 595 amino acids with a predicted molecular weight of 66,182 Dalton (Figure 2). The human ER has the typical structure resembles to the other class I members of ligand-inducible steroid receptor superfamily. It consists of six functional domains designated as A – F: the amino-terminal A/B domain with the hormone-independent activation function (AF-1); the middle C domain contains the DNA binding domain (DBD) consisting of two zinc finger motifs that are responsible for ER binding to estrogen responsive elements (EREs) and dimerization; the D domain – the hinge region is implicated in co-regulatory protein binding; the carboxy-terminal domains E and F contain the ligand binding domain (LBD) that are implicated in modulating the agonist activity. ER is a predominantly nuclear protein regardless of whether or not it is complexed with ligand.
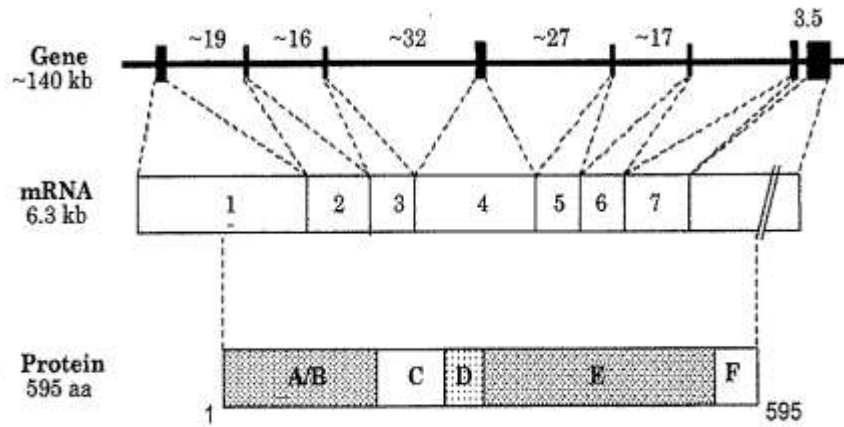
Figure 2. The structure of ESR1 gene at 6q25.1 spans more than 140kb and contains 8 exons. The ERα protein is composed of 595 amino acids and organized into A to F fuctional domains (Shao and Brown, 2004).

1.3 The estrogen receptor subtypes

ER is classified into ERα and ERβ subtypes (Witkowska *et al*, 1997) that composed of highly homologous DNA-binding domain but varies in the ligand-binding domain of which only 58 percent of the amino acid sequence is shared (Figure 3).

The distribution of ERα and ERβ also varies in different tissues. ERβ presents mostly in brain, kidney, intestinal mucosa and prostate gland. In contrast, the classical estrogen target tissues such as breast, endometrium and ovarian stroma contain mostly ERα (Figure 4).

The most striking phenotypes in the female ERα knockout (ERαKO) mice include estrogen insensitivity in the reproductive tract, lack of pubertal mammary gland development and excess adipose tissue, whereas in the male, testicular degeneration and epididymal dysfunction are major phenotypes (Couse JF. and S., 1999). These phenotypes combined with severe deficits in sexual behavior result in complete infertility in both sexes of the ERαKO mice. In contrast, ERβKO males are fertile and show no obvious phenotypes while the female ERβKO mice exhibit inefficient ovarian function and subfertility. Interestingly, combine ERα and ERβ knockout mice (ERαβKO) exhibits phenotypes that mostly resemble those of the ERαKO, suggesting that ERα plays a more predominant role in development.

The diverse composition and responsiveness of ERα and ERβ to different ligands has initiated the search for tissue-selective estrogen receptor modulators (SERMs) with the hopes to establish the therapeutic intervention that enhances the beneficial effects of

estrogen in the target tissues and avoid the unwanted harmful side effects, which will be extremely useful in the treatment of menopausal symptoms, osteoporosis, cardiovascular disease and breast cancer in women.

1.4 The mediation of estrogen by estrogen receptors

There are two models of ligand-dependent activation of ER exist to date: the 'classical' activation induced by agonists which results in direct interaction of the ER with DNA and subsequent transcription activation; and the 'non-classical' activation induced by the agonists which cause the interaction of ER with other proteins which in turn, bind to DNA and modulate transcription.

The classical estrogen signaling pathway reveals that upon the estrogenic stimulation, ligand binding leads to conformational changes of the receptor which causes the dimerization of the receptors. The activated ER will diffuse into the nucleus, bind to its DNA recognition sequences known as estrogen response elements (EREs) and initiate the transcription of its target genes (Figure 5). The 13-bp inverted repeat sequences 5'GGTCAnnnTGACC'3 (n denotes a random nucleotide) are defined as the binding motif of ER that is present in the regulatory regions of the estrogen target genes (Walker *et al*, 1984). Upon binding to an ERE, the ligand-activated ER complex will interact with the co-factors and other histone remodeling enzymes to stimulate or inhibit transcription machinery.
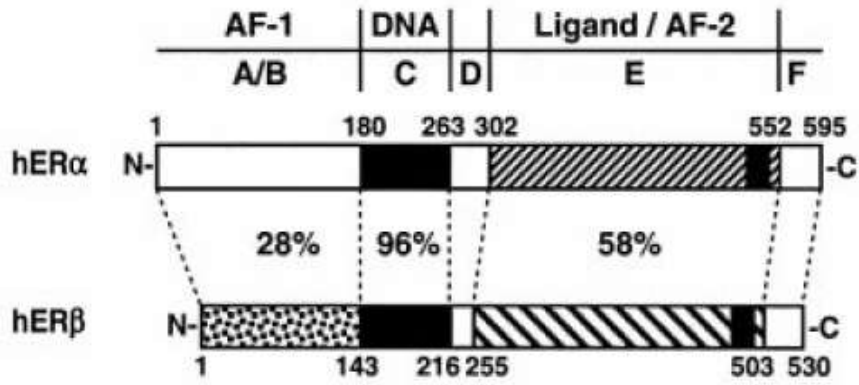
Figure 3. The schematic diagram of human ERα and ERβ (Cheung *et al*, 2003).
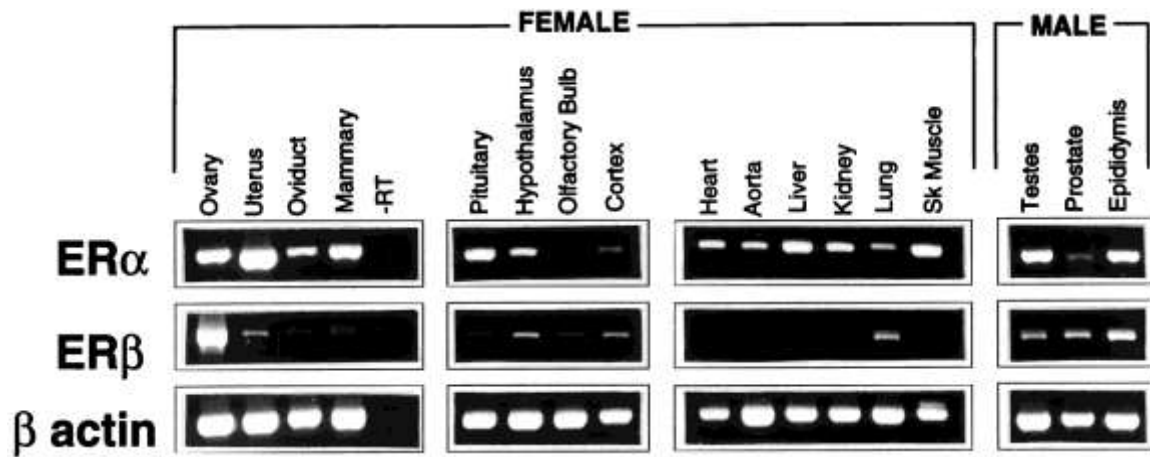
Figure 4. The distribution of ERα and ERβ mRNA in various tissues of the mouse. The ERα mRNA is broadly distributed in many tissues whereas the ERβ transcripts are primarily expressed in the ovary, hypothalamus, lung and male reproductive tract (Couse JF. *et al*, 1999).
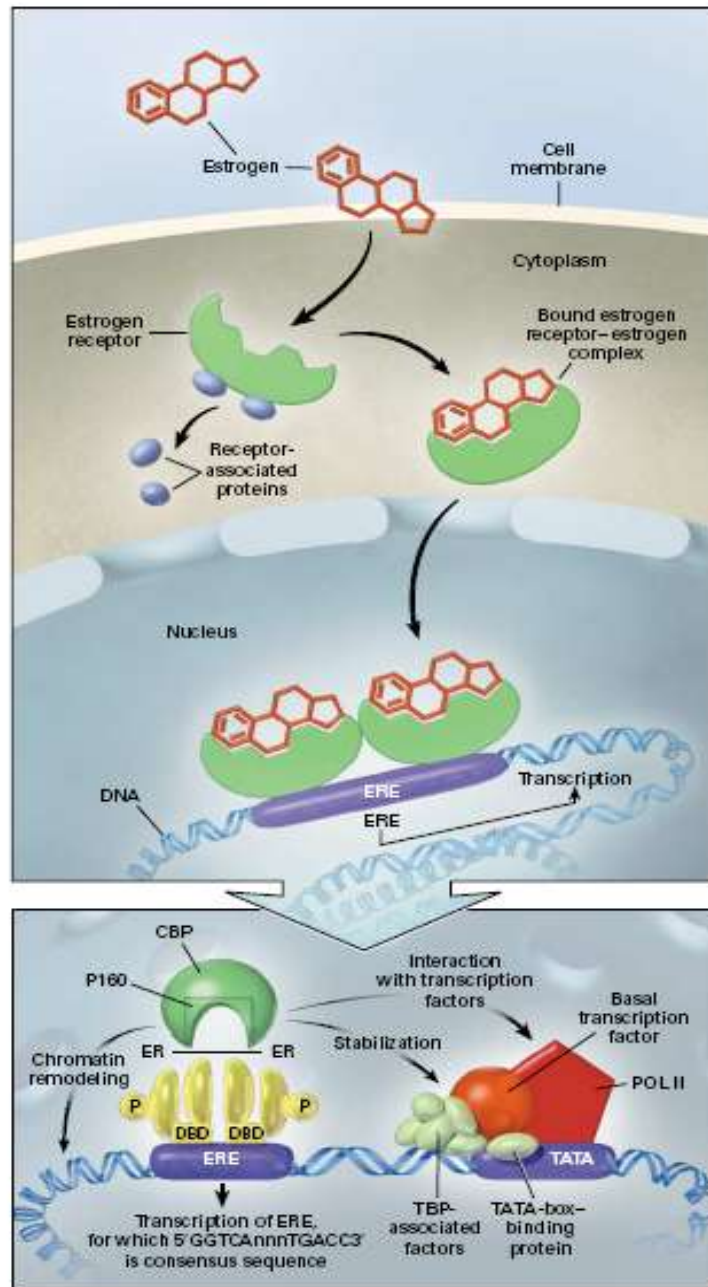
Figure 5. The illustration of classical estrogen signaling pathway mediated by ER. The ligand-activated ER will diffuse into the nucleus, bind to its DNA recognition sequences known as estrogen response elements (EREs) and initiate the transcription of its target genes (Gruber *et al*, 2002).

According to the 'non-classical' model of ER activation, ligand binding leads to the interaction of ER with other transcription factors (TFs) such as AP-1, SP1 or NF-κB (Figure 6). These complexes bind to DNA through direct interaction of the ER-associated TFs, thus influencing the transcription of genes which do not contain EREs but rather have recognition sites specific for AP-1, Sp1 or NF-κB (Webb P. *et al*, 1999). This 'non-classical' model also known as 'transcription factor crosstalk' occurs without the direct ER binding to the DNA.

Another "non-genomic" pathway also exist, whereby the estrogen activates kinase cascades such as MAPK and/or PI3K that activate other transcription factors by phosphorylation and induce gene expression (Figure 6). The ability of estrogen to activate non-genomic kinases such as ERK and AKT may depend on the expression level of peptide growth factor receptors and the signaling kinases in the cells. "Non-genomic" estrogen signaling is transduced by a membrane-localized pool of ER, which can rapidly activate both ERK and AKT in response to estrogen. However, this estrogen-induced cytoplasmic signaling is restricted in cells that express high levels of growth factor receptors (DeNardo DG *et al*, 2007).
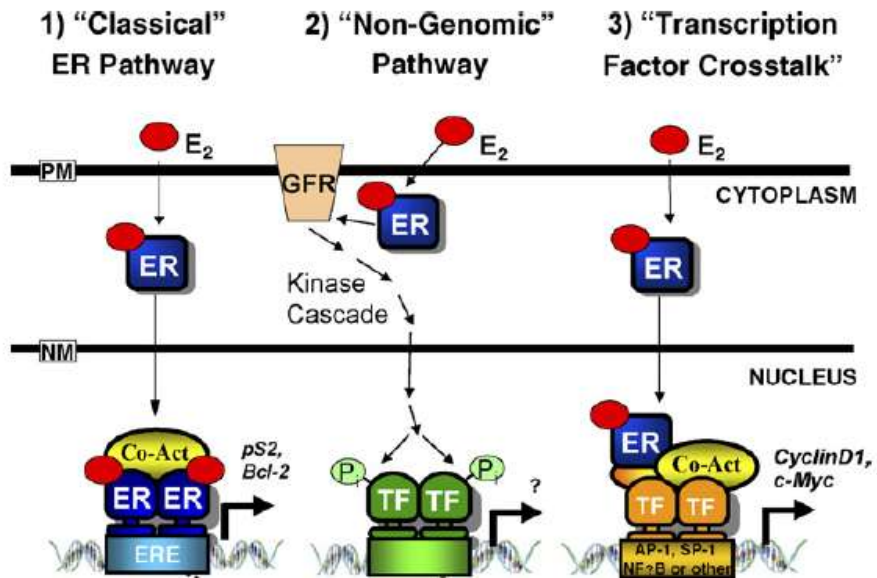
Figure 6. Illustrations of models representing the various modes through which ER can modulate transcription machinery. In the "classical" pathway, estrogen binds to ER, which in turn translocates to the nucleus, binds DNA at the ERE and activates the expression of ERE-dependent genes. In the "non-genomic" pathway, estrogen activates kinase cascades such as MAPK and/or PI3K that activate TFs by phosphorylation and induce gene expression. In the "transcription factor crosstalk" pathway, ER is activated by estrogen and enhances gene transcription through indirect DNA-binding with other TFs such as AP-1, Sp-1, NF-κB, and others (DeNardo DG *et al*, 2007).

1.5 The function of estrogen receptor as the DNA-binding transcription factor in gene regulation

The activity of transcription factors (TFs) was one of the first proposed mechanisms of gene regulation many decades ago. Since the exciting discoveries of various chromatin marks and associated remodeling factors for the last 15 years, epigenetic modifications have largely dominated the general view of how eukaryotic genomes are transcriptionally regulated. However, the recent ground-breaking reprogramming studies to generate the induced pluripotent stem (iPS) cells from the fibroblast with four TFs have reminded the field about the power and supremacy of TFs. Most importantly, this have reinforced the notion that both the TFs regulation and epigenetic modifications are important contributors to eukaryotic gene regulation (Takahashi K and S., 2006; Takahashi K *et al*, 2007).

The complex human body consists of many systems that are specified by their unique transcriptional programs. A finely tuned modulation of transcription activity requires the coordination of numerous regulatory events and mechanisms involving the DNA-binding TFs, coactivators, corepressors and basal RNA polymerase machinery (Figure 7). Nuclear hormone receptors regulate various biologically important processes in the development and homeostasis through their bimodal function as activators and repressors of gene transcription. This precise cell- and time-specific regulation is crucial for the normal development of all organisms.
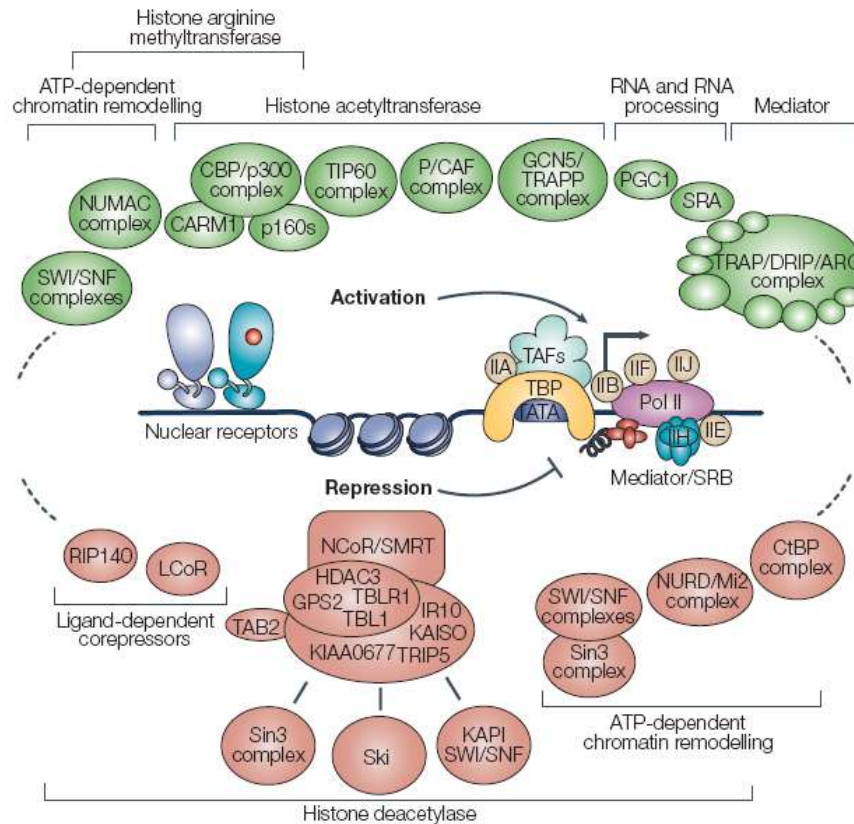
Figure 7. The regulation of gene transcription orchestrated by coactivators, corepressors and basal RNA polymerase machinery. Coactivators complexes include factors that contain ATP-dependent chromatin remodeling activity, histone acetyltransferase, histone arginine methyltransferase that are involved in RNA processing. Conversely, corepressors contains histone deacetylase (Perissi and Rosenfeld, 2005).

The key question in transcription regulation field is deciphering how an organism can achieve such diversity, while maintaining cell specificity and responding dynamically to its environment. One possible explanation is to employ a restricted repertoire of activators as to minimize the complexity required to link the related signaling pathways and orchestrate diverse regulatory cues.

The current model of ER function as a TF implies that ER modulates the transcription machinery via the recruitment of a variety of coactivators or corepressors and chromatin remodeling enzymes to the promoters and enhancers regions (Shang *et al*, 2000). Transcriptional activation involves alterations in chromatin structures mediated by the ATP-dependent chromatin-remodeling enzymes in conjunction with factors that contain histone acetyltransferase (HAT) property that leads to the decondensation of chromatin structure, which is a prerequisite for gene transcription. A large number of co-factors are implicated in these chromatin remodeling processes such as p300 and CBP. The specific and ordered recruitment of multiple protein complexes to the chromatin will then provides the chromatin with the plasticity required for the transcription initiation (Cosma, 2002). However, our understanding in this multifaceted gene regulation is still fragmentary as the range of possible combinations of different players and their remoteness from the transcription start sites (TSS) of regulated genes suggest multiple regulatory mechanisms (Farnham, 2009).

ERα is also known to commonly induce long distance chromatin interactions between ERα binding sites and TSS through chromatin looping (Fullwood *et al*, 2009), suggesting

that higher dimensional structural order beyond the simple receptor-recognition motif interaction is essential in explaining ERα directed transcriptional regulation.

The enhanceosome assembly is dependent on the arrangement of precise component of bound activators, which together generate a network of protein-protein and protein-DNA interactions unique to a given enhancer (Carey M., 1998). As illustrated in Figure 8, the enhanceosome displays an assembly of activator-activator interactions on the naked DNA or chromatin templates. Reciprocity is essential in this process whereby the enhanceosome recruits the RNA Pol II machinery and the machinery reciprocally facilitates assembly of the enhanceosome. This provides an additional specificity and energy required to drive the concerted formation of the final "transcriptome".

The finding that transcription factors (TFs) cluster at juxtaposed binding sites in the genome to form enhanceosomes further suggests that the totality of gene regulation by any transcription factor will be dependent on a complex interaction between specific ERα binding, local configuration of co-occupying TFs, protein co-factors, chromatin conditions, and three dimensional interactions.
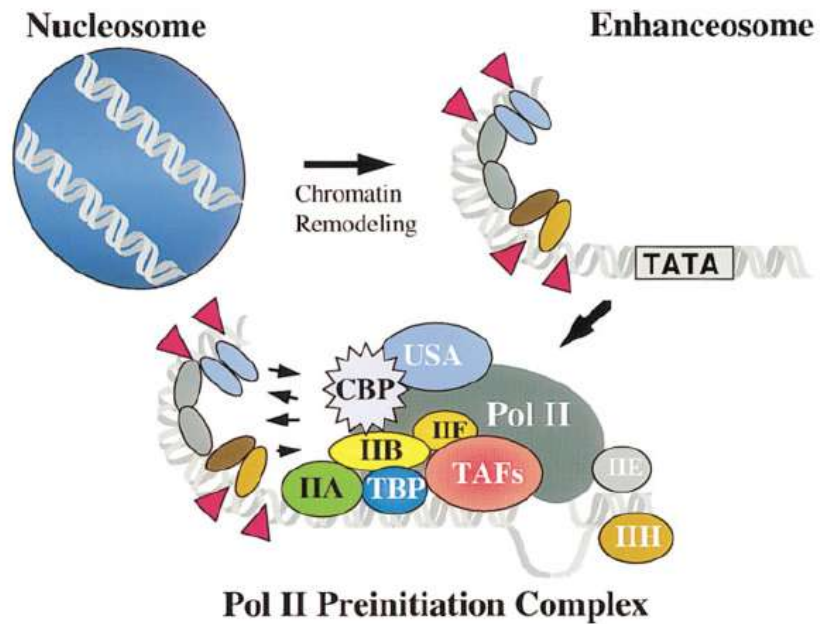
Figure 8. The assembly of enhanceosome components to Pol II preinitiation complex in a multistage process. Initially, the activators engaged to the chromatin in a cooperative manner and formed a stable enhanceosome. This is followed by the recruitment of Pol II and its ancillary factors to DNA and resulted in synergistic transcription that is driven by the reciprocity interactions indicated by the reverse and forward arrows (Carey M., 1998).

1.6 The co-expression of estrogen receptor α, FOXA1 and GATA3 in breast cancer

Breast cancer is a molecularly heterogeneous disease whereby the tumors could be classified into different subtypes of distinct molecular portraits with prognostic significance. This heterogeneity has spawned an era of molecular assays striving to classify the tumor subtypes to predict the disease outcome and provide useful guidelines to the future of targeted personalized treatment strategies.

Since the innovation of microarray technology a decade ago, systematic characterization of gene expression profiles in human breast tumors have provided better understanding in the molecular taxonomy of breast cancers (Figure 9). In particular, these microarray studies have identified distinctive molecular subtypes – luminal A/ B, ERBB2-associated, basal-like, claudin-low subtype and normal-like breast tumors based on gene expression profiling patterns as well as copy number alterations (Perou *et al*, 2000; Prat *et al*, 2010; van't Veer *et al*, 2002).

Patients with ERα-positive breast cancers usually have better prognosis which is partly due to their response to endocrine therapy. Both luminal subtypes A and B are ERα positive with a more favorable outcome in luminal A tumor as compared to luminal B tumors with higher proliferation rate and frequent DNA amplification. The ERBB2 subtype is associated with expression of genes co-amplified with ERBB2 (encoding HER2). The aggressive basal-like tumors are ERα-, progesterone receptor- (PR) and ERBB2-negative, hence also known as triple-negative tumors are associated with poor outcome. The claudin-low tumors lack common epithelial cell features and is enriched

for tumor initiating cell (TIC) features. The normal-like subtype shares expression patterns with normal breast tissue. Due to the unique biology and prognostic features in various breast tumor subtypes, it has become appreciated that breast cancer is not one disease, but in fact represents multiple disease types whereby each of which requires a unique treatment modulation.

Since ERα has been implicated in the etiology of breast cancer, it is a major prognostic marker and therapeutic target in breast cancer. Interestingly, the co-expression of ERα, FOXA1 and GATA3 has been reported in different cohorts of luminal breast cancer tumors (Lacroix M and G., 2004; Oh DS *et al*, 2006; Sorlie T *et al*, 2003). Therefore, it is essential to gain a better understanding in the transcription networks mediated by ERα and its co-expressed genes in breast cancer.
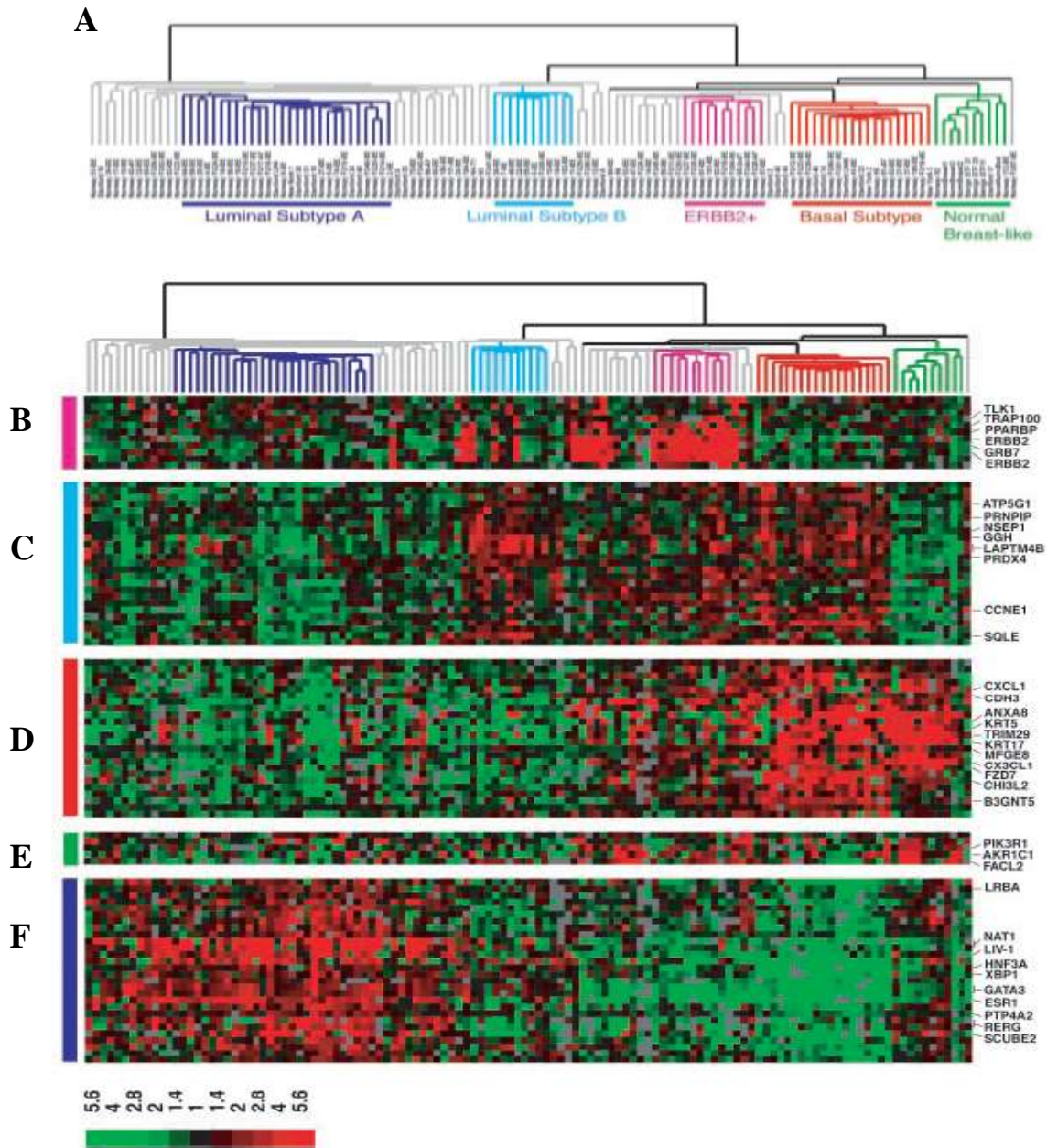
Figure 9. The hierarchical clustering of 115 tumor tissues and 7 non-malignant tissues based on the distinctive molecular signatures. (A) The dendrogram showing the clustering of the breast tumors into five subgroups. (*B*) Gene cluster associated with *ERBB2* amplification. (C) Gene cluster associated with luminal subtype B. (*D*) Gene cluster associated with the basal subtype. (*E*) Gene cluster relevant for the normal breast-like group. (F) Gene cluster associated with luminal subtype A tumors with co-expression of ERα (ESR1), FOXA1 (HNF3A) and GATA3. Scale bar represents fold change for any given gene relative to the median level of expression across all samples (Sorlie T *et al*, 2003).

19

Patients with ERα positive tumors have a longer disease-free interval and overall survival than patients with tumors lacking ERα expression. According to international treatment guidelines for early breast cancer, patients with ERα and/ or PR expression should receive an adjuvant anti-hormonal endocrine therapy. However, the association between ERα expression and hormonal responsiveness is far from perfect, since approximately 30% of ER-positive tumors do not respond to hormonal treatment and 5 to 15% of ER-negative tumors curiously respond to endocrine therapy (Jordan VC *et al*, 1988). Hence, the expression of FOXA1 and GATA3 as the useful prediction marker for patient response to hormonal treatment has been proposed by several studies (Badve *et al*, 2007; Mehra *et al*, 2005; Oh DS *et al*, 2006).

Among the ER-positive tumors, expression of FOXA1 mRNA was noted in tumors that showed favorable outcome (Oh DS *et al*, 2006). When the FOXA1 protein expression in breast cancers was analyzed by immunohistochemistry (IHC), FOXA1[high] (score greater than 3 by IHC) was associated with ERα positivity, GATA3 positivity and luminal A subtype (Badve *et al*, 2007). Most importantly, improved event-free survival was seen in these patients compared with patients in the FOXA1[low] group even at 20 years.

Another clinical study involving over 3500 primary invasive ductal carcinomas demonstrated positive FOXA1 staining in ~86% of all specimens, and expression was positively correlated with favorable prognosis. In consistent with the finding by Badve *et al*, low FOXA1 was correlated with established markers of poor prognosis including high grade, increased tumor size, basal tumors and nodal metastasis (Mehta RJ *et al*, 2011).

The meta-analysis of breast cancer microarray datasets involving 305 breast tumor samples from four breast cancer cohorts (Huang E *et al*, 2003; Lonning PE *et al*, 2001; Sotiriou C *et al*, 2003; van de Vijver MJ *et al*, 2002) revealed that low GATA3 expression was strongly associated with higher histological grade, positive lymph nodes, larger tumor size, ER and PR-negative status, HER2 overexpression and greater risk for recurrence or metastasis (Mehra *et al*, 2005). This observation recapitulated the similar finding in the association of poor prognosis with low FOXA1 level (Figure 10).

Cumulatively, these findings suggested that better delineation on the roles of ERα, GATA3 and FOXA1 in breast cancer will be useful for accurate diagnosis, predicting endocrine responsiveness, form the basis for novel therapeutic strategies and assess the patient's outcome.

1.7 The differential estrogen receptor α binding is associated with clinical outcome in breast cancer

Though ERα is the major determining factor in dictating the cellular growth and endocrine response of breast cancer cells, the comprehensive understanding of ERα function remains fragmentary. The recently published work by Ross-Innes *et al* (Ross-Innes C.S. *et al*, 2012) has mapped the ER binding events in primary breast cancers from patients with different clinical outcomes. Using the differential binding analysis (DBA) and principal component analysis (PCA) approaches, they reported differential ERα binding events in the primary breast tumors with good prognosis, poor prognosis and

distant metastases (Figure 11A). The genes that are associated with these distinguishable ER binding profiles are capable of discriminating the clinical outcome (Figure 11B), suggesting that the ER binding to cis-regulatory elements is functionally and biologically relevant. This sheds lights on the importance and relevance of TF binding in modulating the disease state.
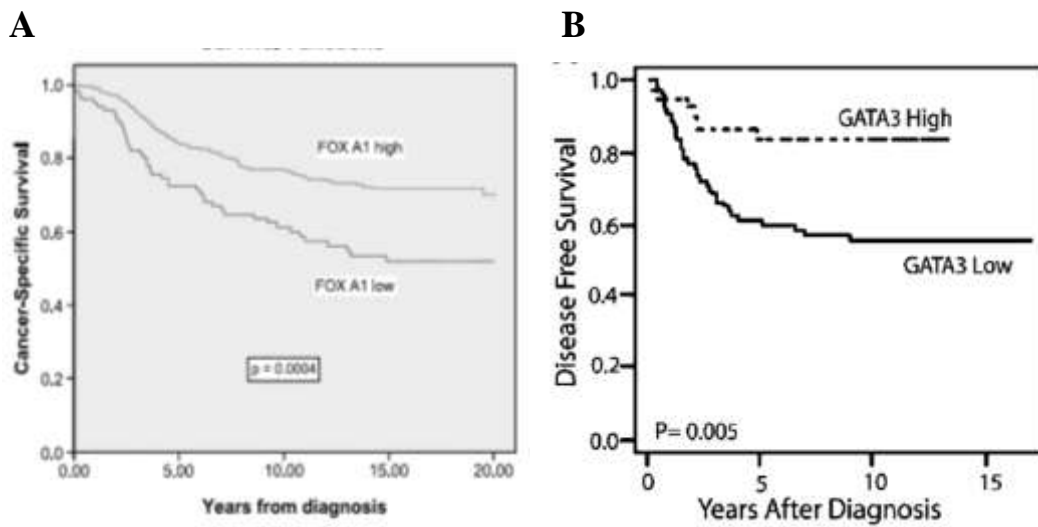
Figure 10. The Kaplan-Meier plot showing the improved cancer-specific survival in patients with (A) FOXA1^{high} score (Badve *et al*, 2007) and (B) high GATA3 expression respectively (Mehra *et al*, 2005).

A

B
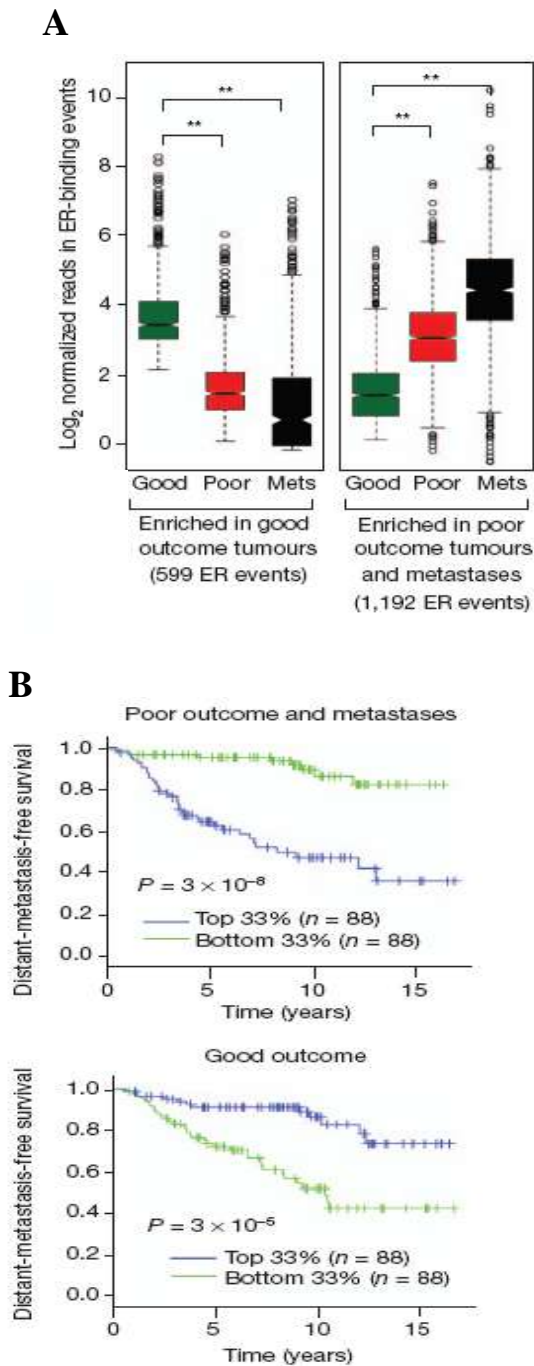
Figure 11. (A) The differential ER binding events that are statistically enriched in either the patients with good outcome (599 ER-binding events) or patients with poor outcome and metastases (1,192 ER-binding events). (B) The genes within 20kb of the differential ER-binding events were useful to predict distant metastases-free survival outcome (Ross-Innes C.S. *et al*, 2012).

1.8 FOXA1 as the key pioneering factor for estrogen receptor activation

The forkhead box A1 (FOXA1) transcription factor was initially discovered to regulate liver and other gut organ-specific genes which are necessary for the development of the endoderm and liver, hence it was named as hepatocyte nuclear factor 3α (HNF3A). Based on the homology of this protein to *Drosophila* forkhead proteins, HNF3A was later renamed to FOXA1 in 2000.

FOXA1 is expressed in the developing mammary gland in conjunction with ERα. Report by Bernado *et al* has demonstrated that FOXA1 is present within the structures that are necessary for mammary morphogenesis and it is expressed in the same developmental stages as ERα (Figure 12). Furthermore, they showed that FOXA1 is required for ERα expression in the mammary epithelium where there is undetectable ERα within the epithelium of FOXA1$^{-/-}$ glands (Figure 13) (Bernado GM *et al*, 2010).

FOXA1 has the winged helix structure which has a helix-turn-helix core of three α-helices flanked by two loops. It binds to the DNA as monomers by recognizing the seven-nucleotide RYMAAYA (R = A or G; Y = C or T; M = A or C) consensus (Pierrou S *et al*, 1994). A winged helix fold remarkably similar to that of forkhead proteins is found in the linker histone H1 and H5, this permits the C-terminus of FOXA1 to interact with histone H3 and H4 (Cirillo *et al*, 2002; Clark *et al*, 1993). A series of elegant papers from Zaret
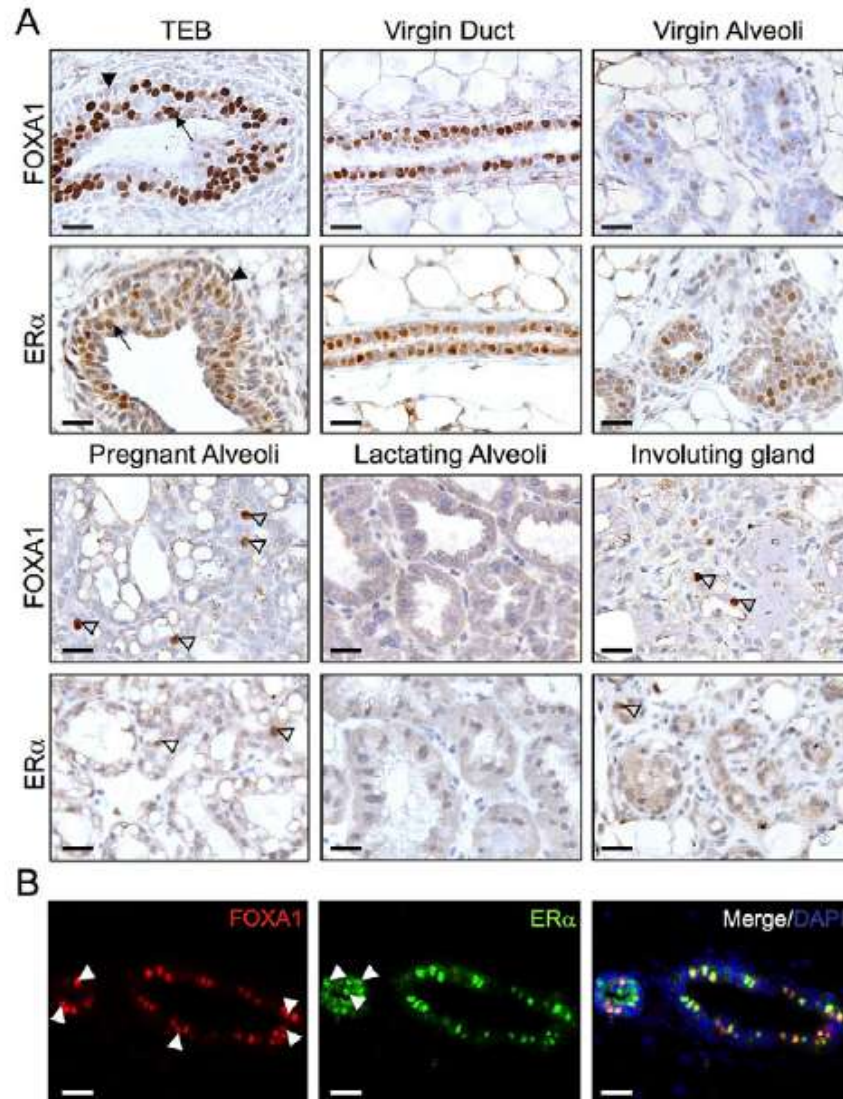
Figure 12. The FOXA1 and ERα expressions at different stages of the developing mammary gland. (A) Representative images of virgin terminal end buds (TEBs) (5 weeks), virgin ductal epithelium (8 weeks), virgin alveoli (20 weeks), pregnant alveoli (day 18), lactating alveoli (day 2) and an involuting gland (day 5). (B) The immunofluorescent image of ERα and FOXA1 in the virgin ductal epithelium (Bernado *et al*, 2010).
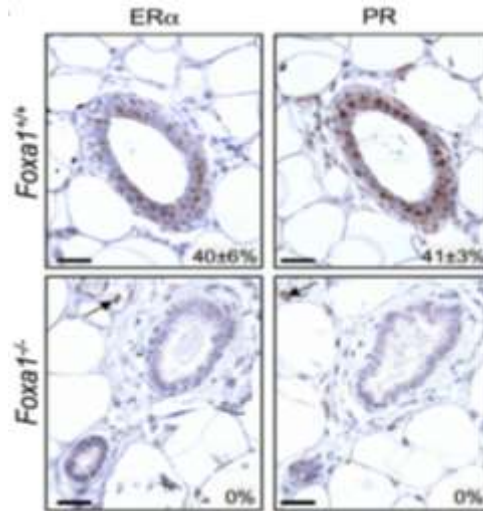
Figure 13. FOXA1 is required for expression of ERα in the normal mammary gland. Images of ERα and PR expressions in the mammary gland of Foxa1$^{-/-}$ mice harvested 4-5 weeks post-transplantation. There is depleted ERα and PR expression in Foxa1$^{-/-}$ mice (Bernado GM *et al*, 2010).

and co-workers suggested that this structural similarity is functionally significant. The binding by FOXA1 to nucleosomes was independent of the histone acetylation, it converted the chromatin to a confirmation that permits binding of additional TFs (Cirillo *et al*, 2002). Consequently, the FOXA1 protein has been proposed to operate as 'pioneer' factor that can displace linker histones from the compacted chromatin and facilitate the binding of other TFs.

Factors that facilitate DNA binding of liganded steroid receptor to chromatinized DNA *in vivo* were poorly understood. The study by Carroll *et al* on the mapping of ERα binding sites in chromosomes has revealed a specific role for FOXA1 as pioneering factor which prepares genomic sites for the recruitment of ERα to ~50% of target genes (Carroll *et al*, 2005).

Recent work by Hurtado *et al* further suggested that FOXA1 is the primary determinant of ER binding and transcriptional activity in breast cancer cells (Figure 14), under both estrogenic and tamoxifen-treated conditions (Hurtado A *et al*, 2011).
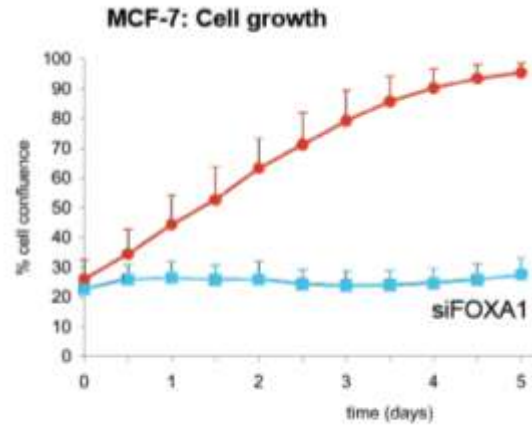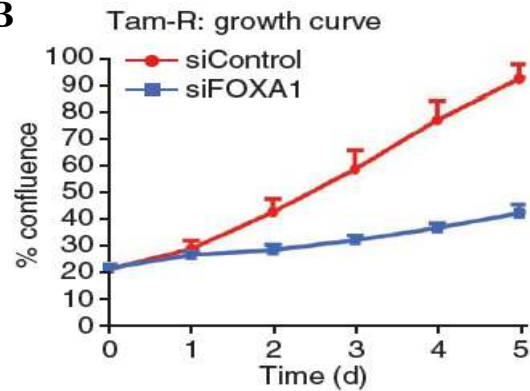
Figure 14. FOXA1 is required for estrogen and tamoxifen-responsive growth. (A) The MCF-7 cells demonstrated repressed growth in the absence of FOXA1. (B) The hormone-depleted Tamoxifen-resistant (Tam-R) MCF-7 cells displayed inhibited cell proliferation in response to Tamoxifen treatment (Hurtado A *et al*, 2011).

Report by Lupien *et al* demonstrated that cell type-specific recruitment of FOXA1 to the chromatin is linked to the breast and prostate cancer transcriptional programs through specific collaborations with ERα in breast cells and androgen receptor (AR) in prostate cells (Lupien *et al*, 2008). Their work has further illustrated how FOXA1 recruitment occurs primarily on H3K4me1/me2 regions and regulates differential transcriptional programs through its collaborations with cell type-specific TFs (ERα and AR) as well as ubiquitously expressed TFs (AP-1 and Sp1) (Figure 15).

Though FOXA1 has been implicated as the pioneer factor with the capability to open the closed chromatin and facilitate the binding of other TFs, a substantial fraction of FOXA1 bound sites still harbor relatively closed chromatin structure with low formaldehyde assisted isolation of regulatory elements (FAIRE) signals (Eeckhoute *et al*, 2009). This indicates that FOXA1 recruitment is required, but it is not sufficient to trigger full opening of the chromatin and induce full functional activity for positive gene regulation (Figure 16). This observation has highlighted that FOXA1 may require a repertoire of collaborating TFs to promote chromatin opening and may impart the cell-type specificity of FOXA1 function. Herein, we hypothesized that ERα and GATA3 are such collaborating TFs in discerning FOXA1 pioneering functionality in the breast cancer cells.
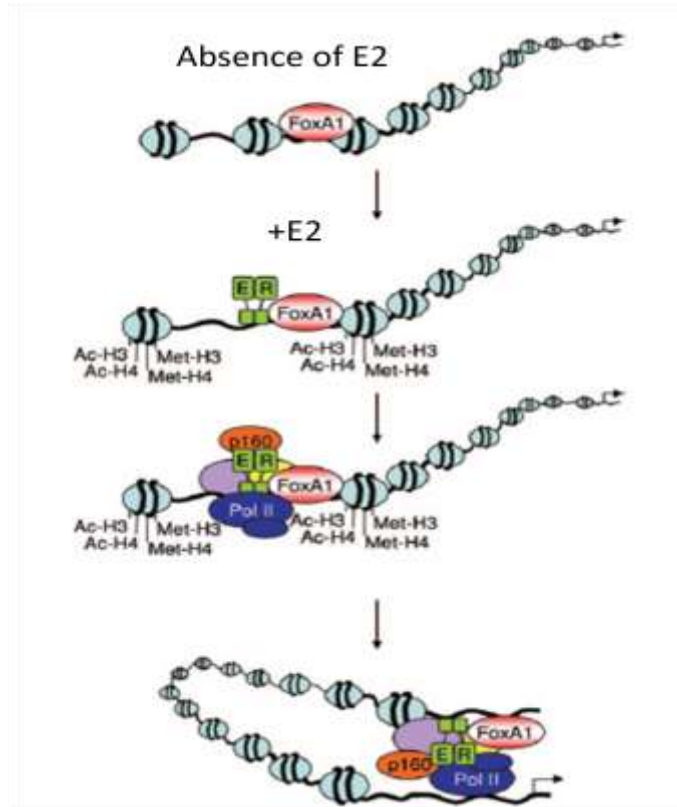
Figure 15. An illustration of ER-mediated transcription in breast cancer cells. FOXA1 interacts with the cis-regulatory regions in heterochromatin and facilitate the interaction of ER with chromatin, follows by p160 cofactors recruitment and histone modification (Carroll and M., 2006a).

31
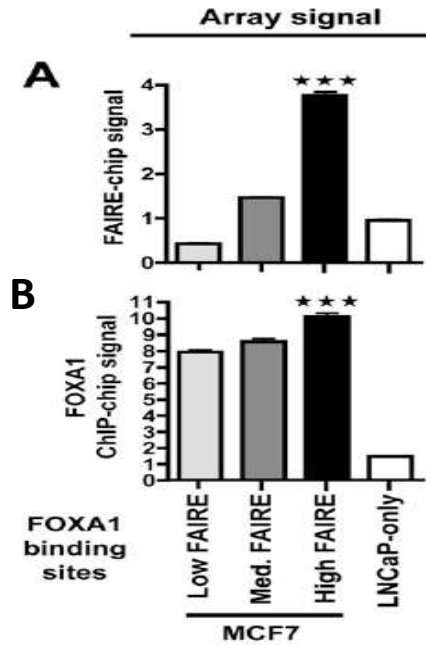
Figure 16. By using FAIRE as the tool to isolate genome-wide nucleosome-depleted DNA, a feature of opened human chromatin, Eeckhoute *et al* has demonstrated that a fraction of FOXA1 bound sites only harbors low FAIRE signal with closed chromatin feature. (A) The average FAIRE array signal was classified as low, medium and high FAIRE in MCF-7 human breast carcinoma cells and the unique LNCaP human prostate carcinoma cells. (B) Signals from FOXA1 ChIP-on-chip in MCF-7 cells were divided into tertiles and associated with the FAIRE-chip signal (Eeckhoute *et al*, 2009).

1.9 GATA3 as the driver of luminal differentiation in mammary gland and co-regulator of ER transcription in breast cancer cells

GATA3 was originally identified as an erythroid cell-specific DNA binding protein that bound to WGATAR consensus sequences (W indicates A/T and R indicates A/G) found in the regulatory regions of many erythroid-specific genes (Wall L. *et al*, 1988). It is important in the development of T-cells, nervous system, kidneys and hair follicle. Its targeted disruption is embryonically lethal with abnormalities including severe aberrations in the nervous system and fetal liver hematopoiesis (Pandolfi PP *et al*, 1995).

GATA3 has been found to be an essential TF for the active maintenance of luminal epithelium in the mouse mammary gland. Targeted deletion of GATA3 resulted in severely diminished mammary epithelial structures, delayed ductal branching (Figure 17), impaired lactogenesis and reduced lobuloalveolar development (Asselin-Labat *et al*, 2007; Kouros-Mehr H *et al*, 2006). The GATA3 deficiency led to an expansion of an epithelial population lacking markers of luminal cells and causing a concomitant block in differentiation. Interestingly, reminiscent of the finding on the inter-dependency of ERα and FOXA1 expression in the mammary epithelium (Bernado GM *et al*, 2010), the ERα expression was also affected in the GATA3-depleted mice, suggesting that these three TFs co-regulate their expression in the mammary gland.
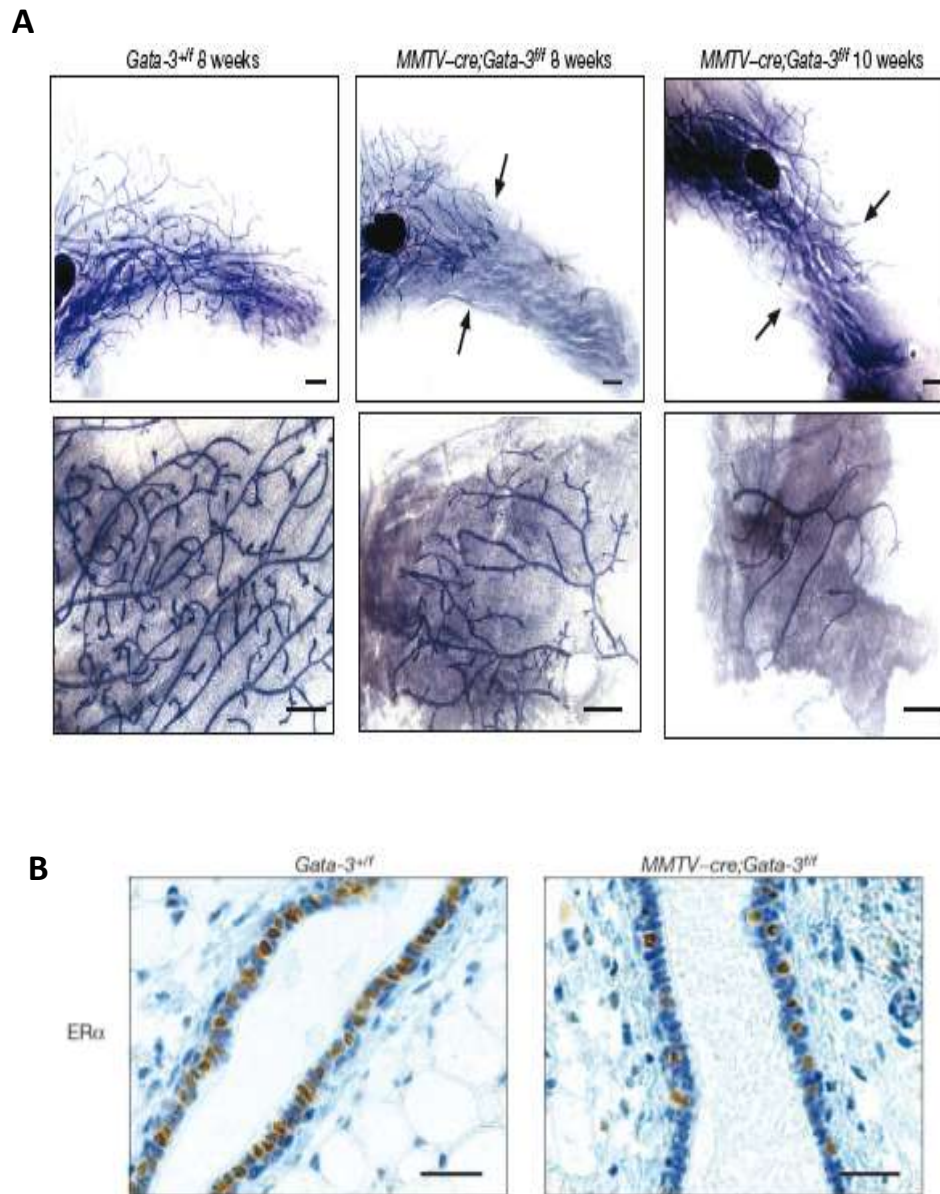
Figure 17. Disrupted development in the MMTV-cre; Gata-3$^{f/f}$ deleted mammary gland comparing to the wild-type control mice. (A) Restricted expansion of the ductal tree within the fat pad of the Gata-3 deleted mice. (B) Immunostaining of ERα in the mammary gland sections from 8-week-old virgin mice revealed that depleted ERα expression in the GATA3-deleted mice (Asselin-Labat *et al*, 2007).

GATA-3 has also been described to possess chromatin remodeling abilities (Shoemaker J *et al*, 2006) whereby ectopic expression of GATA-3 in naïve primary CD4+ cells directly increases chromatin openness at the IL-10 enhancer as well as inducing long-range remodeling of known positive regulatory regions (Figure 18).

GATA-3 is hypothesized to be integral to the ERα pathway supported by the following observation: (1) there is large overlap of the co-expressed genes revealed by meta-analysis between GATA3 and ERα, (2) the highest associated co-expressing gene for GATA3 was ERα and vice-versa, (3) GATA3 and ERα co-expressed with many well-known ERα pathway partners such as *pS2* (Wilson and Giguere, 2008). Furthermore, ERα has been shown to directly stimulate the transcription of GATA3 gene, indicating that these two TFs form a positive cross-regulatory loop that mediate the pro-proliferative signal of estradiol in breast cancer cells (Eeckhoute *et al*, 2007).

Although the coordinate expression and the importance of GATA3 in mammary cell development and differentiation have been well documented, the role of GATA3 in breast cancer is less clear and little experimental data is presently available. Hence, the need for a more complete understanding of the role of GATA3 in the regulation of ERα in breast cancer cell is still warranted.
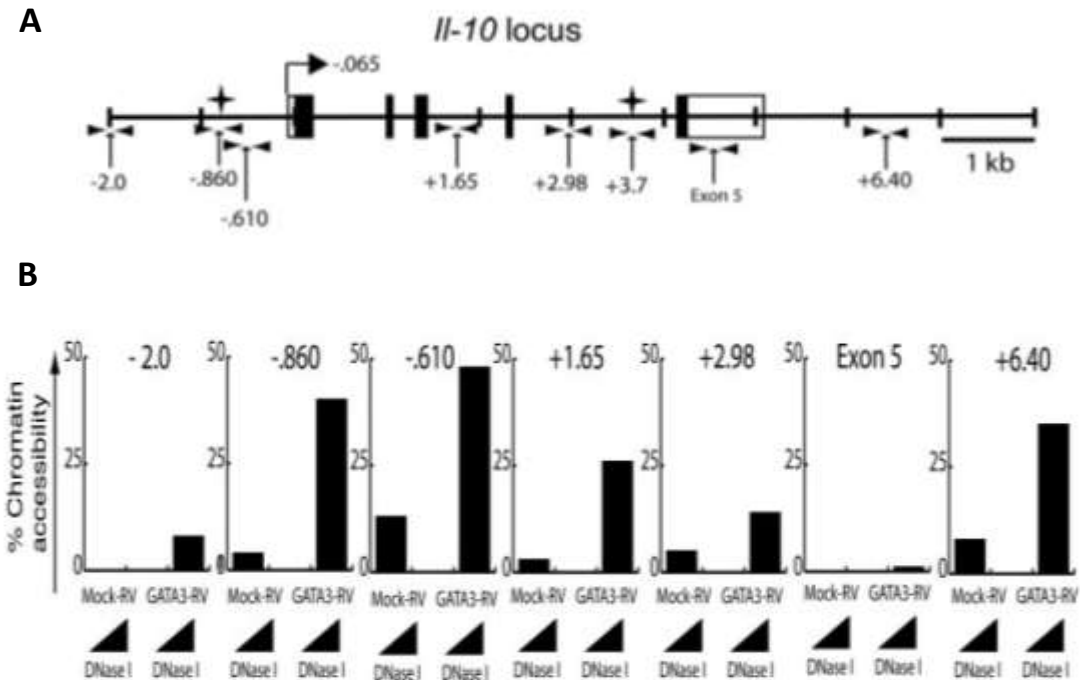
Figure 18. GATA-3 induced changes on the chromatin structure at the *Il-10* locus. (A) Schematic of the fragments of the *Il-10* genomic locus analyzed in this study. (*B)* IL-4$^{-/-}$ naive T cells were transduced with a mock retrovirus (Mock-RV) or a GATA-3-expressing retrovirus (GATA3-RV). In the presence of ectopic GATA-3, the chromatin accessibility at the *Il-10* locus was increased. This effect was most pronounced at sites located in the proximal 5'-region of the *Il-10* gene (HSS -0.860 and HSS -0.610). Chromatin remodeling was also observed at HSS +6.40, suggesting that GATA-3 may also induce long-range remodeling of known positive regulatory regions (Shoemaker J *et al*, 2006).

1.10 The estrogen response cassette in driving the growth and proliferation of breast cancer cells

Estrogen signaling is fundamental to normal mammary gland development and plays a central role in promoting the proliferation of neoplastic breast epithelium. There is long-standing epidemiological evidence that estrogen affects the risk of breast cancer. However, the knowledge and understanding on how this hormone controls human breast development, proliferation and differentiation, and how its action on normal human breast epithelial cells relates to breast cancer risk is still fragmentary.

A hallmark of cancer cells is faulty decision-making: they proliferate when they should be quiescent; they survive when they should be dying, and they invade and move around when they should remain idle. The exact role of estrogen-mediated gene regulation in breast cancer and the manner in which these changes in gene expression affect breast cancer proliferation and progression are far from clear.

ERα is expressed in a subset (only 5-10%) of normal breast epithelial cells and these ERα-expressing epithelial cells do not normally proliferate in response to estrogen (Clarke RB. *et al*, 1997). In contrast to the normal breast, most pre-malignant breast lesions (~70%) express high levels of ERα and proliferate upon estrogenic stimulation. The answer to why do ERα-containing breast cancer cells divide in response to estrogen is still lacking. It is possible that ERα suppresses expression of certain growth factor receptors in normal mammary epithelium and that upon estrogen withdrawal, as occurs during menopause, there is expression of growth factor receptors on ER-positive cells.

Once this has occurred, ER can no longer inhibit the growth factor-stimulated tyrosine kinases and the normal regulation of cell growth is lost (Nilsson S *et al*, 2001).

Current endocrine therapies for ER-positive breast cancers aim at abrogating the ER function at multiple levels. These include ablating the estrogen level, obstructing estrogen action at the ER, and decreasing ER levels. However, the ultimate effectiveness of these therapies is limited by either intrinsic or acquired resistance. Elucidating the factors and pathways responsible for sensitivity and resistance remains a challenge in improving the treatment of breast cancer.

Though estrogen-activated ERα is believed as the prime inducer for the proliferation of breast cancer cell, the finding by Garcia *et al* has yielded unexpected observation where exogenous introduction of ERα into an ERα-negative MDA-MB-231 breast cancer cells shown inhibited growth upon estrogenic stimulation (Figure 19), indicating that factors other than the estrogen receptor are involved in the estrogen-dependent proliferation of breast cancer cells (Garcia *et al*, 1992).

The spatio-temporal expression of ERα and its cofactors is likely to dictate the physiological effects of estrogen. The genetic mechanisms that program cellular growth and proliferation in mammary cells remain to be defined, but it seems likely that these decisions are orchestrated by specific combinations of TFs. Although it is poorly described whether co-regulators levels are related to breast carcinogenesis, it seems likely that changes in the levels or activity of co-regulators would have profound effects on gene regulation that impacts on proliferation and contribute to the development of neoplastic disease.
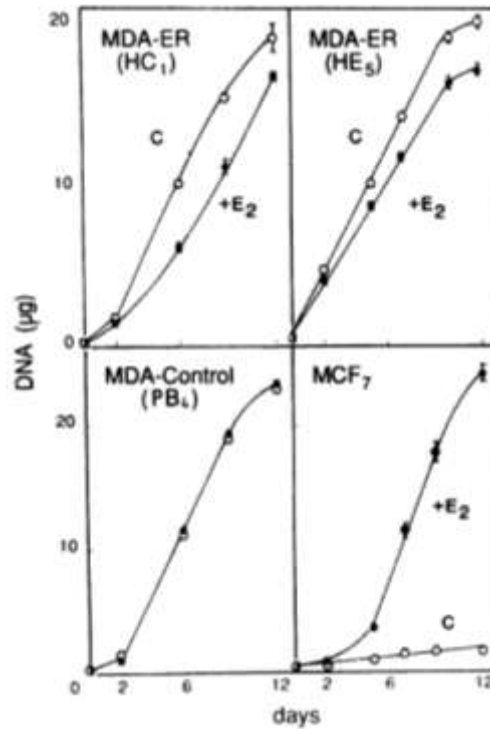
Figure 19. Inhibition of proliferation of MDA-ER cells by estrogen. The MDA-ER cell lines (HC1 and HE5), the control transfected cell line (PB4), and ER-positive breast cancer cell line MCF7 were treated with 20 nM estradiol (•) or ethanol vehicle alone (o) up to 12 days with a medium change every 2 days. The DNA content was determined by the diaminobenzoic acid, and means ±SD of triplicate are represented (Garcia *et al*, 1992).

A positive association between lack of GATA3 and FOXA1 expression and lack of response to hormonal therapy suggests that these TFs may play a role in mechanisms controlling response to estrogen (Hurtado A *et al*, 2011; Parikh P *et al*, 2005). A role for FOXA1 and GATA3 in mediating ERα activity is also suggested by the presence of their specific binding sequences in the promoter of genes involved in both the synthesis (*HSD17B1*) and the degradation (*CYP3A4*) of E_2 (Lacroix M *et al*, 2004).

Harnessing the exact role of many molecular players of ER signaling in regulating the life and death of a breast cancer cell will provide us with the necessary tools to successfully cure the disease.

1.11 Hypotheses and Aims

Aim 1: To decipher the genomics impact of FOXA1 and GATA3 in modulating ERα action

In our previous (Lin *et al*, 2007) and recent (Joseph *et al*, 2010) studies, we have identified high confidence ERα binding sites in MCF-7 human mammary carcinoma cells. With known motif scanning and de novo motif finding methods, we identified that FOXA1 and GATA3 motifs were commonly enriched within ERα binding sites. Moreover, numerous microarray studies have documented the co-expression of ERα, FOXA1 and GATA3 in primary breast tumors (Badve *et al*, 2007; Wilson *et al*, 2008). Though this evidence suggests that these three TFs, ERα, FOXA1, and GATA3 may cluster on DNA binding sites and involved in the breast cancer phenotype, there is little

understanding as to the nature of their coordinated interaction at the genome level or the biological consequences of that detailed interaction.

To address this gap of knowledge, we aim to investigate the ERα-mediated transcriptional networks orchestrated with FOXA1 and GATA3 in breast cancer cells. We will use chromatin immunoprecipitation-sequencing (ChIP-seq) to define the binding profiles of ERα, FOXA1, and GATA3 as to study the interplay among these TFs in breast cancer cells. Specifically, we wish to dissect the roles of FOXA1 and GATA3 in regulating ERα action; to study the progressive recruitment of these TFs to the cis-regulatory elements; and to map the genomic effects of ERα, FOXA1 and GATA3 in altering the transcriptional activation in breast cancer cells.

Aim 2: To investigate the biological consequences by the conjoint action of ERα, FOXA1 and GATA3 in breast cancer cells

ERα is known as a ligand-activated TF that mediates the proliferative effects of estrogen in breast cancer cells. However, some physiologic and cellular contradictions have been previously noted in ERα biology. Garcia *et al* (Garcia *et al*, 1992) showed that the transfection of the ERα alone into ERα negative cell lines has commonly no growth effect or even represses growth. This is also true for an important ERα associated TF, FOXA1, where introduction of this gene represses cell growth (Wolf *et al*, 2007).

We posited that these higher order regulatory mechanisms of ERα function such as the formation and composition of enhanceosomes may explain the establishment of transcriptional regulatory cassettes favoring either growth enhancement or growth repression. As such, we hypothesized that the absence of critical co-regulating TFs such as FOXA1 and GATA3 in the ERα-negative breast cancer cells has resulted inhibited growth in the ERα-negative breast cancer line.

To resolve this, we aim to determine if FOXA1 and GATA3 are essential components of ERα-induced proliferation in breast cancer cells in response to estrogen stimulation.

Aim 3: To study the estrogen-responsive cassettes that drive the growth and proliferation of breast cancer cells

Currently, the most effective therapy in tackling the ERα-positive breast tumors is by interrupting the estrogen-dependent functions in proliferation and survival. Therefore, the discovery of any agent that would influence the growth and proliferation of the breast cancer cells would allow us to gain momentous understanding in the estrogen-responsive growth and ultimately could aid in containing this disease. This motivates us to investigate the estrogen-responsive cassettes that drive the growth and proliferation of breast cancer cells.

To achieve this, we aim to study how the presence of ERα, FOXA1 and GATA3 can orchestrate the symphony of estrogen-inducible growth by evaluating the gene expression

and ERα binding profile of the proliferating estrogen-responsive breast cancer cells. This

allows us to uncover the gene cassettes that drive the growth of breast cancer cells.

CHAPTER 2: MATERIALS AND METHODS

2.1 Cell Culture

The MCF-7, BT-549 and MDA-MB-231 cells were obtained from American Type Culture Collection (ATCC; Rockville MD, USA). The MCF-7 and BT-549 cells were maintained in phenol-red DMEM media supplemented with 10% fetal bovine serum (FBS) and 1% penicillin streptomycin (Invitrogen, CA, USA). The MDA-MB-231 cells were grown in RPMI media supplemented with 10% FBS and 1% penicillin streptomycin. The cells were maintained in $37˚C$ incubator buffered with 5% carbon dioxide (CO_2) where the sub-culturing was performed twice a week using 0.05% Trypsin-EDTA (Invitrogen, CA, USA). The DMEM and RPMI media were purchased from the Biopolis Shared Facilities (BSF), Singapore.

2.2 Chromatin Immunoprecipitation (ChIP) Assay

The MCF-7 cells were grown to approximately 70% confluent before subjecting to ChIP assay. The phenol-red growth media was removed from the cells, followed by three times phosphate buffered saline (PBS) washes for complete removal of phenol-red media before subjecting to the phenol-red free media supplemented with 5% charcoal-dextran treated FBS (CDFBS; HyClone, Utah, USA). The serum-depleted MCF-7 cells were treated with 10nM E_2 (Sigma-Aldrich, St Louis, USA) or vehicle control for 45 minutes. Cells were crosslinked with 1% formaldehyde (Sigma-Aldrich, St Louis, USA) for 10 minutes at room temperature, followed by 125mM glycine (Sigma-Aldrich, St Louis,

USA) treatment for 5 minutes at room temperature to inactivate the crosslinking. The cells were washed with cold phosphate buffered saline (PBS) twice to remove traces of formaldehyde before subjecting to trypsination. The cells were collected from the culture dish using the cell scrapper. The cells were pelleted by refrigerated centrifugation at 3,000 rpm for 5 minutes. The nuclear lysates were collected after three rounds of incubation in Triton-X lysis buffer (0.25% Triton X-100, 10mM EDTA, 10mM Tris-HCl at pH 8, 10mM NaCl and 1X protease inhibitor) with gentle rotation at 4˚C for 10 minutes. The chromatin extracts were fragmentized to an average size of 500bp with sonication using Branson digital sonifier (Branson Ultrasonics, CT, USA) in the SDS lysis buffer (1% SDS, 5mM EDTA, 50mM Tris-HCl at pH 8 and 1X protease inhibitor). The sepharose G-beads (Invitrogen, CA, USA) was blocked with bovine serum albumin (BSA; Sigma-Aldrich, St Louis, USA) for 2 hours at 4˚C. The chromatin extracts were pre-cleared with the BSA-blocked sepharose beads for 2 hours at 4˚C followed by overnight immunoprecipitation in dilution buffer (1% Triton X-100, 2mM EDTA, 20mM Tris-HCl at pH 8, 150mM NaCl and 1X protease inhibitor) added with the ChIP antibody at 4˚C. An aliquot of pre-cleared chromatin was used as the input control. The immunoprecipitated beads were washed with TSE I buffer (0.1% NP-40, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl at pH 8 and 150mM NaCl) , buffer III (0.25M LiCl, 1% NP-40, 1% deoxycholate, 1mM EDTA and 10mM Tris-HCl at pH 8) and TE wash buffer (2mM EDTA and 10mM Tris-HCl at pH 8) at 4˚C for 10 minutes followed by elution (1% SDS, 10mM EDTA and 50mM Tris-HCl at pH 8) at 65˚C for 30 minutes with mixing at 800rpm in the Thermomixer (Eppendorf, Hamburg, Germany). The

protein-DNA complex was subjected to 2 hours pronase (Roche, Mannheim, Germany) treatment at 42˚C followed by de-crosslinked with overnight incubation at 65℃. DNA extraction was performed using phenol/ chloroform (Ambion, Texas, USA) for organic phase separation followed by precipitation with cold ethanol, sodium acetate (Ambion, TX, USA) and GlycolBlue (Invitrogen, CA, USA) as the carrier. The precipitated ChIP and input DNA was measured using pico green quantification (Invitrogen, CA, USA). The quantitative PCR (qPCR) validation was carried out using SYBR Green chemistry on ABI7900 platform (Applied Biosystems, CA, USA). The ChIP samples were then subjected to ChIP-sequencing on Solexa platform (Illumina, CA, USA).

2.3 Preparation of ChIP samples for solexa sequencing.

This protocol describes the preparation of libraries of chromatin-immunoprecipitated DNA using the ChIP-seq Sample Preparation Kit (Illumina, CA, USA) in a format compatible with the Illumina's cluster amplification and sequencing platforms. The objective of the protocol is to add adapter sequences onto the ends of DNA fragments to generate the following template format:

Adapter 1	Sequencing Primer	ChIP DNA fragments	Adapter 2

The 'Adapter 1' and 'Adapter 2' sequences correspond to the two surface-bound amplification primers on the flowcells used in the cluster amplification platform, and the 'Sequencing primer' corresponds to the primer used in the sequencing reaction.

46

The ChIP DNA was end-repaired with T4 DNA polymerase, T4 polynucleotide kinase (PNK) and Klenow enzyme. The exonuclease activity of these enzymes removes 3' overhangs and the polymerase activity fills in the 5' overhangs. This followed by the purification on the QIAquick column using the QIAquick PCR Purification Kit (Qiagen). An 'A' base was added to the 3' end of the blunt phosphorylated DNA fragments using the polymerase activity of Klenow fragment (3' to 5' exo minus). This prepared the DNA fragments for ligation to the adapters which have a single 'T' base overhang at their 3' end.

The 'Adapter 1' and 'Adapter 2' were ligated to the ends of the DNA fragments using DNA ligase to prepare the hybridization of DNA fragments to the flow cell. The adapter-modified DNA fragments were amplified using Pfx polymerase (Invitrogen, CA, USA) in thermacycler machine. The unligated adapater was removed from the amplified DNA by semi-size selection of 200-300bp by the gel electrophoresis on the 2% Ultra low-range agarose gel (Bio-Rad Laboratories, CA, USA) followed by 30mins staining with SYBR Green I Nucleic Acid Gel stain (Invitrogen, CA, USA). The DNA-gel extraction was performed using the Gel Extraction Kit (Qiagen GmbH, Hilden, Germany). The selected DNA fragments were quantified using the DNA 1000 Chip on the Bioanalyzer (Agilent Technologies, CA, USA). The verified DNA fragments were then submitted to sequencing on the Solexa platform (Illumina, CA, USA).

2.4 Gel extraction

The agarose gel was visualized using the UV illuminator, the fragment size ranged from 200-300bp was excised using a sterile blade and placed into a 15ml Falcon tube. The excised gel was weighed and the gel extraction was performed using the Qiagen Gel Extraction Kit (Qiagen GmbH, Hilden, Germany). Three volume of QG buffer was added into the excised gel and incubated at room temperature for 10 minutes until the gel was completely dissolved. One gel volume of isopropanol (Sigma-Aldrich, St Louis, USA) was added to the dissolved sample and mixed thoroughly. The sample was transferred to the QIAquick spin column followed by centrifugation at 13,200 rpm for 1 minute at room temperature. A total of 750μl PE buffer was added into the column and incubated for 5 minutes before subjecting to centrifugation at maximum speed for 1 minute. The DNA was eluted with EB buffer and quantified using the Agilent DNA 1000 chip.

2.5 Quantification assay on Agilent DNA 1000 chip

The quantification of DNA fragment was assayed on the Agilent DNA 1000 Chip on the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). The gel-dye mix was equilibrated to room temperature for 30 minutes in the dark before loading into the Agilent DNA 1000 Chip. The chip was primed on the priming station for 1 minute using the 1ml syringe. The DNA ladder, DNA marker and DNA samples were loaded into the respective wells. The DNA Chip was vortexed for 1 minute at 2,400 rpm before loading onto the Agilent 2100 Bioanalyzer.

2.6 Antibodies used in this study

Antibody	Catalog number	Supplier	Application	Concentration
ERα	Sc-543	Santa Cruz Biotechnology, CA, USA	ChIP; western blot	1μg; 1:500
FOXA1	AB-4124	Chemicon, MA, USA	ChIP	5μg
FOXA1	Ab-5089	Abcam, Cambridge, UK	Western blot	1:500
GATA3	Sc-22205X	Santa Cruz Biotechnology, CA, USA	ChIP	5μg
GATA3	Sc-269	Santa Cruz Biotechnology, CA, USA	Western blot	1:500
RNA pol II	Ab-5408	Abcam, Cambridge, UK	ChIP	5μg
p300	Sc-584X	Santa Cruz Biotechnology, CA, USA	ChIP	5μg
Rabbit IgG	Sc-2027	Santa Cruz Biotechnology, CA, USA	ChIP	1μg
Goat IgG	Sc-2028	Santa Cruz Biotechnology, CA, USA	ChIP	1μg
Rabbit IgG-HRP	Sc-2004	Santa Cruz Biotechnology, CA, USA	Western blot	1:2000
Mouse IgG-HRP	Sc-2005	Santa Cruz Biotechnology, CA, USA	Western blot	1:2000
Goat IgG-HRP	Sc-2020	Santa Cruz Biotechnology, CA, USA	Western blot	1:2000

Table 1. The list of antibodies used in this study.

2.7 Validation of binding sites identified from the ChIP-seq libraries

The binding sites were validated with quantitative Polymerase Chain Reaction (qPCR) using the specific primer sets. The primers are designed around the binding sites and ChIP-qPCR was performed using SYBR Green Chemistry, 1ng of ChIP DNA and 0.5μM of primers (listed in Table 2) in the ABI7500 Read-time PCR System (Applied Biosystems, CA, USA).

No	Primer	Sequence	Coordinates
1	FX-E1	CCGGTTATCACAGGCTGTTC	Chr15: 67674230
		CTGTGTTTGCTCAGCCAATA	
2	FX-E2	ACAATAGCAATCTCAGAGCC	Chr20: 46138302
		GAAGGGGAGAGAGTGAATAT	
3	FX-E3	ATAGCAAGTGGCATTTCAAGGC	Chr2: 220798228
		AAATGGCAGTGGGAGGTTGG	
4	FX-E4	CTCAAGCAGGTTGTTGAAACTTTGGCTC	Chr14: 67393430
		ACATGCCAAACCAGAATAGG	
5	FX-E5	ACAGATCACCAAGAGATAGACC	Chr3: 32126930
		GCTGCCCTAAAGTTCCAAGT	
6	FX-E6	GTATTTCAAGACACTCCTGGTG	Chr20: 14509472
		TGACCTGACCACCTGCTTAA	
7	FX-E7	AAACGGAGATACAGAGACTGAG	Chr9: 12784177
		TTATATTGGCTGCTGCATGG	
8	FX-E8	CCTGTTGTATAACATTGGCC	Chr4: 101634844
		CACCTGACCATGTGTACTTA	
9	FX-E9	AGTCAGAATTGGAGGGGAGC	Chr21: 42669667
		ATCACTCCTTTTCCTGGCTG	
10	FX-E10	TTGGGGTTGTTTGGTTTCAC	Chr13: 25199609
		GCGTTGACAAAAGAGAACAC	
11	FX-E11	AGGGTGAATGAATACCTTATCGGC	Chr8: 30403637
		CCTCAGCCTGTTCAATGCCTTTAAGA	
12	FX-E12	ATTGTACCTTCAAAGTGCCAGG	Chr3: 158013191
		TAGATGAGCCCCAACAGGACCT	
13	FX-E13	AAGAGGCACTGATCTACCAG	Chr11: 131272536
		GTGTGATGACATTTAACTCC	
14	FX-E14	GAATGATGACAAGGTCCCTC	Chr3: 194019535
		GGCAGTTCAGATTCTAGCAA	
15	FX-E15	AATTCTCCAAGACCAAGGTG	Chr22: 17535957
		GATGGCTCTATTCAGATCCC	
16	FX-E16	AGACCTAATGTCTCTAGGACCTAATGGC	Chr8: 29641276
		GGAAAGAGCCACACGAATGTAA	
17	FX-E17	TCCAGGATGACTCATGTGGC	Chr4: 129602041
		CAAGCCATAAAGCCACACGT	
18	FX-E18	GACGGCTTTCAGATTGTGCTAC	Chr17: 23888137
		ATGTGCTGGGAACAGAGCCT	
19	FX-E19	CAGCCATCACTGAGAGTCTTCTC	Chr5: 33586065
		AGGGAAATCAGCAGCCTGCT	

20	FX-E-non binding 1	TCTTTTTCGTCATCTCACCC	chr9:325250
		TGGGCAGGTCTACTTGAAAA	
21	FX-E-non binding 2	GCCTAGACTCAACTGCCTATTTTG	chr8:128856700
		AATACAATGGCTGCCTGATT	
22	FX-E-non binding 3	CTAATTTCTGAGGGGAAGAC	chr20:2169500
		GATAACAACACTGAGGCTCA	
23	FX-V1	ATCACAGGCTGTTCTAATGC	Chr15: 67674230
		CTGTGTTTGCTCAGCCAATA	
24	FX-V2	TATCCAGAGGTTAGTTGTAGGC	Chr17: 56788982
		CAGAAAGGGCTAGAAGAACA	
25	FX-V3	CTACATTTGAAAACAACTGCCCTGCC	Chr20: 39712749
		GTGGAAGTTCTAGCTGTGCA	
26	FX-V4	TCATGTTGCCTGTTCCTTAC	Chr1: 112321182
		CTGCATCTCAACAAGGAATC	
27	FX-V5	TTGGGGACATCCAAGTTGTTTCCAG	Chr1: 230366909
		GCTGACAGACATTCCGCTAGTCACTAAGTA	
28	FX-V6	TCCTACTACCAACACTTGTG	Chr8: 19789795
		TACCTCCATCTCTCTGGTAC	
29	FX-V7	AAAATCTGCATTCCAGGAGC	Chr12: 25568826
		CCACTCTGTAAACCAGTTTGTG	
30	FX-V8	CAGGGAGCCAGCATTCTAGCAAAG	Chr19: 47100617
		AAGACACGCCTCTGTGCAAACACCTG	
31	FX-V9	TTAGGCAAGAGTCATCAGTG	Chr11: 76327997
		CCAGATTTCCCCACTGTTAA	
32	FX-V10	CTGAGGATATTTACCAAGGC	Chr5: 11787705
		CCTGGAGGGATGTGTAATGTAA	
33	FX-V11	CTGGTCTTGTTTATAGAGCC	Chr7: 108041502
		TCAGGTGGAATGAGATCATG	
34	FX-V12	GGTCTTACGTTGAGGAATGG	Chr3: 53527983
		GCGAAGTGGTAAAGCTGTTTGT	
35	FX-V13	ATCAAAGGGGTCAGAGGCTG	Chr7: 101130618
		CAGAGAGGGTGGATGACATG	
36	FX-V14	AAGACATCCACTGGGAAAGC	Chr6: 11385547
		AGTAGAGGCAGAAACCCTTT	
37	FX-V15	CAGCTTGGTGGTTTCTTCTAC	Chr9: 139496534
		TGGACAGAGGTTTTCAGTTC	
38	FX-V16	CTTGTCTCAAAGGCGGAAGG	Chr6: 144241402
		CTCCTTAGGGCAAATCACAG	
39	FX-V17	TGTCCCAGATTGTTTGGTAC	Chr20: 55272806
		TTCAACAGGGGTATGGTGAA	

40	FX-V18	TTTTGCCAGGGTTAATGATGTGCC	Chr13: 68158339
		GATGTCTCATGTCACCCTAA	
41	FX-V19	CCTCTGAGGATGGAGTTTGAG	Chr14: 95277200
		AGCCAAACAGGAAAAGTGGG	
42	FX-V-non binding 1	GAGGTTAGTTTCTTTTCCACCC	chr1: 11216050
		ATCCAAACCCAGGTGTGATC	
43	FX-V-non binding 2	ATCAGGTCCCCAGGTGCCTATATTTC	chr1: 114998951
		CCTGGAATGTGGATGTAATAGCTGGA	
44	FX-V-non binding 3	CACTATTGCTCATCTCTGGAG	chr3: 62880069
		ATCTTTTGCCTGAGCTACCC	
45	GATA3-E1	GAAATCCTTAGAACCAGAGC	Chr11: 1774487
		CATCTTCCAGTGTGGCTAAG	
46	GATA3-E2	ACTCTGGATAGCCTTGGAGG	Chr1: 117613263
		CTGTTCCTGGAAGGATGTTT	
47	GATA3-E3	CCCAGTGGTGTTGACATTCAG	Chr7: 80344311
		GGCTGACTCAGAAGCAAACTGTAGAA	
48	GATA3-E4	GCAACGCCACCTTGACATTC	Chr20: 48814298
		TGCAAAACAGCCACACAAAC	
49	GATA3-E5	TCTAAAACTGCTCTGTCAGTAAGCGC	Chr2: 85340007
		GTTGGGTTTATCGGGCTTTT	
50	GATA3-E6	TGTTTGTGAGCAAAGGAAGG	Chr1: 39381340
		GGCCCTCCTGTGTAAGATAG	
51	GATA3-E7	TTCATCAGCATTTGCGAGTG	Chr22: 44831270
		TTACAGACACCAGCTTCAGGGC	
52	GATA3-E8	TTTGTGGAGAAATGAGCCTG	Chr1: 154853075
		TGCCAACTCAAACAAACAGC	
53	GATA3-E9	GTATCCGCTTGCTTTCCGCATG	Chr8: 12961763
		ATGGGACTGCTGATTGCGCCTT	
54	GATA3-E10	CAAAAGAGGATGAGGATGGG	Chr4: 38143096
		TCCTGTTTACAAGCACGGCT	
55	GATA3-E11	TCTACTAATGCTCATCACTGCC	Chr21: 21541821
		AAAGGAGAAAAGGACGAGGA	
56	GATA3-E12	TTCCCAGCCAGTAATAATCC	Chr20: 37062315
		GTGACCAGGAAAGAGCTGCA	
57	GATA3-E13	AAATGAGTGAAGGTCTGTGC	Chr22: 23696038
		GGGTCTTCTCTGAAGCCTGT	
58	GATA3-E14	AGGAAGGCAAATGATGTCAC	Chr7: 117015702
		TTTCTACTAAGGGACCTCCT	
59	GATA3-E15	ACACTGAGGAAAGGGGAATAAC	Chr11: 101390743
		TCTCTTGCCTCTCAGATTTC	

60	GATA3-E16	TCATCCATTCATTCCCTCAC	Chr9: 78632742
		GAGGCAAGGTTATCTCTAGC	
61	GATA3-E17	GAGTGCGGAGTTTATACTTAGACC	Chr3: 114775880
		CATCTGCTATGAACTCCCTT	
62	GATA3-E18	TCACTTCAACATCTCAGCTC	Chr5: 67358829
		TAGAGATAAGACCAGGCAAG	
63	GATA3-E19	AGTTCTGCTCTTACAGTTCC	Chr12: 82903165
		GGAGCTGTAATAGATTGTGC	
64	GATA3-E-non binding 1	CTTTACTGTGTGTGGTTGAC	chr16:417184
		CCCAAAGGCACCAAAGTAAA	
65	GATA3-E-non binding 2	GTCAACTTCTCTAAGTGTTCTGGG	chr6:144494368
		GCTGACAGCTTTCTTACGTT	
66	GATA3-E-non binding 3	GGATGTTGTAAGGAATGCTG	chr5:136598858
		TCTGAAGGTTTACACTTGCC	
67	GATA3-V1	GGTTGTGTGTGTGTCACTGTCC	Chr1: 199545362
		GAGAGCTTCACCGCCAATGC	
68	GATA3-V2	AAATTCCACCCACAAAGCAG	Chr10: 9270562
		GGTGAGTTATACAATGAAAGGGCC	
69	GATA3-V3	AAGCTAATAGTTCCCAGAGC	Chr9: 136399622
		AAATTACTCTGAGCAGGCAG	
70	GATA3-V4	CTTAGCTTGACAGGCCAGTC	Chr5: 73647995
		GTGTAACAGGACCTCACCAT	
71	GATA3-V5	GGAGCAAAGGACATGAACAG	Chr7: 122670042
		GCCATTTGGACTTGTATGAG	
72	GATA3-V6	ACCCTGAGATCATTCTTCTG	Chr1: 24574381
		CCCAGCAATCACAGGTTTAA	
73	GATA3-V7	CTGGTCACAAGCAGAAATGG	Chr22: 27539809
		TTTCCAGTTTGGAGGAGGGA	
74	GATA3-V8	TTACTACATCTTGCCTGCTC	Chr3: 99129425
		CCAGGATAAGGTTGGAAAGT	
75	GATA3-V9	TGAATCTGCCTTACTCACGC	Chr2: 164658102
		TAGCCTGTGTCAGGTTTGCC	
76	GATA3-V10	CAGTCTCCTTCTGATGCTTAAC	Chr5: 36703083
		GCTGTGAGCACTTGAGTCTTAT	
77	GATA3-V11	TCTCAACAGCCCTTTCTAGTTTGC	Chr11: 78104621
		TGGGCATCATAGATGCAGCT	
78	GATA3-V12	CCCAAGTCTAACACTAGATC	Chr13: 95765479
		GAGTCACTGATTTTCTCCTC	
79	GATA3-V13	GGCAGTTCTGCGCTTTGTC	Chr4: 141665219
		AGGTCAGTGCAGAAAGAGCG	

80	GATA3-V14	ACAGGCTTTCCCTTCTGCAC	Chr21: 21391417
		GCTAACTCTTCAGCACATGGAT	
81	GATA3-V15	CTTTATTTAACCTGGCTCCC	Chr7: 67424263
		GGCTGTCAAAGAAGGATAAGAG	
82	GATA3-V16	TTACCTCTCCAGCAAAATTCCC	Chr6: 10627336
		AGCGCATGAGTCAGCATTTC	
83	GATA3-V17	ACAAATTCCTGTCGCATAGC	Chr9: 136808027
		GGCAAGATAAGAGCATTCTCCA	
84	GATA3-V18	TCACTACACTTACCTTCTCACTCACC	Chr3: 161194854
		AATCTTGCTTGTTTCCCAGC	
85	GATA3-V19	CTTCCTGTGCAAGGCCAAAC	Chr8: 68455991
		CCTTGATAAGCCATCTGAGGAA	
86	GATA3-V-non binding 1	TGATTTAGTCAGCCTTTCCC	chr6:86226853
		AGCGCCATCTCTAGTTGATT	
87	GATA3-V-non binding 2	CACAGTGGGCTTGTCTTCAG	chr7:255747
		GTTCAGTAGGTTAGGGCAGG	
88	GATA3-V-non binding 3	TTCCCTCCACCTACCTTCTC	chr9:1040655
		ACAGAAATCGGGGTCGCTAC	

Table 2. The list of primers used in ChIP-qPCR validation study.

2.8 Sequential ChIP (Re-ChIP)

The ChIP assay was carried out as described previously. The chromatin extracts from the first round of immunoprecipitation with ERα antibody was eluted with 10mM DTT (Bio-Rad Laboratories, CA, USA) at 37˚C for 30 minutes before subjecting to second round of immunoprecipitation using FOXA1 and GATA3 respectively. The sequential ChIP using IgG was employed as the negative control. qPCR was performed to validate co-occupancy of ERα+FOXA1 and ERα+GATA3 to the target sites. The primers used for Re-ChIP are listed in Table 3.

No	Name	Sequence
1	GPR37L1	TAACCCTCCTCTTTGGCTCTG CTAGCCATGTCCTTTCTGCC
2	GREB1	GGCTCCAGTCCAAGTACACAAACTTC TTTTGCTGGGTCACAGTGCTCTCC
3	ITPK1	CTGCCTGCAATCTGTTCCATAC TCAGGTGACGCTGACTGTTG
4	LRRN6A	GTTTGCTGACCAAACACTAGGAAGT CCCACGGAAGCTTAGCTTTA
5	SLC25A25	GCTTTTCCTGTGGAGGCTTC TGGTAGGTACTCGGCAAACC
6	RAD51L1	GTAACAGAACAGGCTGTGCC CAAACAGATGCAAGACAAGG
7	PRKCBP1	GAAGGAACCAAGAGGAAGGAG CCCTGTTTACCTTTGTTTCC

Table 3. List of primers used in the Re-ChIP experiments to investigate the co-occupancy of ERα+FOXA1 and ERα+GATA3 to the target sites.

2.9 Cells synchronization

The MCF-7 cells were grown in phenol-red free DMEM with 5% charcoal dextran-treated FBS (CDFBS) for 3 days before subjecting to 2.5µM α-amanitin (Sigma-Aldrich, CA, USA) treatment for 2 hours. The α-amanitin is able to bind to the large subunit of RNA pol II and block the incorporation of new nucleotides into the nascent RNA chain. The synchronized cells were verified with cell cycle analysis assay. The cells were washed with PBS twice, followed by 10nM E_2 or vehicle control treatment for 45 minutes. Cells were harvested at the 5min interval and subjected to ChIP-qPCR assays. The primers used to study the progressive recruitment of ERα and FOXA1 are listed in Table 4.

No	Name	Sequence
1	GPR37L1	TCTCTGGCAGCTTGTGTCAG
		GGCCACGACAAACAGATTAT
2	KIAA0649	GTGCTTCCTAGCTTTTCTAATGCAGC
		ATGAGAGCAGAGCAGTTGGGAGCTTC
3	DNPEPP	ACAGGGCTGTTTACTTTCAG
		CCCTGGAACTTCATAGACAT
4	RPLP1	AGCTTCCTGTTCTGCTGCCGGTTATC
		TTTGCAGCCTAACACTGGTG

Table 4. List of primers used to study the progressive recruitment of ERα and FOXA1 in synchronized MCF-7 cells.

2.10 Cell cycle analysis

Approximately 1 million synchronized MCF-7 cells were collected by trypsination and washed twice with ice-cold PBS. Cells were then fixed with ice-cold 70% ethanol for 5 minutes at 4°C and then washed twice in ice-cold PBS. The cells were treated with 100μg/ml RNase A in PBS (Roche, Manheim, Germany) for 30 minutes at room temperature before proceeding to 50μg/ml propidium iodide (Sigma-Aldrich, St Louis, USA) stain in dark for 1 hour at room temperature. The cells were filtered through the 60 micron membranes to prevent clumping of cells. The DNA content was determined by flow cytometry (FACScan flow cytometry system, Becton Dickinson, NJ, USA). The proportion of cells in G_0/G_1 phase of the cell cycle was calculated from the DNA histograms.

2.11 Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE)

The MCF-7 cells were grown to ~70% confluence in the 150mm culture dish in phenol-red DMEM media supplemented with 10% FBS. The phenol-red media was removed and the cells were washed with 1x PBS three times for complete removal of phenol-red media before changing into phenol-red free DMEM supplemented with 5% CD-FBS. The serum-depleted cells were grown for 3 days before E_2 stimulation. The cells were crosslinked with 1% formaldehyde at room temperature for 10 minutes followed by the addition of 125mM glycine for 5 minutes to stop the crosslinking. The cells were washed with cold PBS twice and the cells were scrapped from the culture dish and pelleted in the 15ml falcon tube by centrifugation at 3,000 rpm at 4°C for 5 minutes. The nuclear lysates were collected after three rounds of incubation in Triton-X lysis buffer (0.25% Triton X-100, 10mM EDTA, 10mM Tris-HCl at pH 8, 10mM NaCl and 1X protease inhibitor) with gentle rotation at 4°C for 10 minutes. The chromatin extracts were fragmentized to an average size of 500bp with sonication using Branson digital sonifier (Branson Ultrasonics, CT, USA) in the SDS lysis buffer (1% SDS, 5mM EDTA, 50mM Tris-HCl at pH 8 and 1X protease inhibitor). Chromatin DNA was extracted using phenol/ chloroform (Ambion, TX, USA) twice on the Qiagen MaXtract High Density tube (Qiagen GmbH, Hilden, Germany) for phase separation and precipitated using ethanol (Sigma-Aldrich, St Louis, USA), sodium acetate (Ambion, TX, USA) and GlycolBlue (Invitrogen, CA, USA) as the carrier. The precipitated DNA was washed with 70% ethanol twice followed by PicoGreen (Invitrogen, CA, USA) quantification.

2.12 DNA quantification using PicoGreen

The quantification of DNA was performed using Quan-iT PicoGreen dsDNA Assay Kit (Invitrogen, CA, USA). Serial dilutions of the Lambda DNA (100μg/ml) ranged from 2.5 pg/μl to 0.1 ng/μl in TE buffer (10mM Tris HCl and 1mM EDTA) were prepared and added with 1x PicoGreen to generate a standard curve for DNA quantification. The samples were excited at 480nm and fluorescence emission intensity was measured at 520nm using Sunrise microplate reader (Tecan, Männedorf, Switzerland).

2.13 RNA extraction from MCF-7, MDA-MB-231 and BT-549 cells

The RNA extraction was performed by incorporating the use of TRIzol reagent (Invitrogen, CA, USA) and RNeasy Mini Kit (Qiagen, Hilden, Germany) for maximum RNA yield and minimal contamination. The cells were washed with cold PBS twice before subjecting to cell lysis with 1ml of TRIzol reagent. A total of 0.2ml chloroform (Sigma-Aldrich, St Louis, USA) was added followed by vigorous shaking for 15 seconds and a room temperature incubation of 3 minutes before subjecting to centrifugation at 13,200 rpm for 15 minutes at 4°C. The upper aqueous phase was transferred to a fresh 1.5ml tube and 1 volume of 70% ethanol (Sigma-Aldrich, St Louis, USA) was added and mixed well. The sample was then transferred into an RNeasy spin column and centrifuged at maximum speed for 1 minute. This followed by centrifugation with the RW1 and RPE buffers. The sample was eluted in RNase-free water and the RNA quantification was carried out using NanoDrop (Thermo Fisher Scientific, DE, USA).

2.14 Microarray gene expression study on the MCF-7 cells

The MCF-7 cells were grown in phenol-red free DMEM with 5% CDFBS for 3 days before E_2 stimulation. Total RNA was harvested at 3, 6, 9, 12, 24 and 48-hour after E_2 treatment using TRIzol (Invitrogen, CA, USA) and RNeasy Mini Kit (Qiagen GmbH, Hilden, Germany). The quality of RNA samples was verified with Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) before proceeding to Affymetrix microarray experiments on the HG-U133 Plus 2.0 Chip (Affymetrix, CA, USA). The samples were labeled with GeneChip Eukaryotic One-cycle Target Labeling Kit (Affymetrix, CA, USA). The first strand of cDNA synthesis was performed on 1μg of total RNA using oligo dT and SuperScript II (Invitrogen, CA, USA). The IVT labeling was performed at 37˚C overnight followed by cRNA fragmentation at 94˚C for 35 mins and hybridization at the GeneChip Hybridization Oven 640 (Affymetrix, CA, USA) at 45˚C for 16 hours. The GeneChip array was subjected to various washing steps using the GeneChip Fluidics Station 450 (Affymetrix, CA, USA) and scanned with GeneChip Scanner 3000 (Affymetrix, CA, USA).

2.15 Microarray gene expression study on the transfected MDA-MB-231 and BT-549 Cells

The transfected MDA-MB-231 and BT-549 cells were grown in phenol-red free RPMI and DMEM supplemented with 5% CDFBS and selection marker G418 respectively before subjecting to E_2 stimulation. Total RNA was harvested at 6-hour, day 2 and day 10 using TRIzol (Invitrogen, CA, USA) and RNeasy Mini Kit (Qiagen GmbH, Hilden,

Germany). The RNA amplification and biotin-labeling was performed using the TargetAmp^{TM} Nano-g^{TM} Biotin-aRNA Labeling Kit (Epicentre Biotechnologies, MA, USA). The Poly(A) RNA was reverse transcribed into first-strand cDNA using the oligo dT primer and SuperScript III Reverse Transcriptase (Invitrogen, CA, USA). The cDNA:RNA hyrid that was produced during the first strand cDNA synthesis step was converted into double-stranded cDNA containing a T7 transcription promoter followed by the biotin-labeling at 42˚C for 4 hours in a thermocycler (Bio-Rad Laboratories, CA, USA). The sample was then purified using Qiagen RNease Mini Kit (Qiagen GmbH, Hilden, Germany). An equal amount of 750ng RNA was hybridized on the Human HT-12 v4 Expression BeadChip (Illumina, CA, USA) at 58˚C for 16 hours. The BeadChip was then subjected to several washes with E1BC buffer, blocked with E1 Block buffer, stained with streptavidin-Cy3 and scanned on the BeadArray Reader (Illumina, CA, USA).

2.16 Cloning experiment

A linearized pGL4-luciferase vector was generated using SacI and HindIII (New England Biolabs, MA, USA) digestion. The promoter region with the co-occupancy of ERα, FOXA1 and GATA3 bindings at the *GREB1* gene (coordinate - chr2:11588510-11588910) was cloned into pGL4-luciferase construct using the In-Fusion HD Cloning kit (Clontech Laboratories, CA, USA). The In-Fusion PCR primers were designed in a manner that generates PCR products containing ends that are homologous to the vector. The In-Fusion cloning reaction was performed at 50˚C for 15 minutes using the genomic

DNA extracted from the MCF-7 cells and In-Fusion HD Enzyme Premix. The cloned product was verified with sequencing and stored at -20˚C until it is ready for transformation. The In-Fusion PCR primers used are as follows (pGL4-luciferase vector sequences were underlined):

GREB1 forward primer: <u>*CCGGTACCTGAGCTC*</u>GCCAGAGAAGCCCTTTGTAC
GREB1 reverse primer: <u>*CGGATTGCCAAGCTT*</u>AAATGCTGGAGTCGCACCAA

2.17 Site-directed mutagenesis assay

The *in vitro* mutagenesis assay was performed using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, CA, USA) with high fidelity PfuTurbo DNA polymerase. Several oligonucleotide primers containing the mutated ERα, FOXA1, and GATA3 binding motifs were designed (1st BASE, Singapore). The incorporation of mutagenesis primers generates a mutated plasmid. Following thermocycler (Bio-Rad Laboratories, CA, USA) amplification, the product was treated with Dpn I endonuclease to digest the non-mutated parental DNA. The vector DNA containing the desired mutation was stored at -20˚C until it is ready for transformation. The sequences of the *GREB1*-luc construct as well as the mutated-luc constructs were verified with sequencing (1st BASE, Singapore). The mutagenesis primers with the underlined point mutation are listed as follows:

Primers	Sequence 5'-3'
Mutated at GATA3 site	CAGTGTCACCCAGGCCTGGGG<u>G</u>TATTTGCTTCTATTTAG CTCCCAGTAGGTTCCCAG
Mutated at ERE site	GCCTCAGGCCCTGTACTGGAG<u>C</u>TCAGCTCAGTCAGTGTC ACCCAG
Mutated at FOXA1 site	CAGTGTCACCCAGGCCTGGGGGATATTTGCT<u>G</u>CTATTTAG CTCCCAGTAGGTTCCCAG

Table 5. The list of mutagenesis primers.

61

2.18 Chemical transformation experiment

A tube of frozen One Shot Top10 chemically competent cells (Invitrogen, CA, USA) was removed from the -80°C freezer and placed on ice. Approximately 5μl of the cloned DNA was added into the competent cells and mixed by flicking the tube gently. The tube was placed on ice for at least 30 minutes before heat shock for 30 seconds in the 42°C waterbath. The tube was placed immediately on ice for 2-5 minutes before adding 500μl of SOC medium (Invitrogen, CA, USA) and incubated for 1 hour at 37°C with shaking at 250rpm. The cells was pelleted by centrifugation at 3000rpm for 5 minutes and then resuspended in 100μl of SOC medium and plated on the pre-warmed LB plates added with ampicillin. The plate was incubated at 37°C for 16 hours. A single bacterial colony was picked and inoculated in lysogeny broth (LB) to extract the plasmid DNA for downstream work.

2.19 Expression clones

The expression clones encoding human *ESR1, FOXA1* and *GATA3* were purchased from GeneCopoeia (MD, USA). Transformation experiment was performed using these expression clones. After verification with agarose gel electrophoresis (Bio-Rad Laboratories, CA, USA) and sequencing analysis, the plasmid DNA were expanded and purified with Plasmid Maxi Kit (Qiagen GmbH, Hilden, Germany). The plasmid DNA was quantified using NanoDrop (Thermo Fisher Scientific, DE, USA).

2.20 Western blot

The nuclear lysate was extracted using SDS lysis buffer (1% SDS, 5mM EDTA, 50mM Tris-HCl at pH 8 and 1X protease inhibitor). Equal amounts of protein were mixed with 2x Laemelli loading buffer (Bio-Rad Laboratories, CA, USA) and boiled for 5 minutes. The samples were then separated by sodium dodecyl sulfate (10%) polyacrylamide gel electrophoresis gels (SDS-PAGE; Bio-Rad Laboratories, CA, USA) and transferred to nitrocellulose membrane (Bio-Rad Laboratories, CA, USA). The membrane was blocked with 5% non-fat milk (Bio-Rad Laboratories, CA, USA) in Tris-buffered saline (1st BASE, Singapore) containing 0.1% Tween 20 (TBST; Sigma-Aldrich, St Louis, USA) at room temperature for 1 hour. The membrane was washed twice with TBST to remove excess milk and membrane was incubated with the relevant primary antibody in TBST with 5% non-fat milk at 4˚C overnight. The membrane was subjected to four washes with TBST to remove unbound primary antibody prior to incubation with the appropriate HRP-conjugated secondary antibody for 1 hour at room temperature. The membrane was washed four times using TBST to remove excess unbound secondary antibody. The detection for the protein of interest was performed using ECL Prime Western Blotting Detection Reagent (GE, UK) and Kodak Biomax X-ray film (Kodak, NY, USA).

2.21 Transfection experiments

The MDA-MB-231 and BT-549 cells were stably transfected with ERα, FOXA1, and GATA3 using Lipofectamine 2000 (Invitrogen, CA, USA) according to the manufacturer's instructions. The cells were seeded into 6-well plates one day before

transfection experiment so the cells can reach ~90% confluent at the time of transfection. A total of 2μg of plasmid DNA and 6μl Lipofectamine 2000 at the ratio of 1:3 was diluted and mixed gently in Opti-MEM Medium (Invitrogen, CA, USA) without serum respectively. After the 5 minutes incubation at room temperature, the diluted plasmid DNA and Lipofectamine 2000 were combined and incubated at room temperature for 20 minutes. The mixture of plasmid DNA and Lipofectamine 2000 was added into the cells and rocked gently. The cells were incubated in a 37˚C CO_2 incubator with a medium change after 6 hours. The cells were passaged into 10cm culture dish 24 hours after transfection. The selective antibiotic, Geneticin (G418; Invitrogen, CA, USA) was added into the medium the following day. An empty vector transfection was included as a negative control. The G418-selected clonal cells were verified for their ERα, FOXA1, and GATA3 expression using western blot.

2.22 Luciferase Reporter Assays

The MDA-MB-231 cells were transfected with the reporter construct together with different combination of TFs. The MCF-7 cells were transfected with the reporter construct together with the mutated ERα, FOXA1 and GATA3 constructs. A *Renilla* luciferase plasmid was co-transfected as an internal control. Dual-Luciferase Reporter Assay System (Promega, WI, USA) was employed to measure the relative *Renilla* activity according to the manufacturer's instructions. The growth media was removed from the cells and rinsed with 1x PBS. The cells were lysed with 1x PLB and rocked at room temperature for 15 minutes. The Luciferase Assay Reagent was added into each

sample and the light intensity produced was measured using GloMax 96 Microplate Luminometer (Promega, WI, USA).

2.23 Cell proliferation WST-1 assay

The transfected MDA-MB-231 and BT-549 cells were seeded in 96-well plate and subjected to 10nM E_2 or vehicle treatment. The culture media was changed every 3 days and the cell proliferation was assayed with Cell Proliferation Reagent WST-1 (Roche, Manheim, Germany). A total of 10μl Cell Proliferation Reagent WST-1 was added into the cells cultured in 100μl/well (1:10 final dilution). The cells were incubated at 37°C incubator in dark for 1 hour before proceeding to measurement with Sunrise Microplate Absorbance Reader system (Tecan, Männedorf, Switzerland). A blank with the same volume of culture medium and Cell Proliferation Reagent WST-1 was recruited as the background control.

2.24 Cell proliferation with cell count assay

Another cell growth assay assessed by cell number count was performed. The cells are fixed with 2% paraformaldehyde (Sigma-Aldich, St Louis, USA) and incubated at room temperature for 45 minutes. The cells were washed with 1x PBS three times before permeabilization with 0.1% Triton-X (Bio-Rad Laboratories, CA, USA) at room temperature for 10 minutes. The cells were washed with 1x PBS three times followed by Hoechst staining (2.5μg/ml) in dark at room temperature for 10 minutes. The number of

cells were counted using Cellomics ArrayScan VTi machine (Thermo Fisher Scientific, DE, USA).

2.25 Short reads mapping

The reads of 25 bp from each library were mapped independently to the reference genome hg18 using the BatMan package (Tennakoon *et al*, manuscript in preparation). BatMan is a Burrows-Wheeler-Transform-based (BWT) method which maps short sequences to a reference genome at a very high speed.

Up to 2 bp mismatches were allowed. The non-mappable reads and reads with more than 2 "best mapping" genomic locations were removed. The best mapping means, if a read has mapping locations without mis-match (as 0 mis-match), the best mapping will be the locations without mis-match; if a read has no mapping location with 0 mis-match, the mapping locations with 1 mis-match will be the best mapping locations; and the same for 2 mis-matches, etc. Only the best unique mapped location for each mappable read was kept for further processing.

2.26 Binding sites analysis

The short reads from ChIP-seq libraries were aligned to the human genome hg18 using Batman with at most 2 mismatches, and only the uniquely mapped reads were extracted for further analysis. Here, we used the Model based Analysis for ChIP-Seq (MACS) to call the peaks for all the three TFs (Zhang *et al*, 2008) with default parameters. The peaks

were reported as the summit of the enriched regions. The overlap of peaks from two libraries is defined as the peaks within genomic distance 200bp.

Cell type	ChIP-seq libraries	Number of binding sites	Number of uniquely mapped reads
MCF-7	ERα-45 min E_2	19412	8.0 millions
MCF-7	ERα-45 min vehicle	1990	13.6 millions
MCF-7	FOXA1-45 min E_2	15852	13.5 millions
MCF-7	FOXA1-45 min vehicle	9337	19.6 millions
MCF-7	GATA3-45 min E_2	38530	23.6 millions
MCF-7	GATA3-45 min vehicle	20707	16.8 millions
MCF-7	RNA pol II-45 min E_2	Not applicable	7.6 millions
MCF-7	RNA pol II-45 min vehicle	Not applicable	9.6 millions
MCF-7	P300-45 min E_2	Not applicable	12.8 millions
MCF-7	P300-45 min vehicle	Not applicable	16.9 millions
MCF-7	FAIRE-45 min E_2	Not applicable	12.6 millions
MCF-7	FAIRE-45 min vehicle	Not applicable	12.3 millions
MDA-MB-231+ER	ERα-45 min E_2	934	13.2 millions
MDA-MB-231+ER	ERα-Day 10 E_2	940	14.4 millions
MDA-MB-231+ER+FOXA1+GATA3	ERα-45 min E_2	2666	11.0 millions
MDA-MB-231+ER+FOXA1+GATA3	ERα-Day 10 E_2	3777	9.8 millions
BT-549+ER	ERα-45 min E_2	2465	44.1 millions
BT-549+ER	ERα-Day 10 E_2	1544	43.2 millions
BT-549+ER+FOXA1+GATA3	ERα-45 min E_2	2962	49.6 millions
BT-549+ER+FOXA1+GATA3	ERα-Day 10 E_2	4284	44.0 millions

Table 6. The tabulation for the number of reads and binding sites from all the ChIP-seq libraries used in this study.

2.27 Fold change of the numbers of binding sites before and after E$_2$ treatment

As shown in Table 6, there may be concerns that the sequencing depth of the library has some effects on the number of binding sites. To minimize the effect of the sequencing depth on the fold change of the numbers of binding sites before and after E$_2$ treatment, we normalized the fold change of the numbers of binding sites with the sequencing depth. Using ERα library as an example, the sequencing depths are 13.6 million reads before E$_2$ treatment and 8 million reads after E$_2$ treatment. There are 1990 binding sites before E$_2$ treatment and 19412 binding sites after E$_2$ treatment. We computed the fold change of the numbers of binding sites as follows:

$$normalized\ fold\ change\ of\ the\ number\ of\ binding\ sites = \frac{\frac{19412}{1990}}{\frac{8}{13.6}} = 16.58$$

2.28 Motif analysis

The motif scanning was done with the program CentDist (Zhang *et al*, 2011) with motif position weight matrixes from TRANSFAC version 11.3 (Matys *et al*, 2003) with FDR < 1E-3.

2.29 Association of Transcription Factor Binding with Gene Expression

The binding sites were associated with the nearest transcription start sites within 20kb. In order to assign peaks to genes, we look for the nearest transcription start site (TSS) from the Genome Browser by University of California Santa Cruz (UCSC) RefSeq genes for each peak. If the distance from the peak to the nearest TSS is less than 20kb, the peak will be assigned to the TSS. In this way, each peak is assigned to at most one TSS. However, each TSS may have a few peaks assigned from the same TF. If a TSS is the nearest TSS to at least one ERα peak, we will associate this TSS to the ERα peak, regardless of the number of possible ERα peaks within the distance of 20kb. Similarly, the same TSS can be associated to FOXA1 and GATA3 peaks. Based on the condition whether the TSS is associated with ERα, FOXA1 and GATA3 peaks, the TSS will be grouped into 8 different categories: 1) with ERα+FOXA1+GATA3 peaks; 2) with ERα+FOXA1 peaks; 3) with ERα+GATA3 peaks; 4) with FOXA1+GATA3 peaks; 5) with ERα peaks only; 6) with FOXA1 peaks only; 7) with GATA3 peaks only; and 8) without any peaks.

For the category with ERα+FOXA1+GATA3 peaks, the situation can be sub-divided into two sub-categories with/without ERα+FOXA1+GATA3 conjoint peaks as shown in the following illustrations. In the left figure, we associated the TSS with ERα+FOXA1+GATA3 non-overlapped peaks. In the right figure, we associated the TSS with overlapped (or conjoint) ERα+FOXA1+GATA3peaks.

69
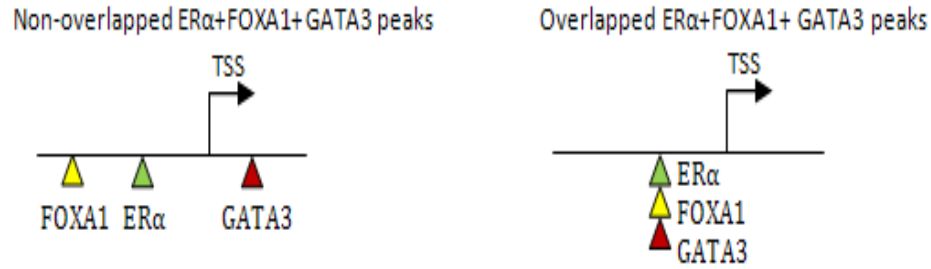
Figure 20. These illustrations demonstrated the two different scenarios where the ERα+FOXA1+GATA3 peaks were associated with the TSS.

2.30 Analysis of the gene expression profiles of MCF-7, MDA-MB-231 and BT-549 transfectant cells

We have three different MDA-MB-231 and BT-549 transfectant cells: transfected with vector control, ERα-only, and ERα+FOXA1+GATA3. These different transfectants were subjected to estrogen and vehicle treatment, the cells were collected at day 2 and day 10 before proceeding to microarray experiment performed on 3 biological replicates.

The gene expression was calculated in the following procedure. The expression of the probes is measured by the fold change of the raw intensities from E_2 treatment over the raw intensities from vehicle treatment, and followed by median-normalization in each replicate. The normalized expression from the probes was averaged from three replicates. The expression of the probes from the same genes was averaged to generate the final expression level of the genes. The genes with fold change over 1.2 (which corresponds to 0.263 after log2 transformation) are called as up-regulated genes, and the genes with fold change lower than 0.83 are called as down-regulated genes. The E_2-regulated genes from the vector-control MDA-MB-231/ BT-549 cells, ERα-expressing MDA-MB-231/ BT-

70

549 cells and ERα+FOXA1+GATA3-expressing MDA-MB-231/ BT-549 cells are overlapped with the E_2-regulated genes from MCF-7, and this list of genes are used to compare the expression from different cells.

For the comparison of MDA-MB-231/ BT-549 transfectants and MCF-7, we used a growth normalized strategy. This means that the time points that are compared are selected by the time of equivalent phenotype. This strategy was used in defining *MYC* effects on apoptosis where different conditions gave the same gene responses but with different phasing (Yu Q. *et al*, 2002). In our experimental case, we found that MDA-MB-231/ BT-549-ERα+FOXA1+GATA3 triple transfectant exhibited E_2 stimulated growth only after day 7 and maximally by day 10. By contrast, MCF-7 showed near optimal growth induction after 24 hours of E_2 exposure. Therefore, our array comparison was between the 24 hours time-point for MCF-7 and 10 days for MDA-MB-231/ BT-549- ERα+FOXA1+GATA3.

2.31 Analysis on the Expression Levels of Luminal and Basal Marker Genes

Based on the recently published luminal and basal marker gene list from Kao *et al* (Kao *et al*, 2009), we included 45 luminal marker genes and 49 basal marker genes (referred to as the luminal/basal cassette) that are present in both our MCF-7 and transfected MDA-MB-231/ BT-549 expression data. We generated the box plots for the expression levels of these luminal and basal marker genes in the MCF-7 and transfected MDA-MB-231/ BT-549 cells after estrogen stimulation. Additionally, we assessed the expression changes of

the luminal/basal cassette from MDA-MB-231/ BT-549 cells transfected with vector control to the MDA-MB-231/ BT-549 cells transfected with ERα+FOXA1+GATA3. The expression level of a gene from the MDA-MB-231/ BT-549 cells transfected with ERα+FOXA1+GATA3 was deducted with the expression level of the same gene from the MDA-MB-231/ BT-549 cells transfected with vector control.

2.32 Genome-wide co-motif analysis using Pomoda

Peak Oriented Motif Discovery Algorithm (Pomoda) (Zhang *et al*, in preparation) is a position-weight-matrix-optimization method for de novo motif finding from ChIP-seq peaks. Pomoda takes into consideration the peak locations and intensities to give more weights to the motifs near the center of the peaks and from the peaks with high peak intensities. It utilizes all the peaks for motif finding, which is possible to identify all the potential co-factors for a transcription factor of interest.

CHAPTER 3: THE CARTOGRAPHY OF ESTROGEN RECEPTOR α, FOXA1 AND GATA3 BINDINGS IN BREAST CANCER CELLS

3.1 The genome-wide mapping of transcription factor binding events

We mapped the genome-wide *in vivo* binding sites of ERα, FOXA1 and GATA3 using the massively parallel ChIP-seq in MCF-7 cells before and after estrogen (E_2) exposure. Using the peak calling algorithm MACS (Zhang *et al*, 2008), we found a total of 1,990 high confidence ERα binding sites, 9,337 FOXA1 binding sites and 20,707 GATA3 binding sites in the vehicle-treated cells (i.e., without ligand) (Figure 21A).

Upon E_2 stimulation, we found a total of 19,412 high confidence ERα binding sites (Figure 21B; an increase of ~16.58 fold after normalization of library size, see details in Materials and Methods section 2.27 and Table 6), 15,852 FOXA1 binding sites (an increase of ~2.46 fold after normalization) and 38,530 GATA3 binding sites (an increase of ~1.32 fold after normalization).
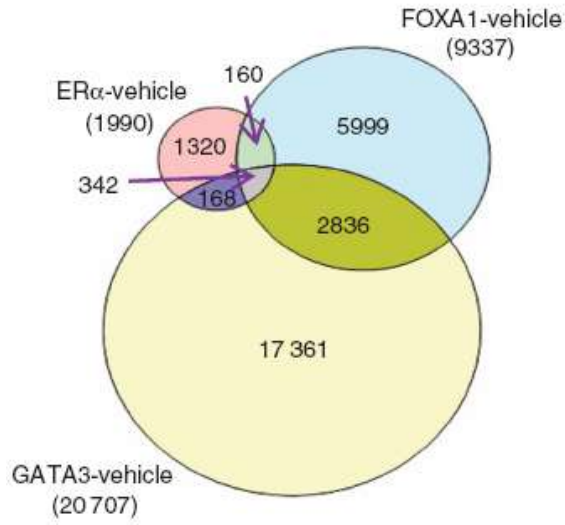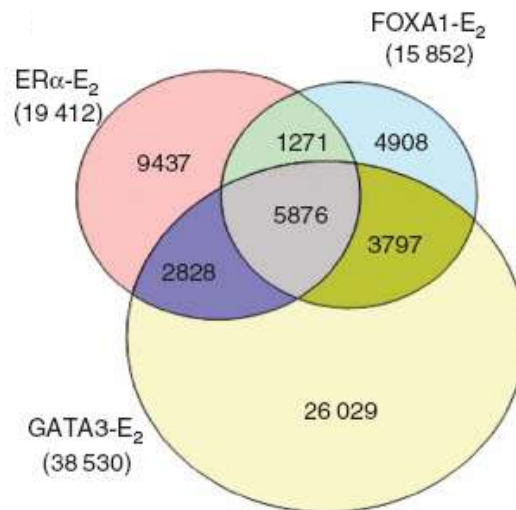
Figure 21. The portrait of ERα, FOXA1 and GATA3 binding using the massive parallel ChIP-seq approach in the human mammary carcinoma, MCF-7 cells in (A) vehicle-treated and (B) estrogenic condition.

In order to validate the binding events, several primer pairs were designed around the randomly-selected binding sites with different binding intensity and a few non-binding sites were also included as the negative control. The ChIP-qPCR results showed 100% concordance in calling bound sites (Figure 22 and 23).

Quantitatively, we found that the ChIP enrichment assayed by ChIP-qPCR for FOXA1 and GATA3 binding sites were correlated well with the binding intensity measured by ChIP-seq (correlation coefficient R= 0.63-0.81, Figure 24).
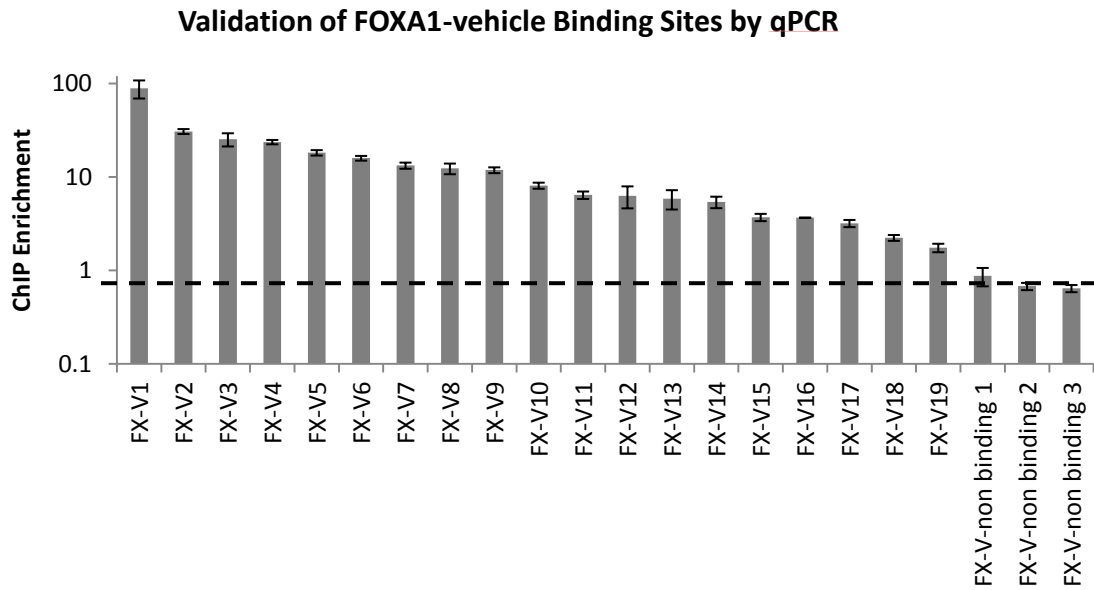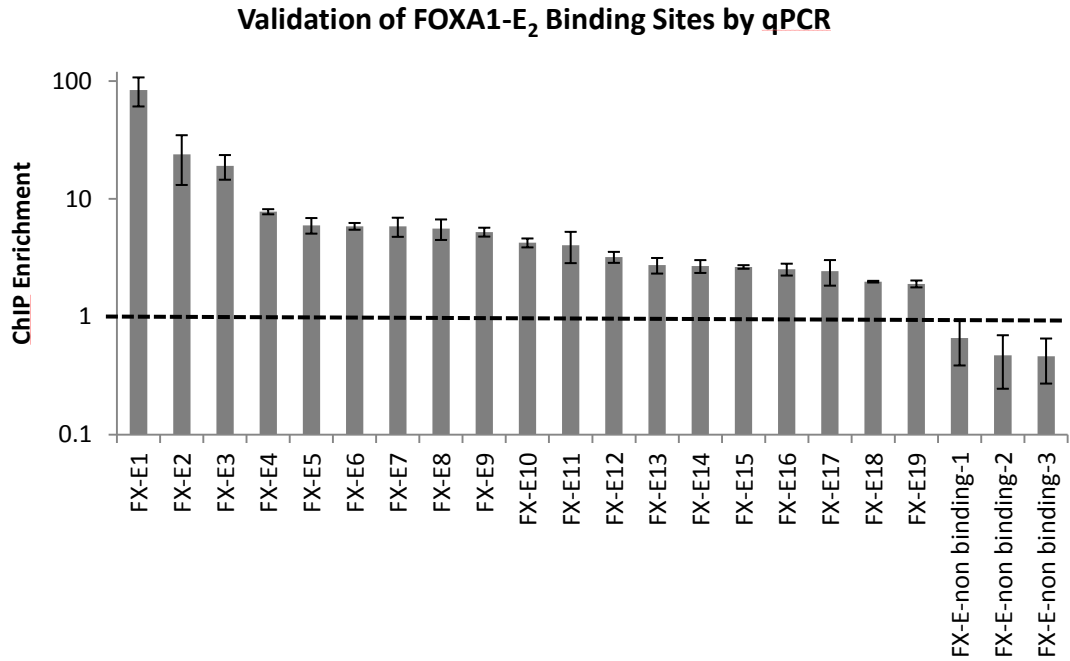
Figure 22. The validation of FOXA1 binding on randomly selected sites with different binding intensity by ChIP-qPCR. A few non-binding sites are included as negative control. The error bars show the standard errors of the means of binding enrichments from at least two biological replicates.
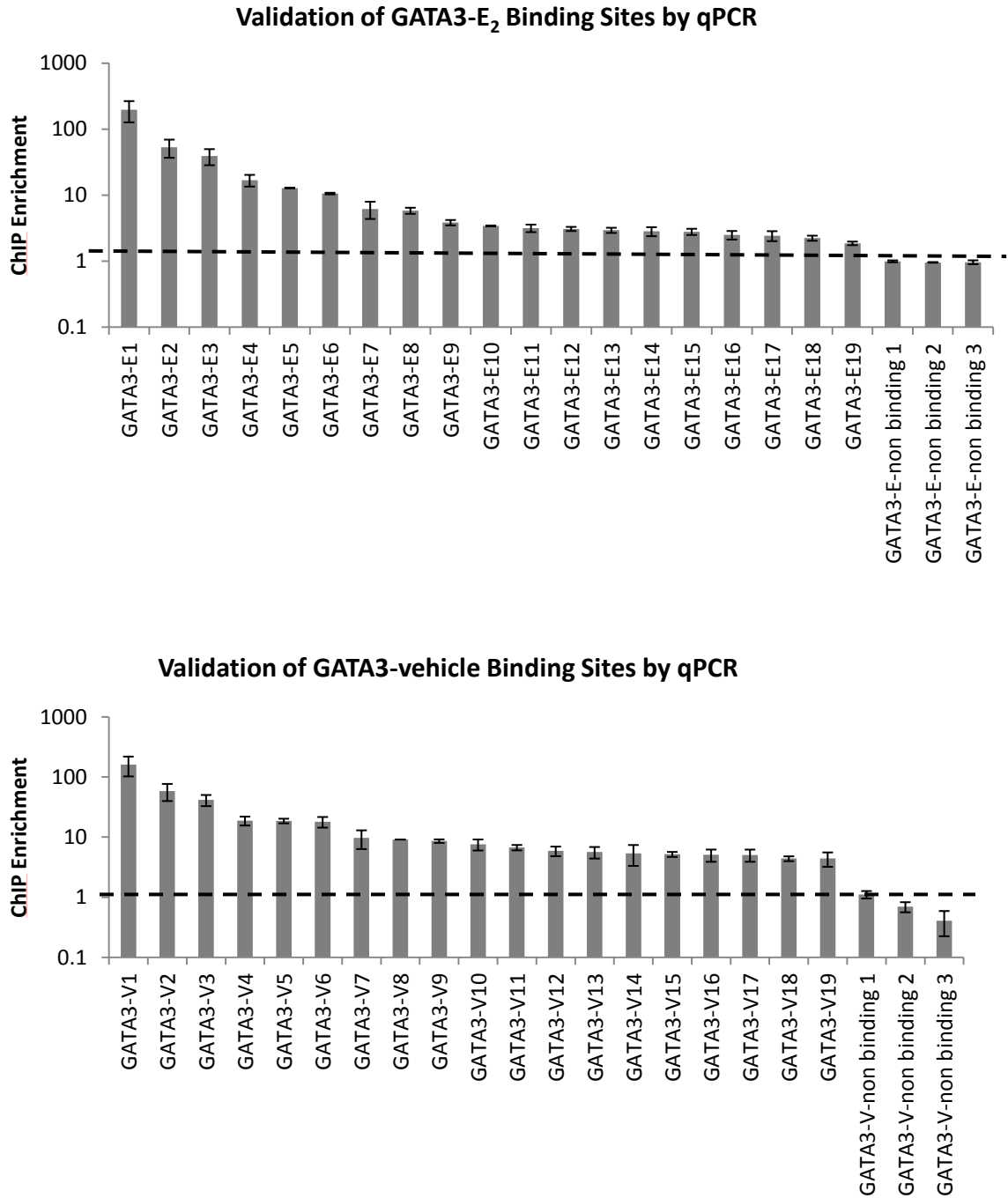
Figure 23. The validation of GATA3 binding on randomly selected sites with different binding intensity by ChIP-qPCR. A few non-binding sites are included as negative control. The error bars show the standard errors of the means of binding enrichments from at least two biological replicates.
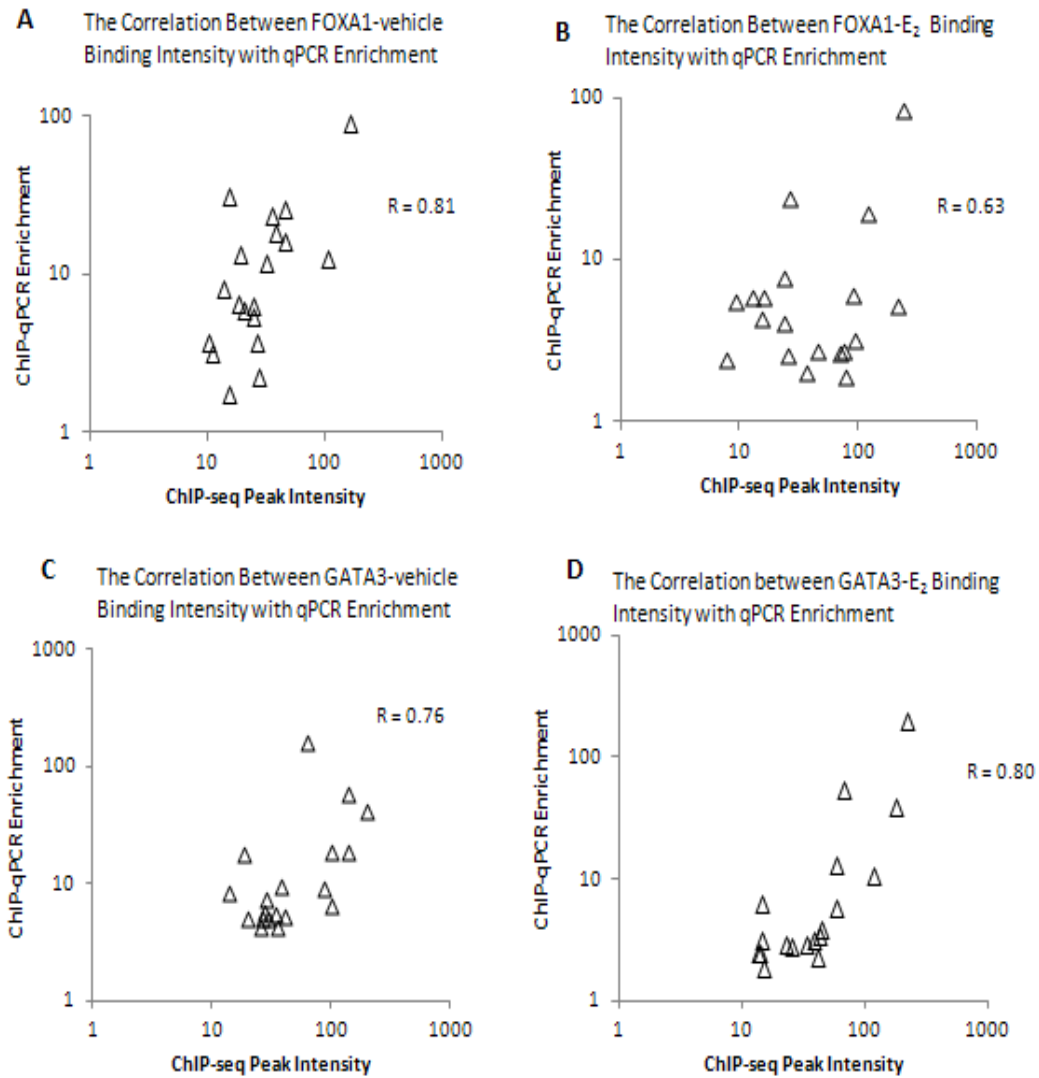
Figure 24. The scatter plots on the correlation between (A-B) FOXA1 and (C-D) GATA3 peak intensity measured by ChIP-seq after normalization to input control and binding enrichment assayed by ChIP-qPCR.

To assess how these TFs individually interact, we overlapped their binding profiles and found 37% of ~19k ERα binding sites showed FOXA1 co-localization, while 45% of ERα binding sites overlapped with GATA3 binding sites, and as much as 30% of ERα binding sites were co-occupied by both FOXA1 and GATA3 (Figure 21). Interestingly, the number of sites with occupancy of all three TFs increased from 342 (before estradiol exposure) to 5,876 (after estradiol exposure).

Our data revealed that FOXA1 and GATA3 bindings are symmetrically distributed within 200bp around the 5,876 ERα, FOXA1 and GATA3 conjoint binding sites (Figure 25). The relative intensity of bindings as measured by TF occupancy at these conjoint sites was highly correlated amongst the three TFs (R = 0.48-0.63, Figure 26). These results suggest that ERα, FOXA1, and GATA3 bind in a coordinated fashion at ~30% of all ERα binding sites after stimulation by ligand.
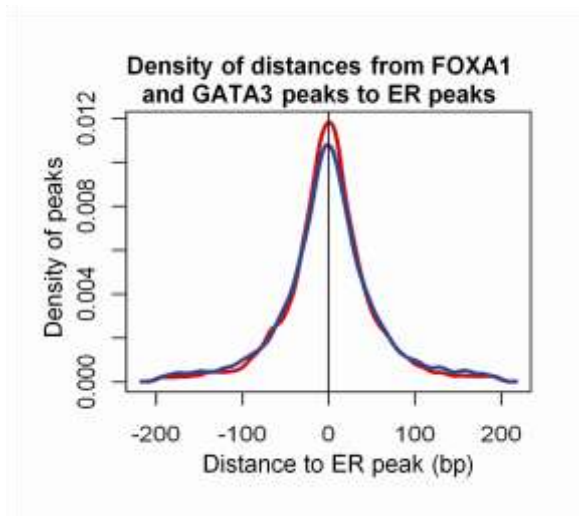
Figure 25. The distance distribution of FOXA1 and GATA3 bindings around ERα sites demonstrating that FOXA1 and GATA3 have similar distance distribution around ERα sites.
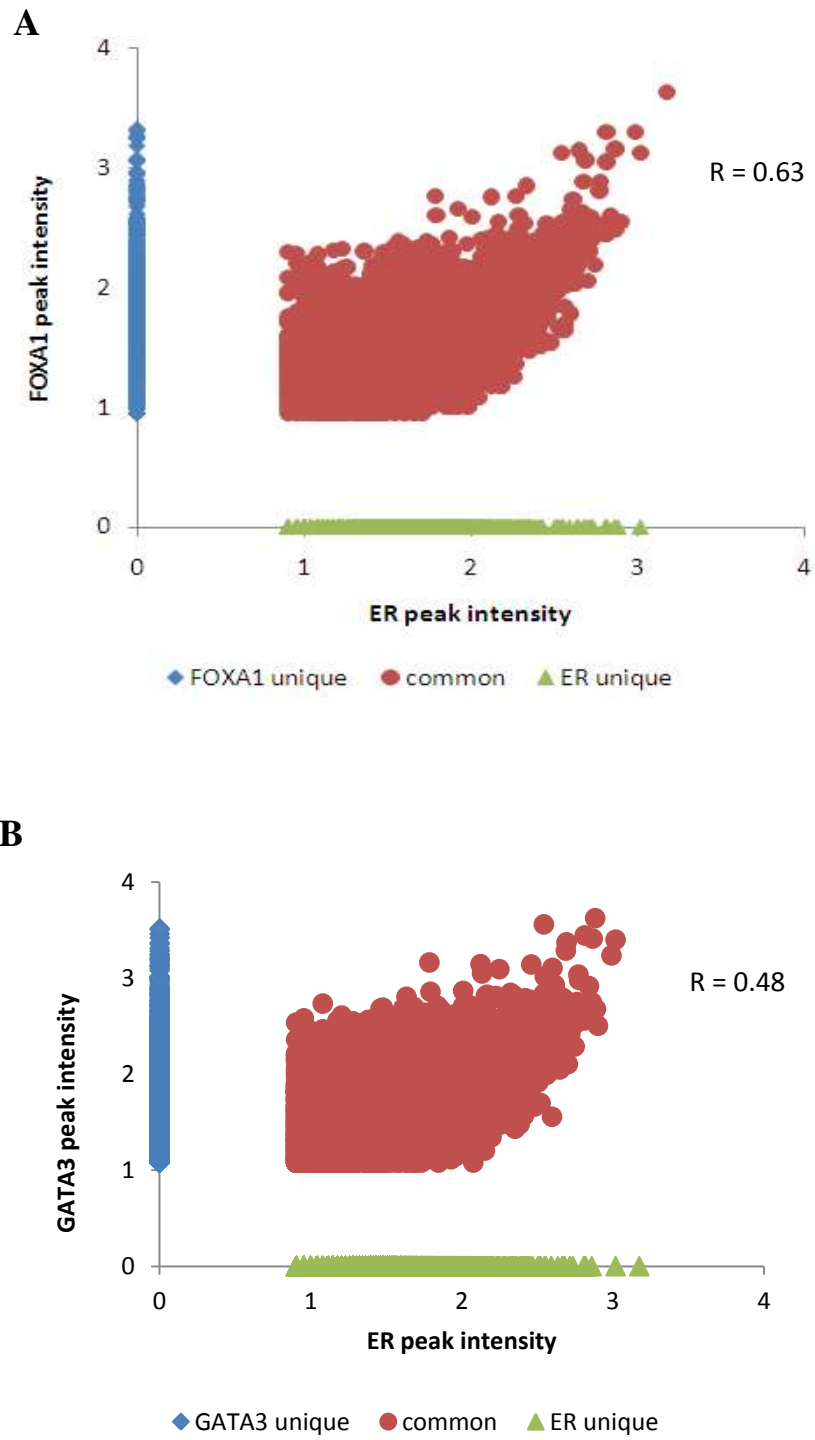
Figure 26. The correlation of ERα peaks intensity with (A) FOXA1 and (B) GATA3 peaks intensity. The peak intensities are in log10 scale.
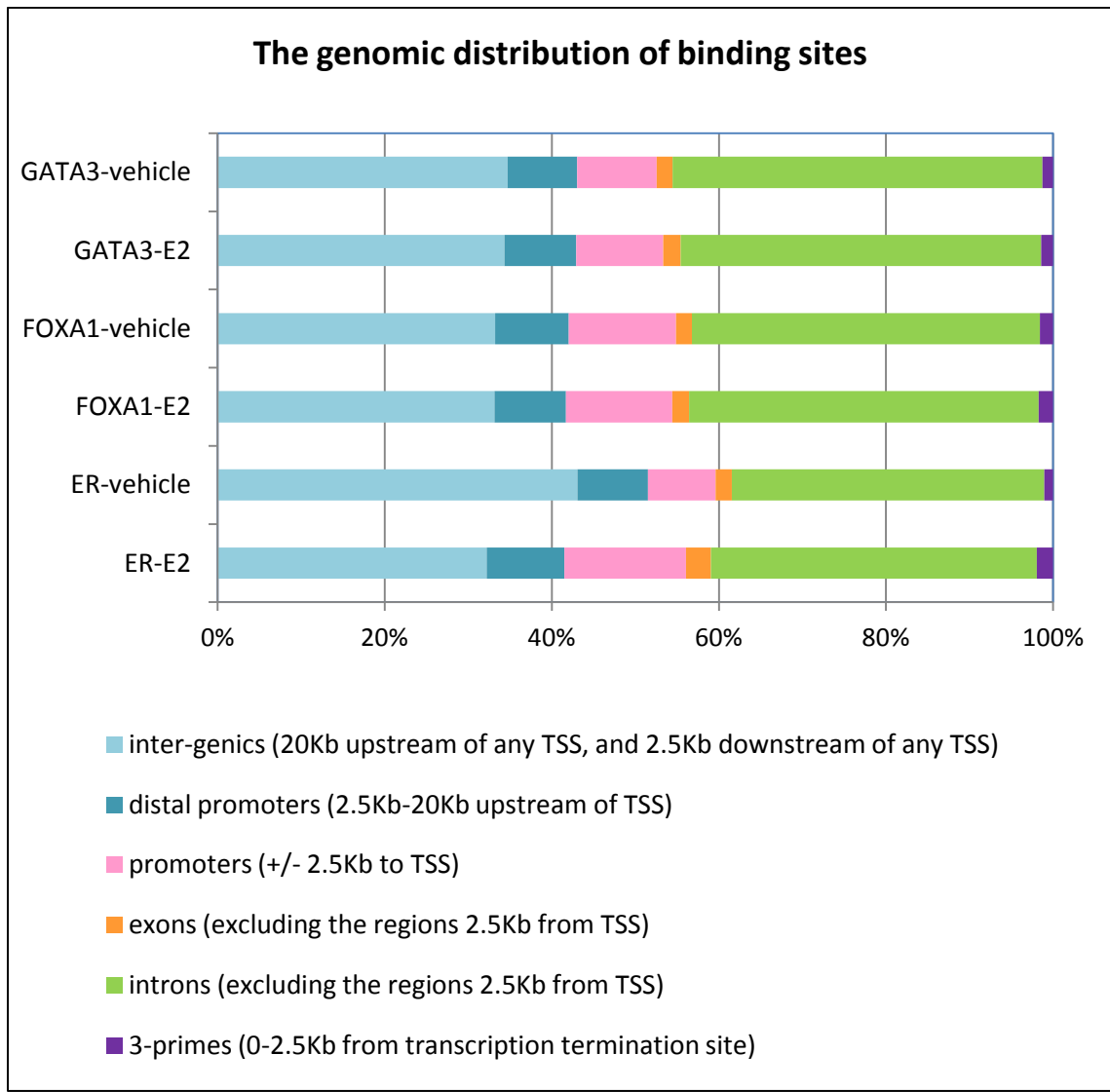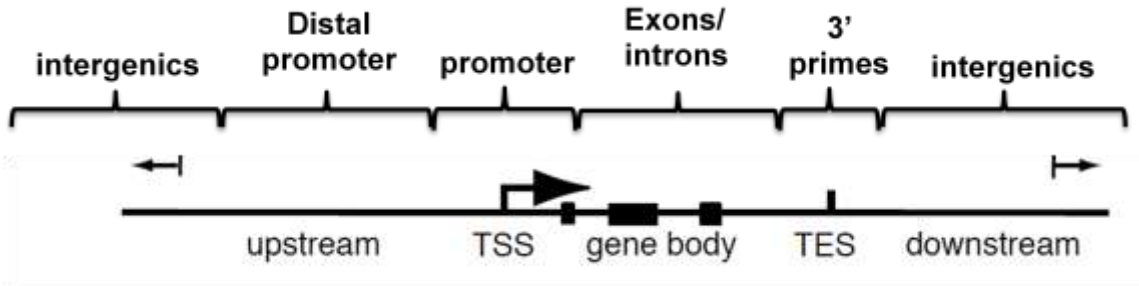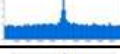
Figure 27. The illustration of the distribution of binding sites to various genomic locations. Majority of the binding events was found in the enhancer and gene body regions (TSS = transcription start site; TES = transcription end site).

We investigated the distribution of the binding sites to various genomic locations and found that majority of the binding events occurred at the enhancer regions (Figure 27). This observation is in agreement with previously reported works where ERα and FOXA1 were recruited mainly to the enhancers (Carroll *et al*, 2006b; Lupien *et al*, 2008).

3.2 Motif analyses of the TFs binding

In order to determine the *in vivo* sequences enriched in the ERα, FOXA1 and GATA3 occupied sites, we used an in-house program CentDist (Zhang *et al*, 2011) for known motif scanning. This program not only allows for the identification of specific binding motifs, but also displays the position-distribution around the binding sites to indicate binding specificity. The motif position weight matrixes (PWM) from TRANSFAC (Matys *et al*, 2003) version 11.3 were used and the cutoff of PWM score was set to 1E-3. As expected, we found significant enrichment of individual ERE, FOXA1 and GATA3 motifs in the ERα, FOXA1 and GATA3 ChIP-seq libraries respectively (Figure 28). We also observed that the three ERE, FOXA1 and GATA3 binding motifs emerged together as the top enriched motifs in each set of the individual TF binding sites, suggesting a bias for recognition motifs for all three factors to be clustered together. Besides FOXA1 and GATA3 motifs, AP1 and BACH motifs were also enriched in the ERα binding sites, which is in agreement with the previous finding reported by Bhat-Nakshatri *et al*. (Bhat-Nakshatri P. *et al*, 2008). Because of prior genetic data suggesting a role for FOXA1 and GATA3 in breast biology, we pursued the interaction of these three factors.

Motif scanning around ERα peaks

Rank	Family	Transfac	Logo	Score	Histogram
1	ERE	V_ER_Q6		3149.04418409 out of 19412(16.22%)	
2	BACH	V_BACH2_01		1414.29586222 out of 19412(7.29%)	
3	AP1	V_AP1_C		1409.68897897 out of 19412(7.26%)	
4	AP2	V_AP2ALPHA_02		867.22051178 out of 19412(4.47%)	
8	FOX	V_FREAC4_01		537.677461461 out of 19412(2.77%)	
21	GATA	V_GATA1_05		257.6736209 out of 19412(1.33%)	

Motif scanning around FOXA1 peaks

Rank	Family	Transfac	Logo	Score	Histogram
1	FOX	V_FREAC4_01		2139.71271235 out of 15852(13.5%)	
2	BACH	V_BACH2_01		1586.27460573 out of 15852(10.01%)	
3	AP1	V_AP1_01		1569.32988538 out of 15852(9.9%)	
4	ERE	V_ER_Q6_02		1172.75931914 out of 15852(7.4%)	
14	GATA	V_GATA_Q6		256.237253682 out of 15852(1.62%)	

Motif scanning around GATA3 peaks

Rank	Family	Transfac	Logo	Score	Histogram
1	GATA	V_GATA_Q6		3260.03137672 out of 38530(8.46%)	
2	AP1	V_AP1_Q6		3202.38201603 out of 38530(8.31%)	
3	BACH	V_BACH2_01		3104.17775623 out of 38530(8.06%)	
4	FOX	V_FREAC4_01		1628.99388434 out of 38530(4.23%)	
6	ERE	V_ER_Q6_02		1429.60275937 out of 38530(3.71%)	

Figure 28. Motif scanning around ERα, FOXA1 and GATA3 peaks using CentDist (Zhang Z. *et al*, 2011) .

Next, we investigated the de novo motifs enriched in the respective individual ChIP-seq libraries. The results revealed that those de novo motifs emerging from the ERα, FOXA1 and GATA3 binding sites resembled the conventional ERE, FOXA1 and GATA3 motifs (Figure 29) with minimal changes comparing the binding sites prior or after E_2 stimulation.

We specifically assessed the frequency of ERE, FOXA1 and GATA3 binding motifs around ERα binding sites. Figure 30 showed that 72% of the ERα sites have an ERE motif, 71% of ERα sites contain a FOXA1 motif, 43% of ERα sites contain a GATA3 motif, and 23% of the ERα sites contain all three motifs.
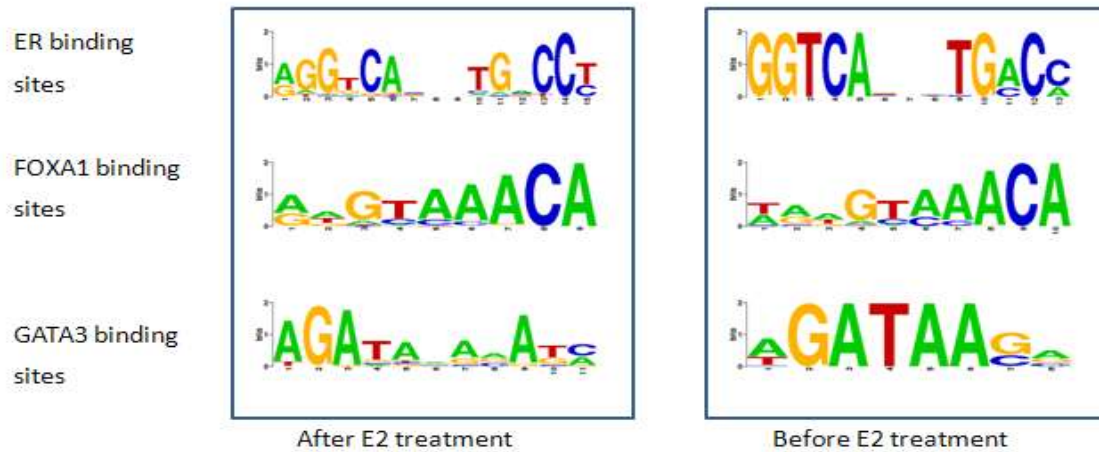
Figure 29. The de novo motifs enriched in ERα, FOXA1 and GATA3 binding sites prior or after E_2 stimulation.
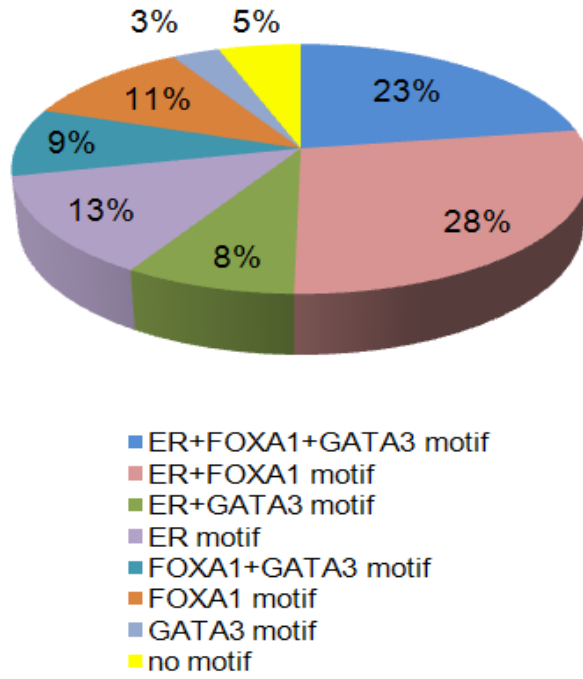
Figure 30. The distribution of ERα, FOXA1 and GATA3 motifs around ERα binding sites.

3.3 Co-occupancy of ERα+FOXA1 and ERα+GATA3

Since a large portion of ERα binding sites was bound by FOXA1 and GATA3, we sought to dissect the interaction between these three TFs by investigating whether these factors are physically co-localized. Using sequential ChIP followed by qPCR in randomly selected sites conjointly occupied by ERα+FOXA1 and ERα+GATA3, we found co-occupancy of these TFs at the overlap binding sites (Figure 31).

These results suggest that the co-localization of the FOXA1 and GATA3 at ERα occupied sites occur through sequence recognition and not solely through a tethering mechanism involving only protein-protein interaction. Indeed, extensive co-immunoprecipitation assays failed to demonstrate direct interaction between ERα and FOXA1; ERα and GATA3 in various conditions and cell lines (Eeckhoute *et al*, 2007; Wolf *et al*, 2007).
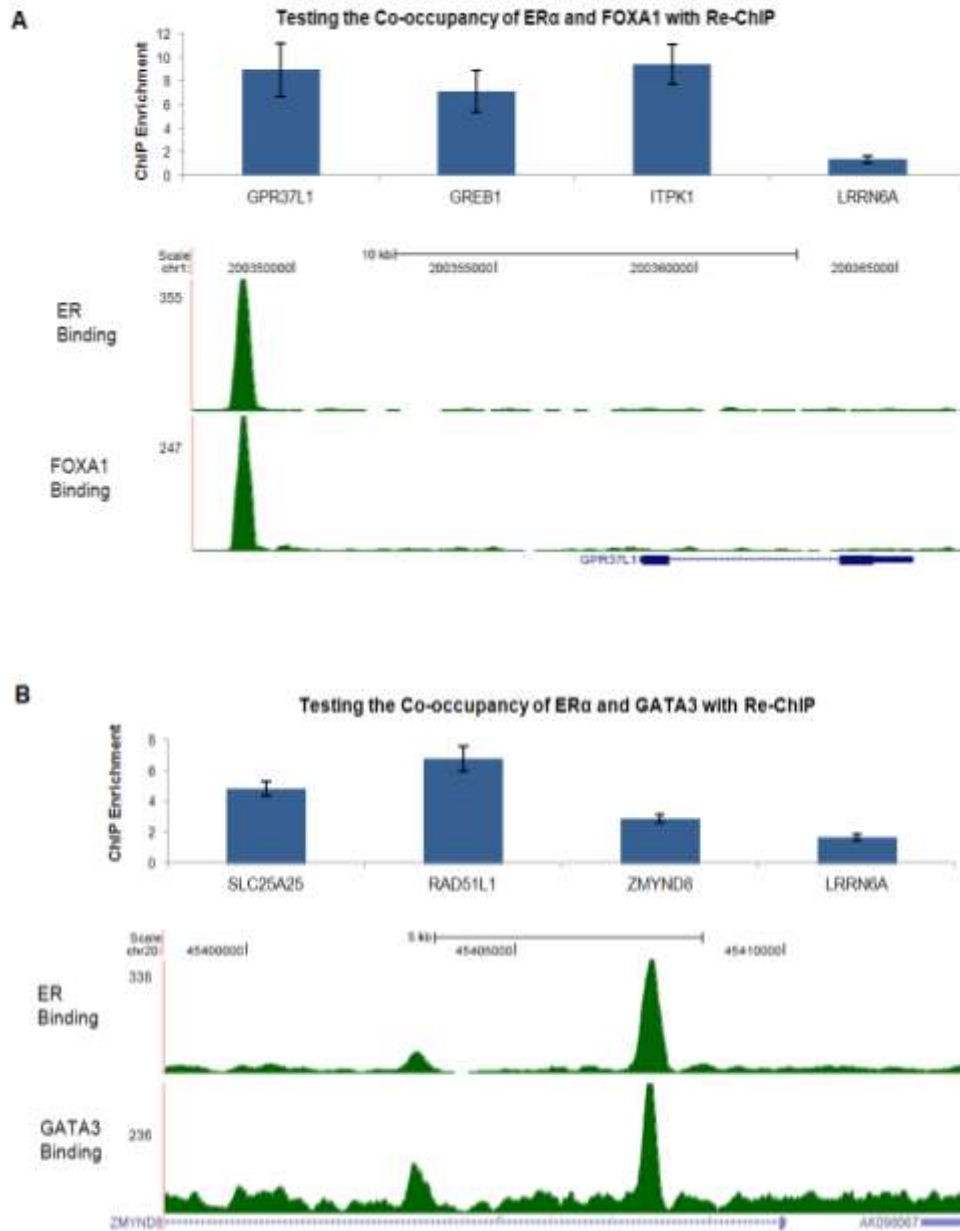
Figure 31. The co-occupancy of ERα+FOXA1 (A) and ERα+GATA3 (B) to the target cis-regulatory regions as validated by sequential Re-ChIP assay. Genes nearby are used to label the peaks, and the tag densities around gene GPR37L1 with ERα+FOXA1 peaks and gene PRKCBP1 with ERα+GATA3 peaks are shown as examples. A site near LRRN6A gene with only unique ERα binding is recruited as a negative control for the sequential Re-ChIP assay. The ChIP enrichment was computed by comparing to the IgG pull-down control. Means of at least two independent experiments are compared and standard errors are shown.

3.4 Progressive recruitment of ERα and FOXA1 to the cis-regulatory elements

Given the diversity of coregulatory complexes and their mutually exclusive mode of interaction with the TFs, it is essential to establish whether the series of TF regulation occur randomly or in a sequential fashion. The elegant study by Gannon and colleagues (Metivier R *et al*, 2003) presented an ordered pattern of recruitment and release of cohorts of regulatory complexes by examining the ERα-mediated transcriptional regulation to the *pS2* promoter in breast cancer cells. Hence, it is fascinating to depict the enthralling picture of the ordered recruitment of TFs to the regulated units.

It was previously described that FOXA1 is a pioneering factor characterized by this sequence of events: FOXA1 binds to the condensed chromatin in the absence of E_2 and opens the chromatin to facilitate the ERα binding upon E_2 stimulation (Carroll *et al*, 2005; Hurtado A *et al*, 2011). In addition, it was reported that FOXA1 bound at ERα sites prior to ligand stimulation followed by diminished FOXA1 occupancy after E_2 exposure potentially through displacement by the activated ERα (Carroll *et al*, 2005). However, we observed both an increase in the number of FOXA1 binding sites (9337 to 15852; Figure 21) and in the average level of occupancy at each site after ligand stimulation (on average there were 2.58 reads per peak per million reads in the FOXA1 ChIP-seq library after E_2 stimulation, compared to 1.74 reads per peak per million reads before E_2 stimulation).

We found that 37% (7196/19412) of the ERα binding sites after E_2 stimulation were co-occupied by FOXA1 where 25% (508/1990) of the ERα binding sites were co-occupied by FOXA1 in the absence of ligand (Figure 32). This is in accordance to the earlier

observations that FOXA1 is preferentially associated with E_2-bound ERα (Zhao *et al*, 2001).

If FOXA1 were a true pioneering factor, FOXA1 occupancy would be present in a significant percentage of ERα bound sites prior to estradiol exposure. However, our data revealed that only 11% (2218/19412; Figure 31) of E_2-induced ERα sites are occupied by FOXA1 prior to ligand exposure. When we eliminate the number of basal ERα-bound sites before E_2 stimulation, the percentage of FOXA1 sites that can recruit ERα is 19% (=(130+1598)/9337). This means that FOXA1 is a potential pioneering factor to recruit only a subset of ERα binding sites.

Interestingly, view from a different prospective, after excluding the number of basal FOXA1-bound sites before E_2 stimulation, the percentage of ERα sites that can recruit FOXA1 after E_2 exposure is 29% (=(573+9)/1990; Figure 31). Thus, ERα can also "pioneer" a site for FOXA1 as efficiently as the converse even though, because of a much larger starting denominator (9337 sites vs. 1990), it appears that FOXA1 is a better pioneering factor.
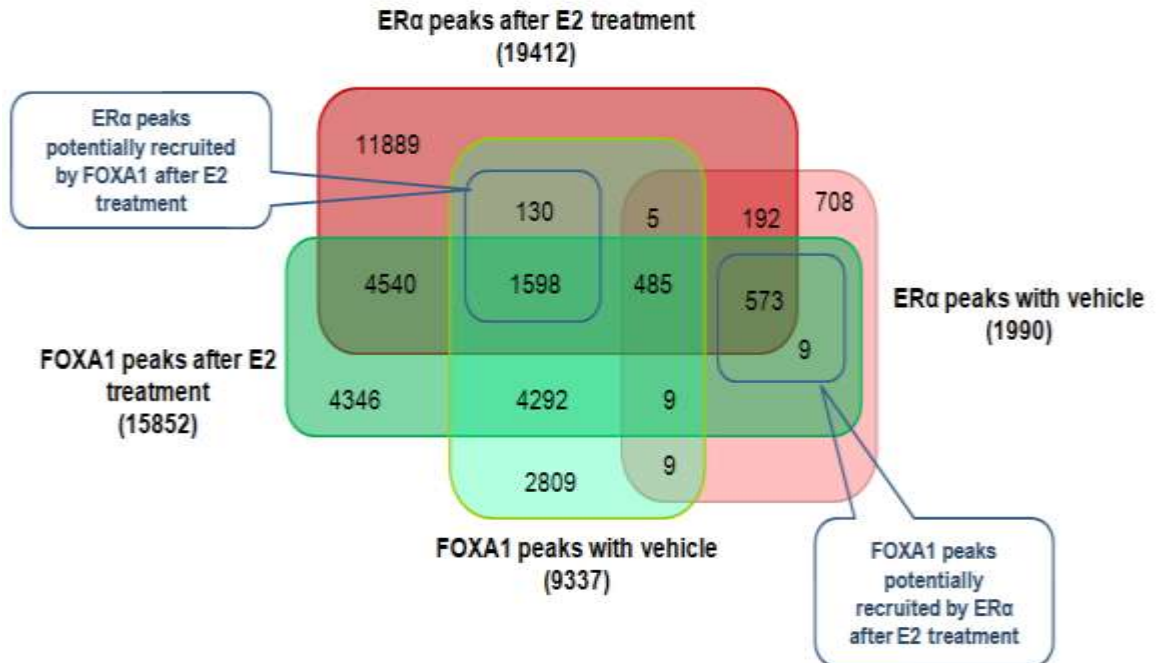
Figure 32. The progressive recruitment of ERα and FOXA1 to the cis-regulatory elements. Venn diagram of ERα binding sites with vehicle, ERα binding sites after E_2 treatment, FOXA1 binding sites with vehicle, and FOXA1 binding sites after E_2 treatment. We found a total of 1728 (130+1598) ligand-stimulated ERα binding sites that could be potentially pioneered by FOXA1. Conversely, a total of 582 (573+9) FOXA1 binding sites found in E_2-treated condition could be recruited by ERα. This suggests that both ERα and FOXA1 could be the recruiting factor for one another.

To confirm this ChIP-seq based observation, we synchronized the cells with α-amanitin treatment followed by E_2 stimulation and performed ChIP-qPCR over time on the nucleus lysates. The DNA content of the synchronized cells was assessed by cell cycle analysis (Figure 33). In the sites where FOXA1 functions as the recruiting factor, we observed enrichment of FOXA1 occupancy as early as 5 minutes upon E_2 stimulation, followed by progressively increasing ERα occupancy at the later time points (e.g., after 10-15minutes, example in Figure 34A). Conversely, in those sites where ERα functions as the recruiting factor, we show high ERα occupancy at early time points followed by increasing FOXA1 occupancy at the later time points (example in Figure 34B). Thus, we show that ERα and FOXA1 can equally function as recruiting factors for the other.

A

B
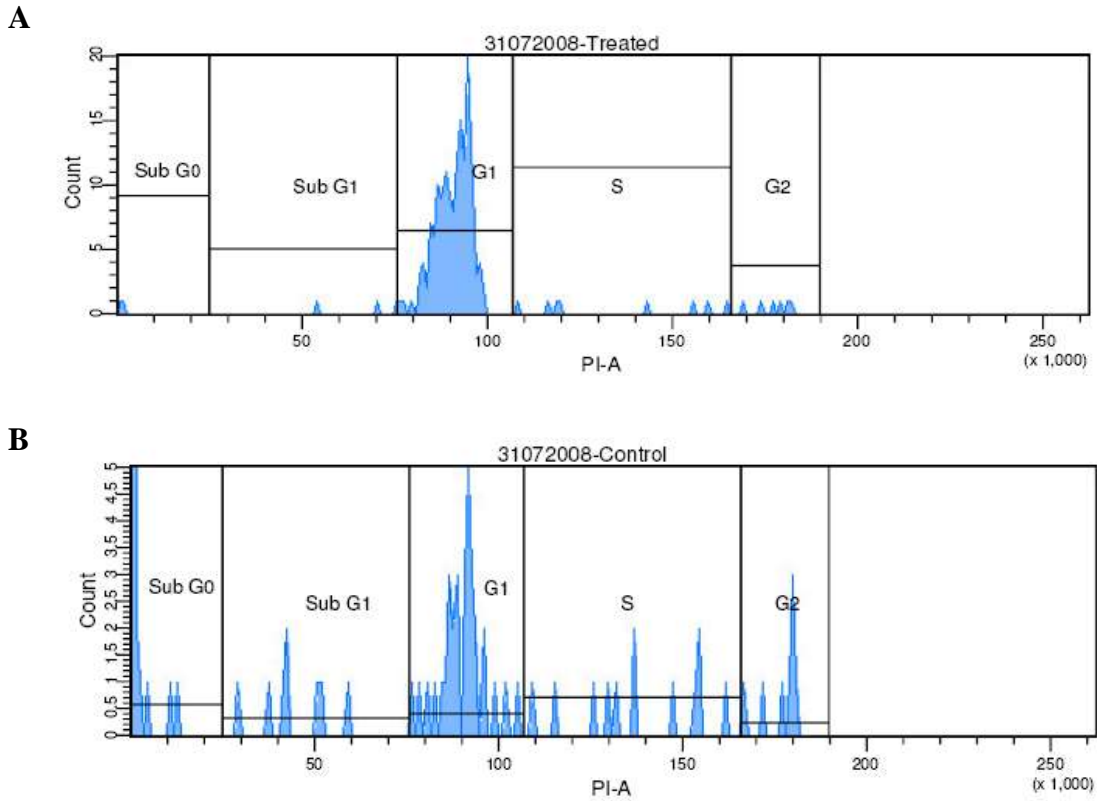
Figure 33. The synchronized MCF-7 cells (A) treated with α-amanitin for 2 hours before subjecting to 10nM E$_2$ treatment. The synchronized cells were analyzed with FACS (Beckton) revealing that these cells were synchronized to a single G1 phase as compared to non-synchronized control cells without α-amanitin treatment (B). The nucleus lysates were harvested at a 5-minutes interval.

94
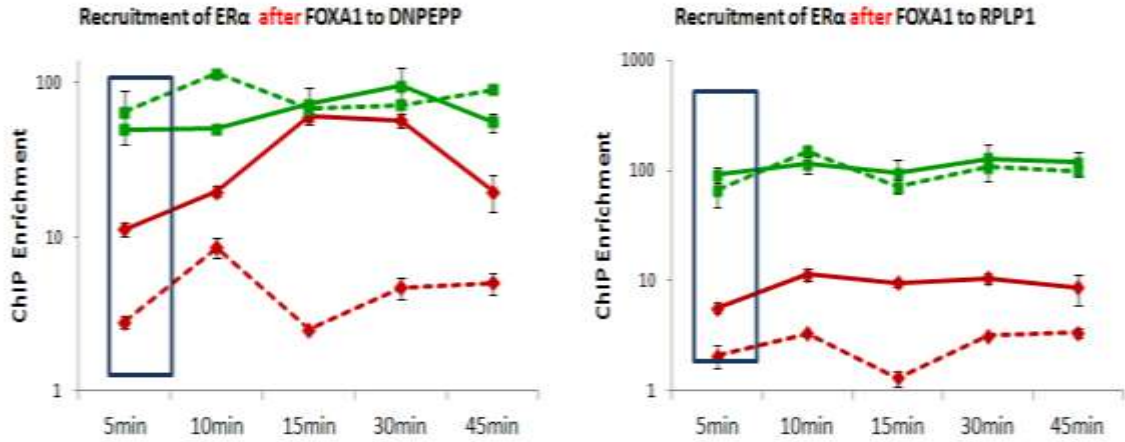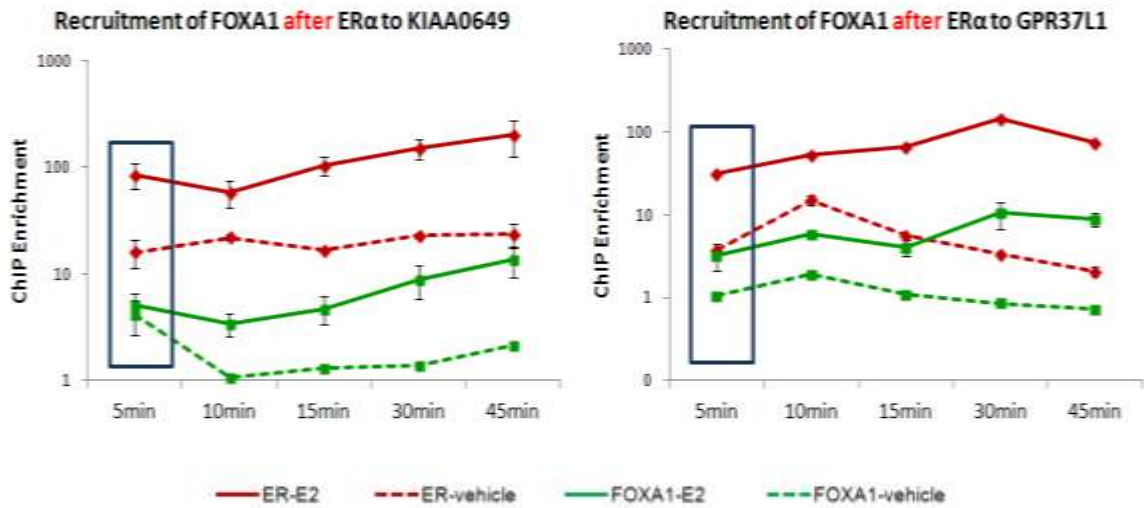
Figure 34. (A)The recruitment of ERα after FOXA1 binding in the synchronized MCF-7 cells upon E$_2$ stimulation as validated by ChIP-qPCR in various time points. Means of two independent experiments with consistent reproducible results are compared and standard errors are shown. (B) The recruitment of FOXA1 after ERα binding in the synchronized MCF-7 cells upon E$_2$ stimulation as validated by ChIP-qPCR in various time points. Means of two independent experiments with consistent reproducible results are compared and standard errors are shown.

3.5 The formation of enhanceosome consisting of ERα, FOXA1 and GATA3 in breast cancer cells

We have observed the co-localization of ERα, FOXA1 and GATA3 at genomic sites after ligand stimulation. We then wished to assess the dynamics of this recruitment by the three TFs in response to E_2 stimulation. First, we grouped the different subsets of binding sites before E_2 treatment as ERα unique, FOXA1 unique, GATA3 unique, ERα+FOXA1 overlap, ERα+GATA3 overlap, FOXA1+GATA3 overlap and ERα+FOXA1+GATA3 overlap sites as shown in the Venn diagram of Figure 21. We investigated how these different subsets of TF bindings clustered after E_2 stimulation. We found (Figure 35) a dramatic shift of single and double TF binding sites to sites occupied by the three TFs: more than 89% of vehicle-treated ERα+FOXA1 overlap sites, 86% of ERα+GATA3 overlap sites, 30% of ERα unique site, and 28% of the FOXA1+GATA3 overlap sites were shifted to the ERα+FOXA1+GATA3 overlap sites in response to E_2 induction. By contrast, the FOXA1 unique and GATA3 unique sites (before ligand) showed little to no shift to the conjoint three factor occupancy state after E_2 treatment. This suggests that estradiol activation induces the recruitment of FOXA1 and GATA3 with ERα at ERα binding sites.
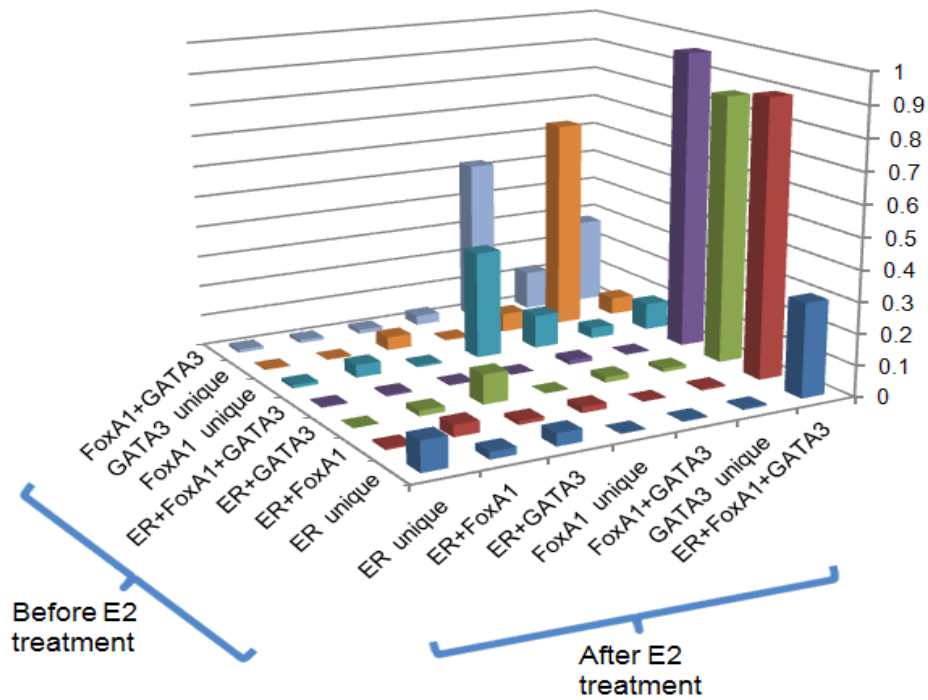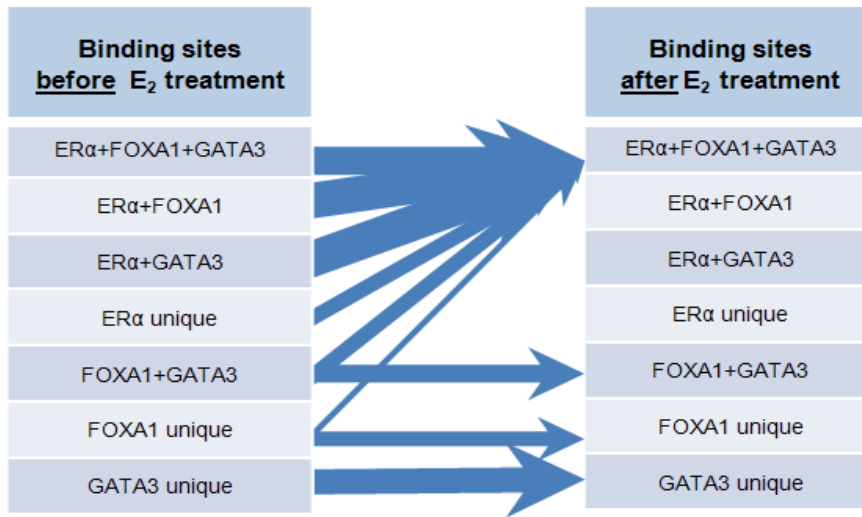
Figure 35. The dynamics of TFs binding before and after E_2 stimulation. The different categories of ERα, FOXA1 and GATA3 binding sites before E_2 stimulation will converge to ERα+FOXA1+GATA3 overlapped binding sites after E_2 stimulation.

It has been previously shown that the functional utility of an ERα binding site is higher when these sites are marked by specific and quantitative chromatin signatures: functionally active sites have higher ERα occupancy, more open chromatin, and more likely to show p300 and RNA polymerase II (RNA Pol II) occupancy. Figure 36-37 shows that the normalized tag profiles of ERα, FOXA1 and GATA3 at the binding sites are strongly enriched after E_2 treatment for all the above marks with the ERα+FOXA1+GATA3 overlapping conjoint sites having the highest tag occupancy profile above all other co-localized categories. In addition, the triple TF overlap sites show the greatest occupancy of each individual TF.
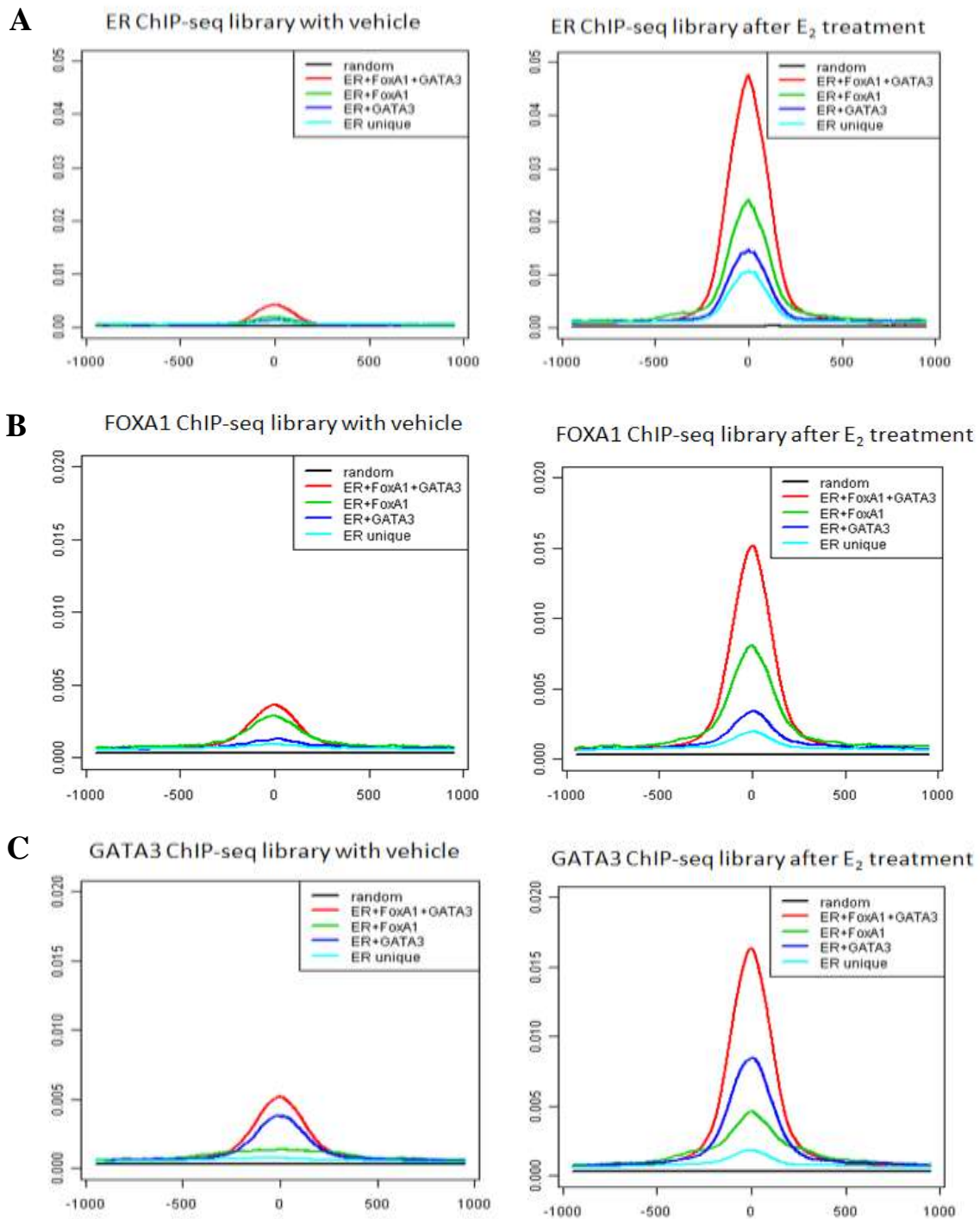
Figure 36. The tabulation of ERα unique, ERα+GATA3, ERα+FOXA1, ERα+FOXA1+GATA3 sites with the (A) ERα occupancy (B) FOXA1 occupancy, (C) GATA3 occupancy before and after E_2 stimulation.

99

In most instances, compacted chromatin is permissive to direct binding by TFs. One of the strategies adopted by TFs to overcome chromatin barrier is to recruit proteins such as p300 that can directly remodel chromatin structures *in vivo*. The p300 coactivator possesses intrinsic histone acetyltransferase (HATs) activity capable of modifying the chromatin organization, leading to chromatin decondensation and facilitate transcriptional initiation (Heintzman *et al*, 2007). The p300 enrichment is also commonly found at the enhancer regions. We observed ERα+FOXA1+GATA3 overlap sites have the highest p300 coactivator occupancy (Figure 37A).

In addition to ERα binding and targeted chromatin modification, the initiation of transcription at estrogen-responsive promoters also requires the recruitment of RNA Pol II. Previously, we have determined that RNA Pol II co-binding at ERα binding sites is related to distant interactions linking the enhancer sites with the transcription start sites (Fullwood *et al*, 2009). In Figure 37B, we show that ERα+FOXA1+GATA3 overlap sites have the highest RNA Pol II occupancy with ERα+FOXA1 double overlap sites following closely.

Chromatin is a well-known obstacle to transcription as it controls DNA accessibility, which directly impacts on the activation of transcriptional machinery. The degree of chromatin compaction is intimately related to its functionality and these can be identified using formaldehyde-assisted isolation of regulatory elements (FAIRE) that allows for enrichment of nucleosome-depleted genomic regions when cross-linked chromatin is subjected to phenol-chloroform extraction (Boyle AP. *et al*, 2008). Understanding how the chromatin landscape is shaped is essential to decipher the spatial and temporal fine-

tuning of ER transcriptional programs. When we assessed the chromatin state with FAIRE (Joseph *et al*, 2010), we found that ERα+FOXA1+GATA3 and ERα+FOXA1 overlap sites have the highest association with chromatin opening (Figure 37C). This suggests that these triple overlap sites (ERα+FOXA1+GATA3) are potentially the most active enhancers affecting ERα transcriptional regulation and that there appears to be a hierarchy of associative effect: FOXA1 contributing the most to ERα enhancer function and GATA3 being less impactful (Figure 37C).

Next, we investigated the distribution of the triple overlap/ enhanceosome sites in the ERα binding dataset. Interestingly, we found that majority of the enhanceosome sites (57%) present in the 1st quartile of ERα binding sites with highest ERα binding intensity, while only 7% of the enhanceosome sites reside in the 4th quartile of ERα binding sites with lowest binding intensity (Figure 38).
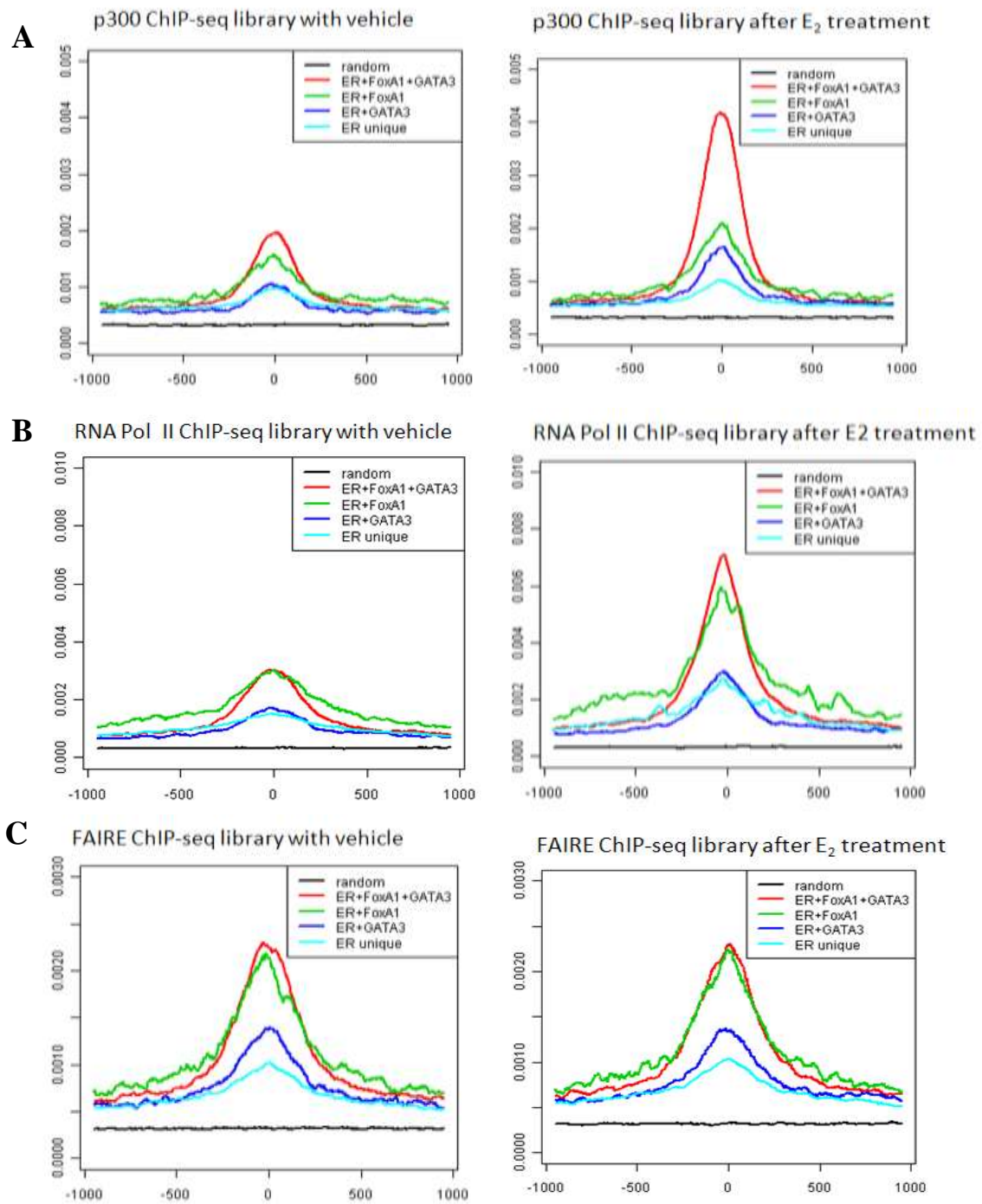
Figure 37. (A)The enhanced p300 coactivator recruitment to enhanceosome sites after E_2 stimulation. (B) The highest association of RNA Pol II recruitment with enhanceosome sites after E_2 stimulation. (C) The enhanceosome is correlated with chromatin opening as measured by FAIRE after E_2 stimulation.

Distribution of ER+GATA3+FOXA1 overlap sites in different quartiles of ER sites

- 1st quartile
- 2nd quartile
- 3rd quartile
- 4th quartile

57%
25%
11%
7%
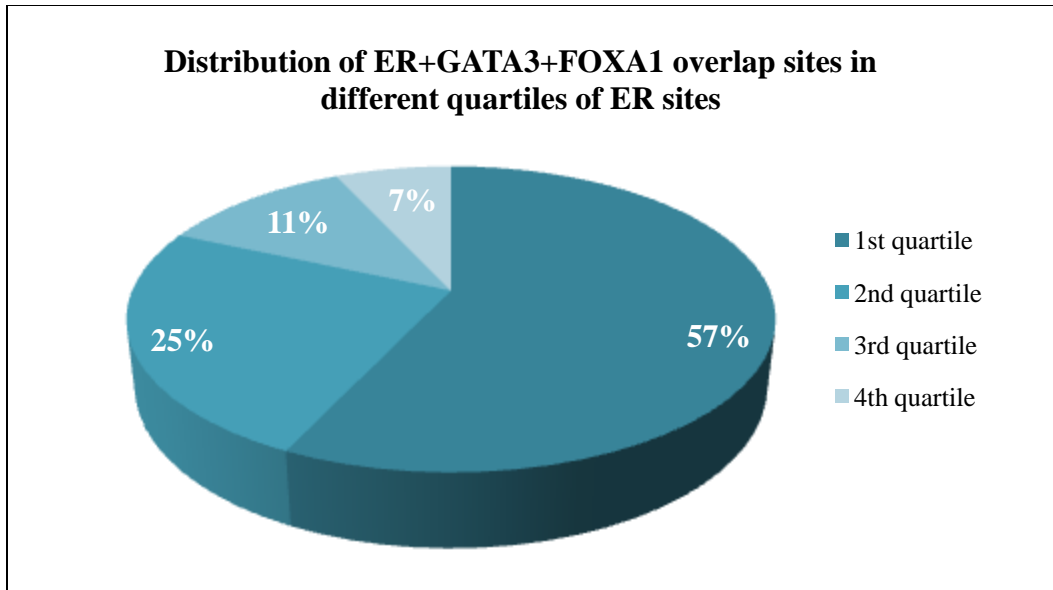
Figure 38. The distribution of enhanceosome consists of ERα+FOXA1+GATA3 in the ~19k ERα binding sites in MCF-7 cells. Majority of the enhanceosomes were found in the top quartile ERα binding sites with the highest binding intensity (1st quartile).

This observation raises the question whether the enhanceosome impact is solely driven by the strong ERα binding instead of the cooperative regulation by ERα+FOXA1+GATA3 TFs. In order to interrogate this, we separated the 1st quartile of ERα binding sites of highest binding intensity into two categories based on the presence or absence of triple overlap/ enhanceosome sites. We found the enhanceosome impact is independent of ERα occupancy intensity since the triple overlap/enhanceosome sites (ERα+FOXA1+GATA3) bear the marks for optimal enhancer with the highest individual TF, p300 and RNA Pol II occupancy while the non-enhanceosome sites have less association with the enhancer marks though all these sites were from the top quartile of ERα sites of highest binding intensity (Figure 39). Hence, we concluded that the enhanceosome impact is indeed driven by the cooperative regulation by ERα, FOXA1 and GATA3.

Next, we asked the question whether the enhanceosome found in MCF-7 cells can be generalized in other breast cancer cells. In our recent work (Joseph *et al*, 2010), we have provided the evidence that the common ERα sites found in both MCF-7 cells and another breast cancer cell line, T47D cells represented the sites with the highest association with ERE score, H3K4me1, FOXA1 occupancy and chromatin opening. Interestingly, we observed that as much as 52% of the common ERα sites found in MCF-7 and T47D cells exhibit enriched FOXA1 and GATA3 binding (Figure 40A-B). Furthermore, the common ERα sites overlapped with FOXA1 and GATA3 bindings displayed higher ERα binding intensity as compared to non-overlapped common ERα sites (Figure 40C).

Recent publication by Carroll and co-workers has revealed a list of core ERα binding events that is present in the ERα-positive breast tumors and absent in the ERα-negative

104

tumors (Ross-Innes C.S. *et al*, 2012). In order to address whether the genomic impact by enhanceosome in MCF-7 cells would be relevant in primary tumors, we overlapped our enhanceosome binding sites with the core ERα binding sites found in ERα-positive breast tumors. Remarkably, we found that more than 82% of these core ERα binding events in breast tumor overlapped with our enhanceosome sites (Figure 41). This suggests that the impact of enhanceosome beyond MCF-7 cell model. Moreover, when we overlapped our enhanceosome sites with the ERα binding events that were differentially enriched in tumors with good or poor prognosis/ distant metastasis, we found that there is better overlap between the MCF-7 enhanceosome sites with the poor prognosis ER binding events. This could be explained by the fact that MCF-7 cell line was derived from metastatic mammary tumor and hence it is more similar to tumor with poor prognosis/ distant metastasis.
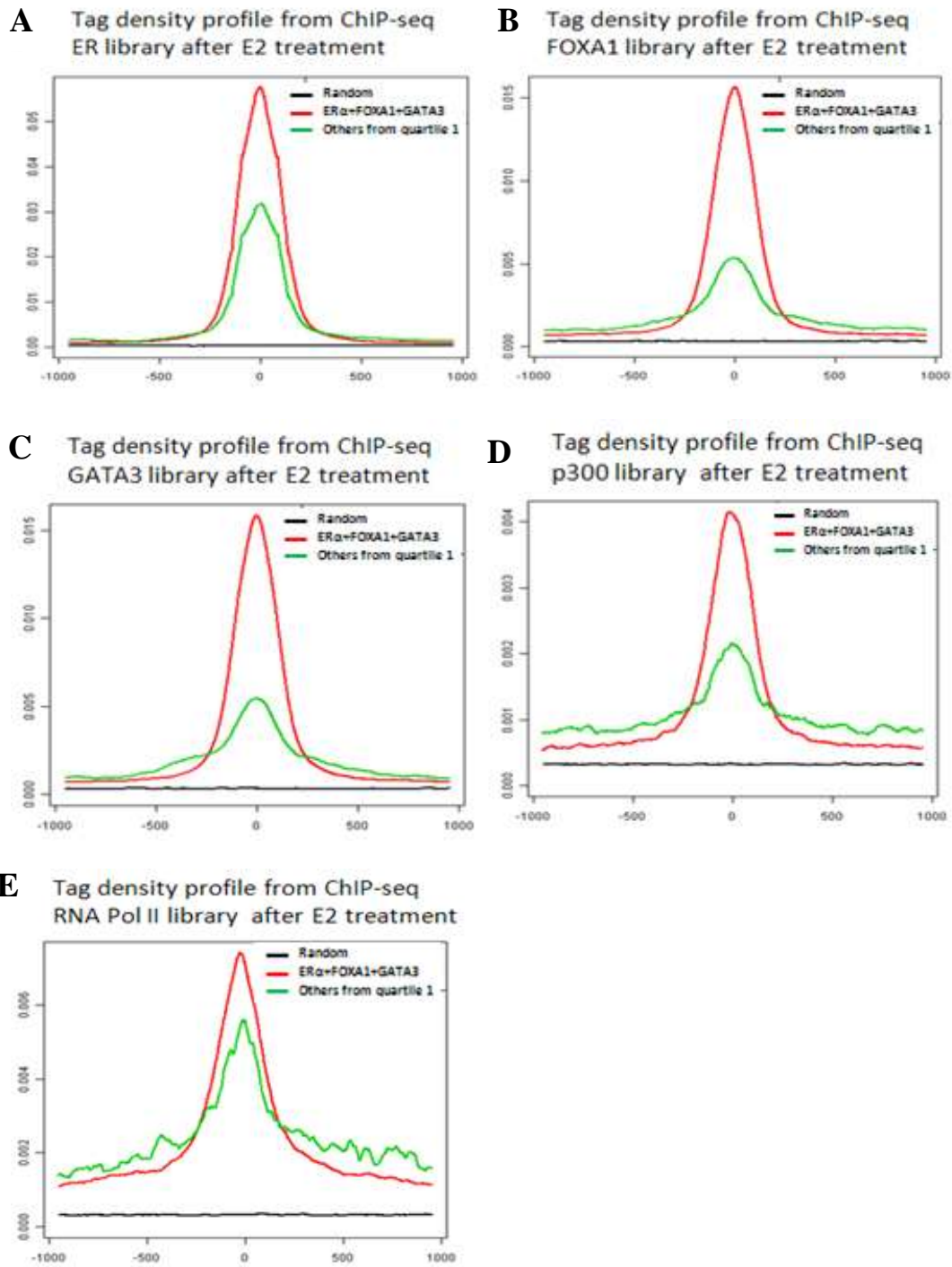
Figure 39. The comparison between enhanceosome (red) vs non-enhanceosome (green) from the top quartile ERα sites for the occupancy of ERα, FOXA1 and GATA3 (A-C); (D) p300 recruitment and (E) RNA Pol II recruitment.
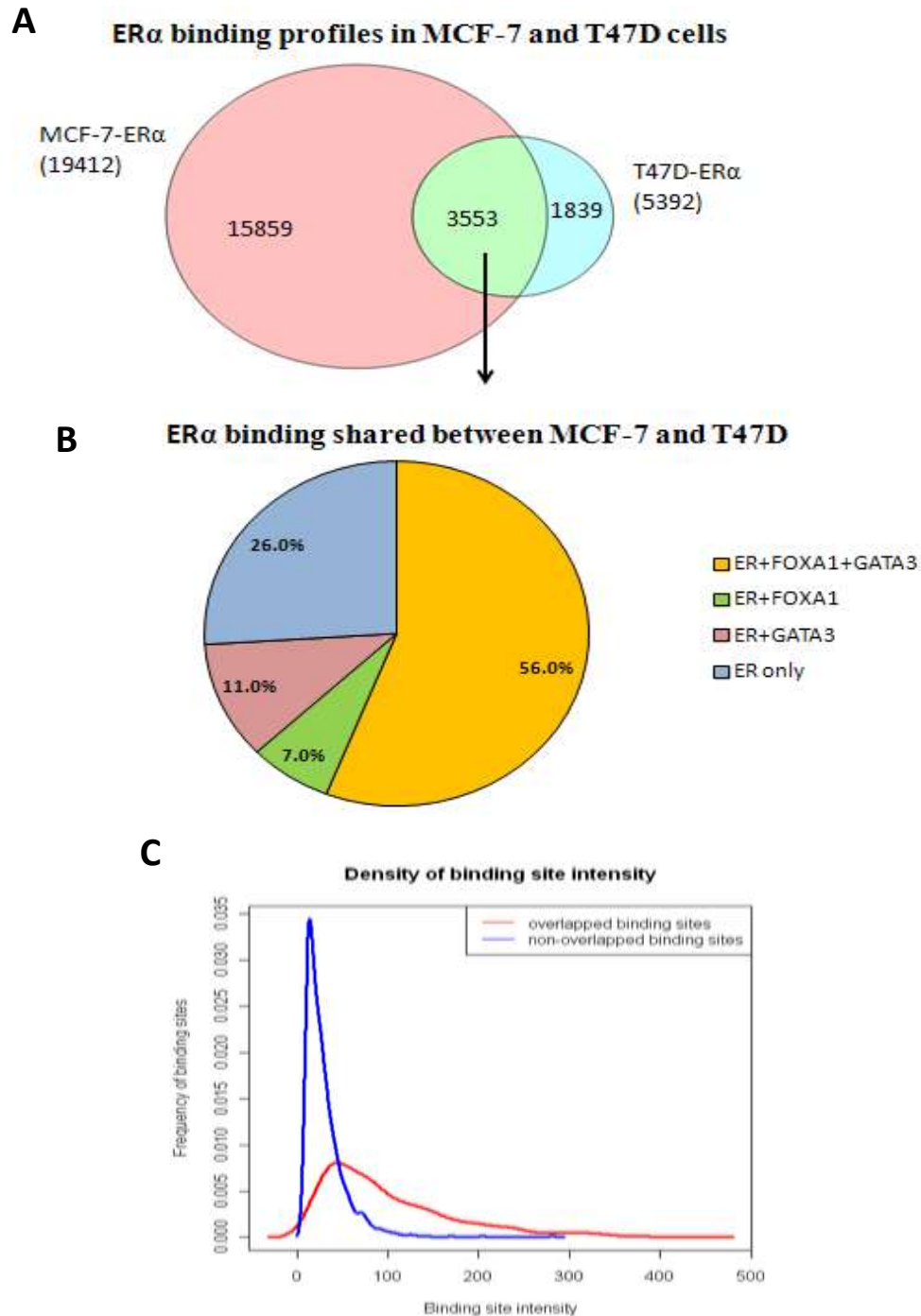
Figure 40. The enhanceosome found in MCF-7 cells can be generalized in another breast cancer cell line, T47D. (A) There was 52% of the common ERα sites shared between MCF-7 and T47D cells. (B) More than half of the common ERα sites exhibited enriched FOXA1 and GATA3 bindings. (C) The common ERα sites overlapped with FOXA1 and GATA3 bindings displayed higher ERα binding intensity as compared to non-overlapped common ERα sites.

107
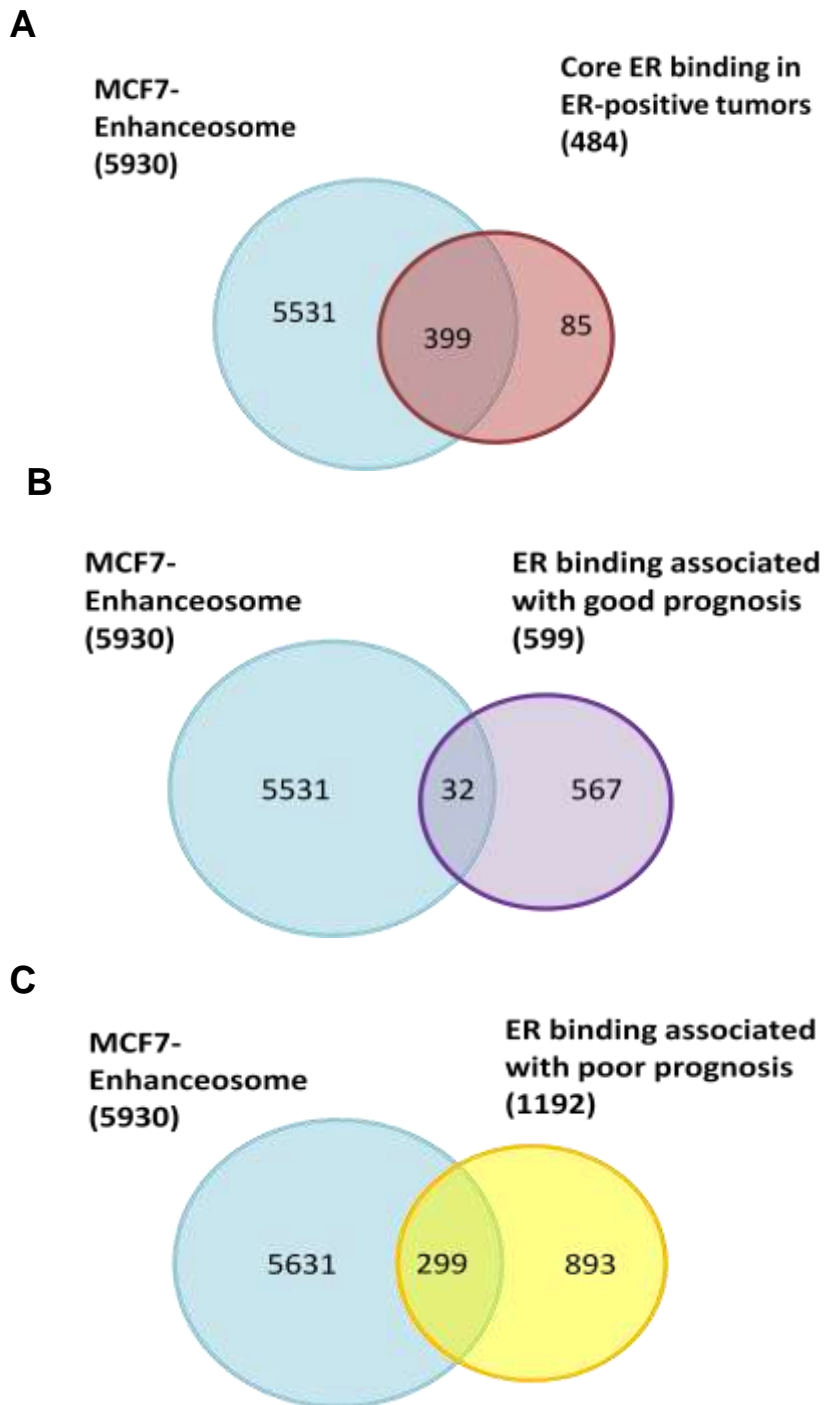
Figure 41. (A) The overlap between MCF-7 enhanceosome sites with core ER binding sites in ERα-positive breast tumors. (B) The overlap between MCF-7 enhanceosome sites with ER binding events enriched in tumors with good prognosis. (C) There is more overlap between MCF-7 enhanceosome sites with ER binding events enriched in tumors with poor prognosis.

108

3.6 The enhanceosome is associated with three-dimensional ERα-regulated long-range interactome

The chromatin landscape is defined as the combination of parameters that influence its function that include the higher-order folding and related three-dimensional organization. It is known that TFs can interact through long-range chromatin interactions to regulate the transcriptional networks. Recently, a new method known as Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) has been developed (Fullwood *et al*, 2009; Li *et al*, 2010) to characterize the long-range chromatin looping mediated by ERα in MCF-7 cell line. After re-analyzing the ERα ChIA-PET data, we observed that 81% of the 5067 interaction clusters have at least one anchor region (an ERα binding site associated with a distant chromatin interaction forming at least one loop) characterized by ERα+FOXA1+GATA3 co-localization. Furthermore, the interaction clusters from ChIA-PET can form complex clusters that organize the local genomic region into multiple loops. These complex interaction clusters also demarcate the most significantly ERα regulated genes (Fullwood *et al*, 2009). Of the 5067 ERα mediated long-range interaction clusters, 4500 clusters are involved in complex interaction clusters, and 567 clusters are involved in duplex interaction clusters (Figure 42). 88% of the complex interaction clusters are associated with ERα+FOXA1+GATA3 overlapped binding sites, while 51% of the duplex interaction clusters have the support of ERα+FOXA1+GATA3 overlapped binding sites (Figure 43).
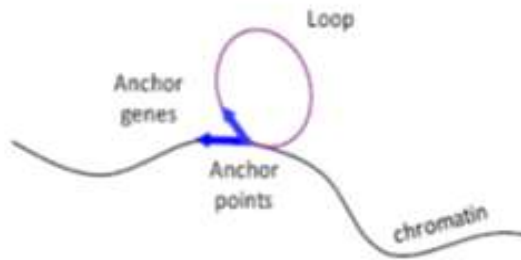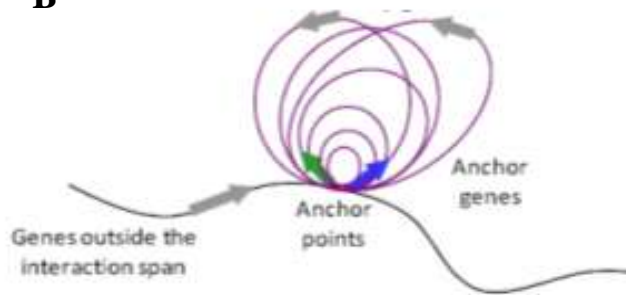
Figure 42. The illustration of an example of (A) duplex and (B) complex interaction is shown (modified from (Fullwood *et al*, 2009).
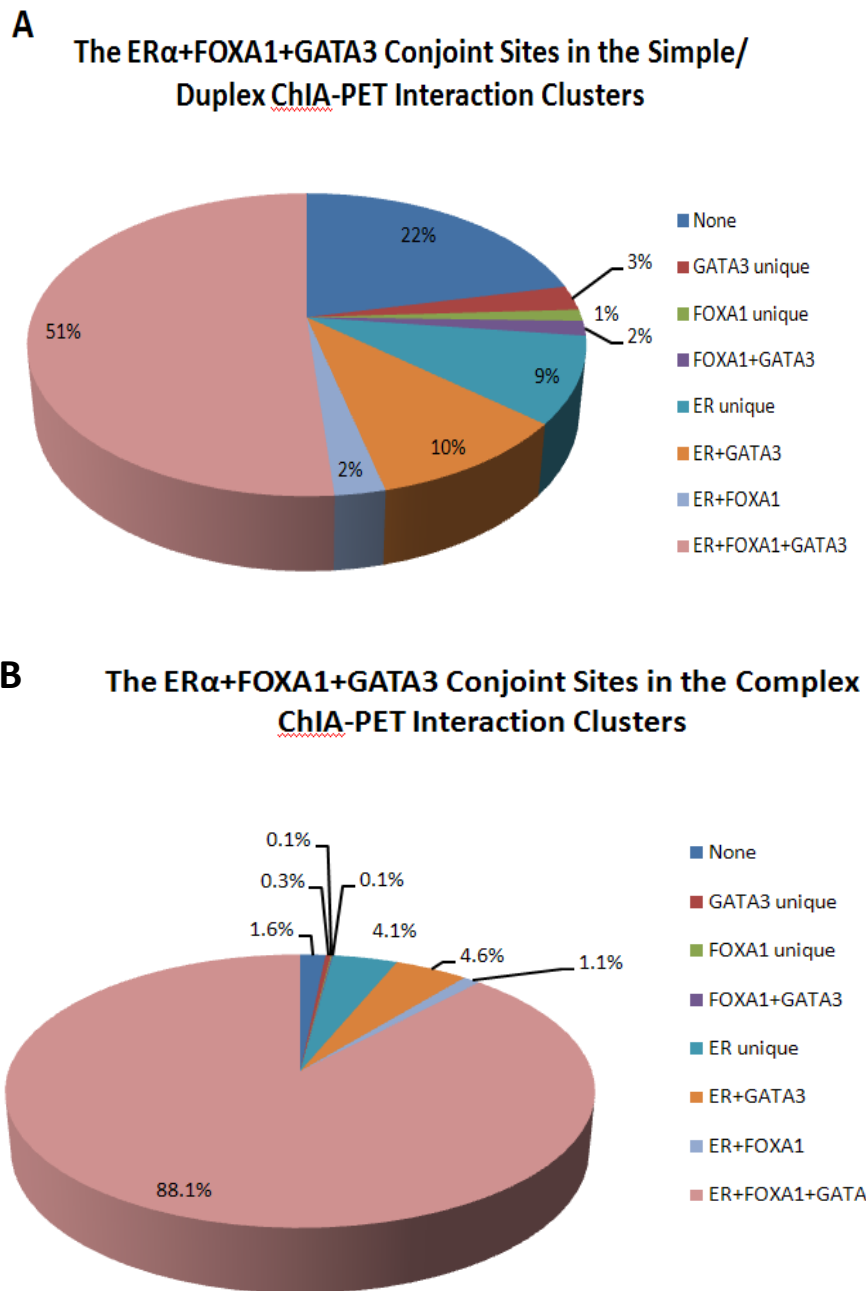
Figure 43. Majority of the ERα interactomes display ERα+FOXA1+GATA3 co-localization. (A) The distribution of simple/ duplex ChIA-PET interaction clusters with ERα+FOXA1+GATA3 conjoint bindings. (B) The complex ChIA-PET interaction clusters with ERα+FOXA1+GATA3 conjoint bindings. We found that the enhanceosome has better association with complex ERα interactome, indicating that interactome of higher complexity requires multiple TFs.

111

Since complex interaction clusters mark genes most responsive to E_2 as compared to duplex clusters, these data support the notion that the presence of the ERα, FOXA1, and GATA3 putative enhanceosome is associated with genes that are most responsive to E_2. A specific example of the ERα mediated long-range interactions involving conjoint ERα, FOXA1 and GATA3 binding sites around the highly E_2 responsive *GREB1* gene is shown in Figure 44. These triple TF conjoint binding sites are highly represented at the sites involved in frequent long range chromatin interactions spanning 50 kb.

Figure 44. An example of E$_2$-regulated gene with complex ERα interactome was shown to have co-localization of ERα, FOXA1 and GATA3 bindings.

3.7 The impact of ERα+FOXA1+GATA3 enhanceosome in regulating E$_2$-responsive genes

We have provided evidence that the clustering of ERα, FOXA1 and GATA3 at ERα binding sites is associated with chromatin characteristics of the most active ERα enhancers. To assess the effect of this enhanceosome presence on direct gene regulation, we performed a detailed microarray expression analysis to determine E$_2$-responsive genes in MCF-7 cells. We found a total of 653 up-regulated and 1249 down-regulated genes in response to E$_2$ stimulation. We assigned a specific known gene to a binding site occupied by any combination of the three TFs if the peak of each TF category is the nearest and within 20kb of TSS of that E$_2$ responsive gene. Figure 45A shows that, for the genes with ERα, FOXA1 and GATA3 peaks or any combination, the biggest proportion of either up- or down-regulated genes are from genes with adjacent ERα+FOXA1+GATA3 conjoint binding sites within 20kb of their TSS (28% and 30% respectively). Since the different binding sites are present in the genome at different frequencies, the ratio of regulated vs. non-regulated genes for each binding site class can be used to normalize the differences. The proportion of up-regulated genes with ERα, FOXA1 and GATA3 conjoint peaks is 2.3 fold of the non-regulated genes with the same configuration. This is contrasted by genes adjacent to other combinations of ERα, FOXA1 and GATA3 binding which do not show significant changes as compared to non-regulated genes. The only exception is the proportion of genes close to FOXA1+GATA3 co-bound sites which are associated with greater down-regulated genes. Despite this association, the percentage of down-regulated genes putatively controlled jointly by FOXA1 and GATA3 is relatively small.

114

Finally, the presence of the three TFs relative to a regulated gene may have two configurations: one where the ERα binding site has conjoint and therefore overlapping occupancy by all three TFs, and the other where the binding of the individual TFs are in proximity with each other and within 20kb of a gene, but the binding sites are not overlapping (Figure 45B).

When we analyzed the association of E_2-regulated genes with these two categories (overlapping and non-overlapping), we found that the predominant association is between the conjoint binding sites and regulated genes (Figure 45B). Our results imply that ERα-regulation of gene expression is closely linked to adjacency with sites that show conjoint binding with ERα, FOXA1 and GATA3 putatively forming an enhanceosome.
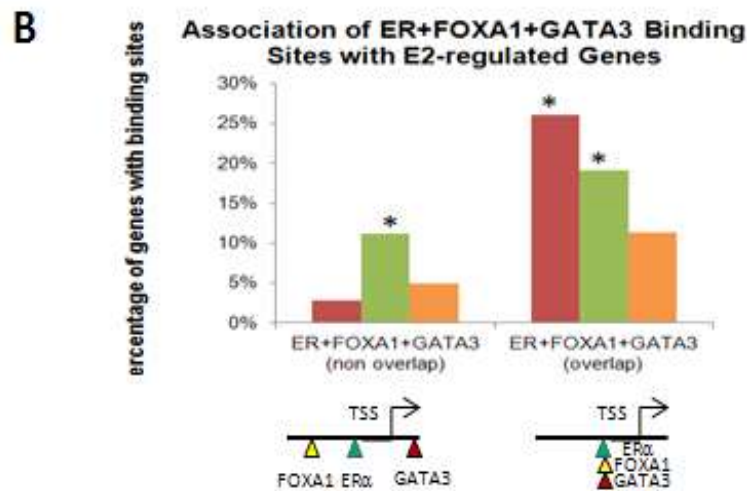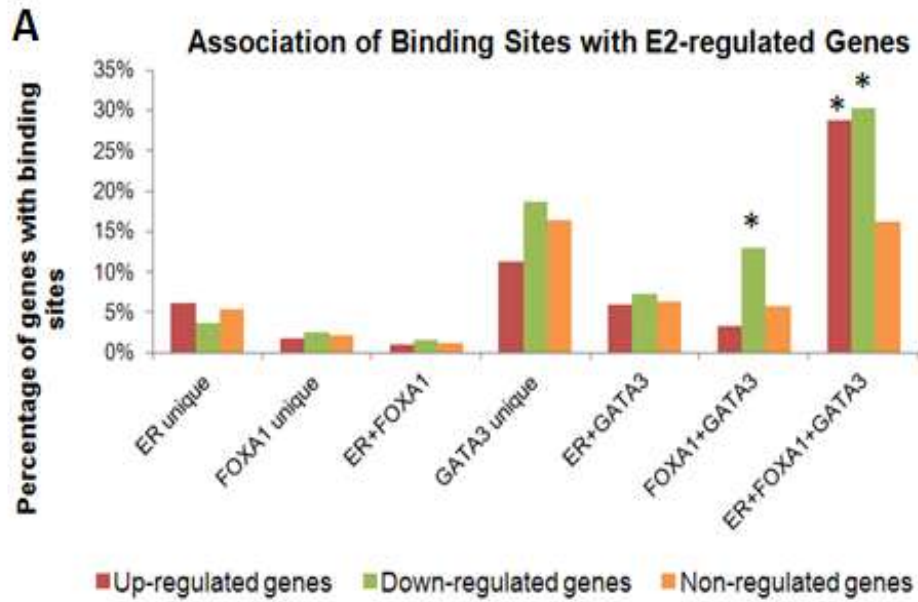
Figure 45. The classification of ERα+FOXA1+GATA3 conjoint binding sites into two configurations: (A) non-overlapped ERα+FOXA1+GATA3 peaks and, (B) overlapped ERα+FOXA1+GATA3 peaks within 20kb of the TSS.

Using Gene Ontology (GO) analysis (Thomas *et al*, 2003), we sought to further ascertain the importance of genes in proximity with enhanceosome binding as compared to binding of the individual TF components. We found that only genes associated with ERα+FOXA1+GATA3 binding have significant association (with P values up to 6.7E-26) with specific biology processes known to be involved in ERα signaling (eg. signal transduction, cell proliferation), molecular function (eg. kinase, protein binding) and signaling pathways (eg. PDGF signaling pathway, inflammation mediated by chemokine and cytokine signaling pathway) (Tables 7-9). Thus, the identification of the ERα enhanceosome associated genes allows for the identification of a "core" set of ERα regulated genes that are strongly associated with the cognate cellular functions previously known for ERα. This result implies that the cooperativity of the enhanceosome components not only facilitate binding of the activators to DNA but also position them to create an interface for the transcription machinery.

Biological Process	genes with ER+FOXA1+GATA3 overlapped peaks (P-value)	genes with ER unique peaks (P-value)	genes with FOXA1 unique peaks (P-value)	genes with GATA3 unique peaks (P-value)
cellular process	6.71E-26	1.14E-05	N. S.	1.61E-07
cell communication	1.48E-19	8.06E-08	N. S.	9.69E-05
developmental process	2.08E-19	N. S.	N. S.	N. S.
signal transduction	1.58E-17	6.34E-08	N. S.	N. S.
system development	7.60E-13	N. S.	N. S.	N. S.
intracellular signaling cascade	1.75E-12	N. S.	N. S.	N. S.
system process	3.44E-12	N. S.	N. S.	N. S.
cell motion	3.61E-12	N. S.	N. S.	N. S.
ectoderm development	1.02E-11	N. S.	N. S.	N. S.
cellular component morphogenesis	3.96E-11	N. S.	N. S.	N. S.
anatomical structure morphogenesis	3.96E-11	N. S.	N. S.	N. S.
transport	3.96E-11	N. S.	N. S.	N. S.
neurological system process	4.17E-11	N. S.	N. S.	N. S.
nervous system development	1.69E-10	5.95E-05	N. S.	N. S.
cellular component organization	1.20E-09	N. S.	N. S.	N. S.
cell adhesion	2.31E-09	N. S.	N. S.	N. S.
mesoderm development	4.05E-09	N. S.	N. S.	N. S.
sensory perception	1.06E-08	N. S.	N. S.	N. S.

Note: N.S. means "not significant with p-value larger than 1E-04"

Table 7. Gene ontology analysis of genes associated with different categories of ERα, FOXA1, GATA3 bindings. The genes associated with ERα+FOXA1+GATA3 overlapped binding sites have significant functions, compared to genes only with individual unique ERα, FOXA1 and GATA3 binding.

Pathway	Genes with ERα+FOXA1+ GATA3 overlapped peaks (p-value)	Genes with ERα unique peaks (p-value)	Genes with FOXA1 unique peaks (p-value)	Genes with GATA3 unique peaks (p-value)
PDGF signaling	1.41E-07	N.S.	N.S.	N.S.
Heterotrimeric G-protein signaling	3.92E-05	N.S.	N.S.	N.S.
Inflammation mediated by chemokine and cytokine signaling	2.81E-05	N.S.	N.S.	N.S.
Histamine H1 receptor mediated signaling	1.86E-05	N.S.	N.S.	N.S.

Note: N.S. means "not significant with p-value larger than 1E-04"

Table 8. The Gene Ontology analysis (Pathway) for genes with ERα+FOXA1+GATA3 bindings as compared to genes with ERα unique, FOXA1 unique or GATA3 unique binding.

Molecular Function	Genes with ERα+FOXA1+GATA3 overlapped peaks (p-value)	Genes with ERα unique peaks (p-value)	Genes with FOXA1 unique peaks (p-value)	Genes with GATA3 unique peaks (p-value)
Protein binding	1.59E-15	3.90E-07	N.S	7.29E-06
Small GTPase regular activity	6.29E-15	N.S	N.S	N.S
Enzyme regulator activity	2.51E-12	N.S	N.S	1.55E-05
Catalytic activity	7.68E-09	5.30E-08	N.S	3.13E-11
Guanyl-nucleotide exchange factor activity	9.82E-09	N.S	N.S	N.S
Binding	1.09E-08	N.S	N.S	1.76E-06
Transporter activity	6.08E-08	N.S	N.S	N.S
Transmembrane transporter activity	1.23E-07	N.S	N.S	N.S
Transferase activity	7.96E-07	N.S	N.S	3.37E-07
Kinase activity	2.91E-06	N.S	N.S	8.11E-05
Structural constituent of cytoskeleton	8.41E-06	N.S	N.S	N.S
Calcium ion binding	1.71E-05	N.S	N.S	N.S

Note: N.S. means "not significant with p-value larger than 1E-04"

Table 9. The Gene Ontology analysis (Molecular Function) for genes with ERα+FOXA1+GATA3 bindings as compared to genes with ERα unique, FOXA1 unique or GATA3 unique binding.

To further validate that ERα+FOXA1+GATA3 co-binding represents an optimal configuration for E_2-mediated transcriptional activation, we have performed luciferase reporter assays on *GREB1* locus that actively engages ERα enhanceosome sites in gene regulation (Figure 46A). We cloned the promoter region of *GREB1* that includes a ERα+FOXA1+GATA3 enhanceosome binding site into the pGL4-luciferase reporter construct and then transfected *GREB1*-luciferase promoter construct into ER negative MDA-MB-231 cells, followed by transfection and over-expression of ERα, FOXA1 and/or GATA3. The individual presence of FOXA1 and GATA3 or combination of both only produced subtle changes to the *GREB1* luciferase activity, demonstrating that the presence of FOXA1 and GATA3 alone or combination of both do not activate the transcription of *GREB1* gene (Figure 46B). The presence of ERα induced the *GREB1* luciferase activity to ~246% (as compare to the control construct). The combination of ERα+FOXA1 and ERα+GATA3 has increased the luciferase activity to ~330% (an increment of 26-32%). Interestingly, the assemblage of ERα+FOXA1+GATA3 provided the optimal ER responsiveness to 370% representing an additional 12-14% increment. This suggests that ERα provides the fundamental gene regulatory module but that FOXA1 and GATA3 incrementally improve ERα regulated transcriptional induction.
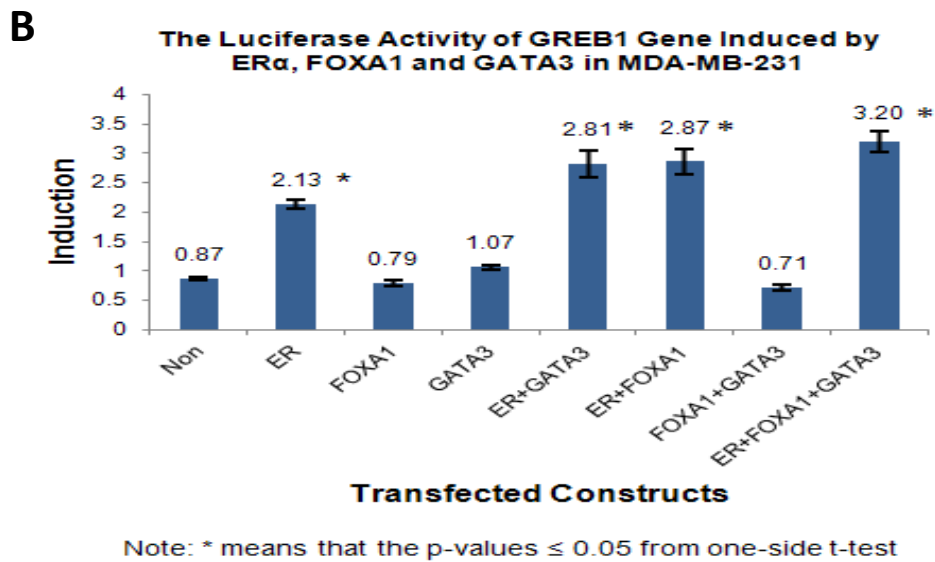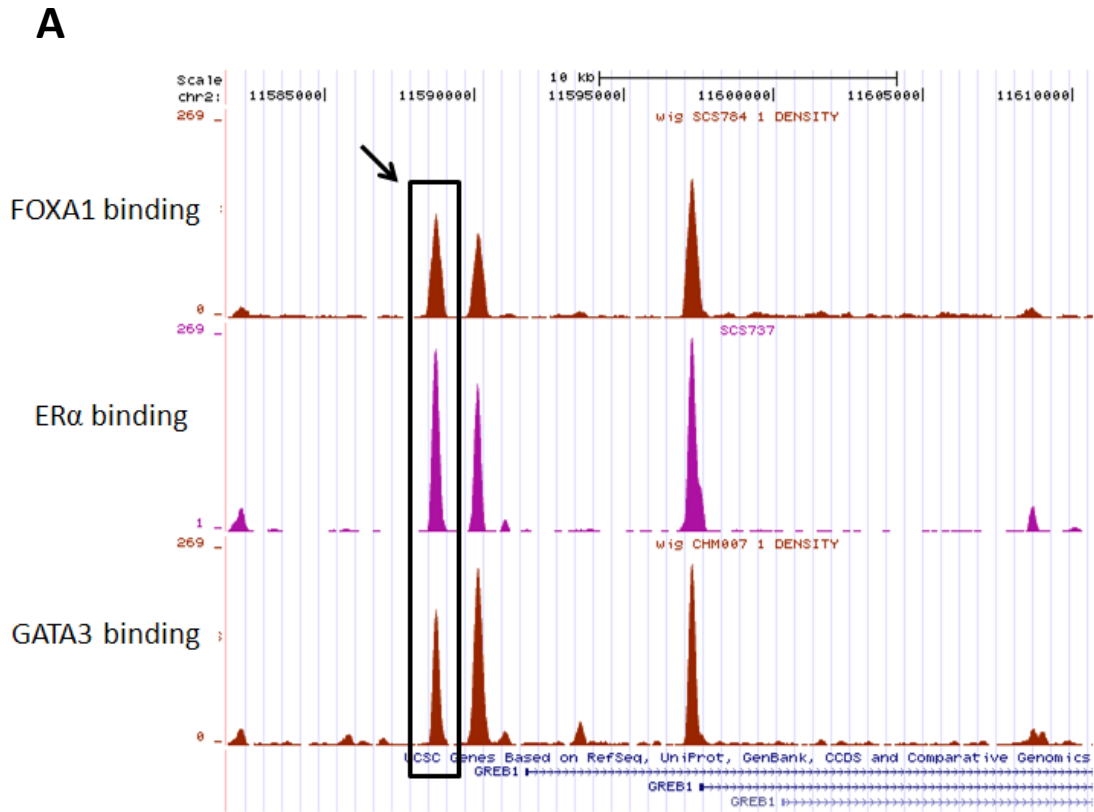
Figure 46. (A) The *GREB1* locus that is engaged with enhanceosome recruitment. (B)The presence of ERα, FOXA1 and GATA3 has induced the luciferase activity of *GREB1* gene in MDA-MB-231 cells. The basal luciferase activity of *GREB1* in MDA-MB-231 cells is used as the control reference. Means of three independent experiments are compared and standard errors are shown.

Such artificial transfection reporter systems accentuate TF responses because of unnatural stoichiometries of the TFs. To further assess the interplay among ERα, FOXA1 and GATA3, we perturbed the binding of these TFs in MCF-7 cells through the site-directed mutagenesis assay and asked whether the loss of individual binding motifs would alter gene regulation under physiologic concentrations of the three TFs. Different *GREB1*-luciferase constructs with mutated ERE, FOXA1 or GATA3 motif at the specific ERα, FOXA1 and GATA3 binding sites were generated. The results revealed that individually mutated FOXA1 or GATA3 motif only imposed 25-30% loss of *GREB1* luciferase activity (Figure 47). Mutated ERE alone has repressed the luciferase measurement to ~50%. Interestingly, combinatorial ERE+FOXA1 and ERE+GATA3 mutation further reduced the luciferase activity by ~65-70% suggesting that the effects of the individual TFs on this putative enhanceosome are additive. Here, we can build a hierarchy of TFs control, showing that ERα accounts for 50% of the transcriptional control while FOXA1 and GATA3 individually account for another 20% transcriptional control at the *GREB1* gene regulatory locus.

Next, we asked if the enhanceosome impacts on gene regulation is merely driven by induced expression of ERα, FOXA1 and GATA3 upon E_2 stimulation. We performed western blot on MCF-7 lysates prior and after estrogenic stimulation, the data revealed unaltered expression of these TFs upon E_2 treatment (Figure 48). This suggests that the enhanceosome impacts that we observed is not caused by the mass action of induced TFs expression, instead it is motivated by the activated and enhanced recruitment of TFs to the transcription hot spots.

The luciferase activity of *GREB1* gene repressed by mutations of ERα, FOXA1, and GATA3 in MCF-7
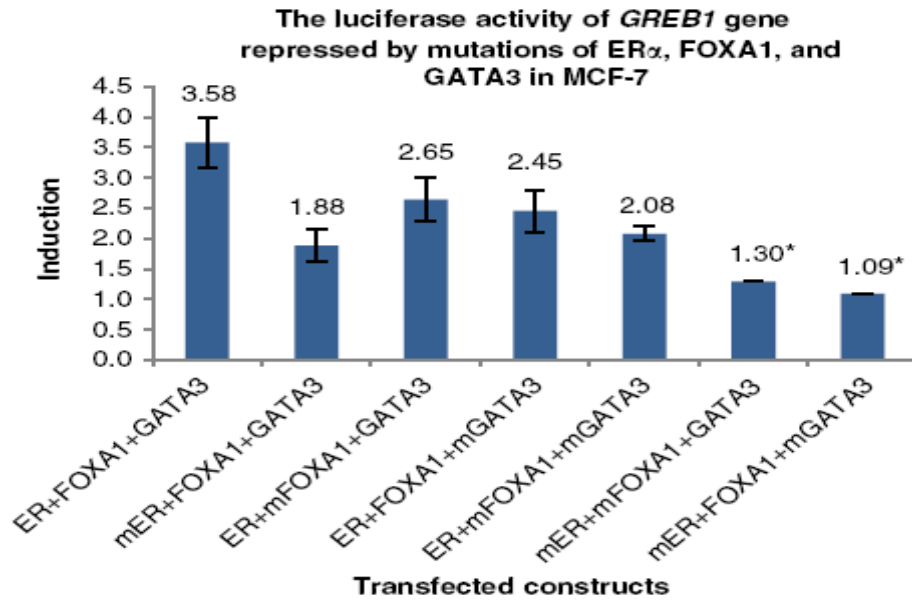
Note: * means that the p-values ≤ 0.05 from one-side t-test

Figure 47. The loss of FOXA1 and/or GATA3 bindings has reduced the luciferase activity of *GREB1* gene in MCF-7 cells. "mER", "mFOXA1" and "mGATA3" denote mutated ERE, FOXA1 and GATA3 motif sequences around their respective binding sites near the *GREB1* promoter. The basal luciferase activity of *GREB1* in MCF-7 wild-type cells is served as the control reference. Means of three independent experiments are compared and standard errors are shown.
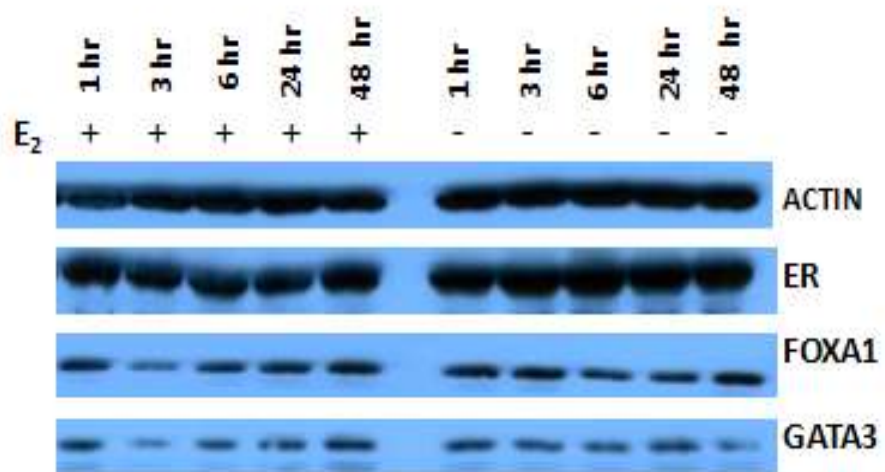
Figure 48. There were unaltered ERα, FOXA1 and GATA3 expression levels prior and after estrogen stimulation at various time points.

CHAPTER 4: FOXA1 AND GATA3 ARE ESSENTIAL COREGULATORS IN MEDIATING THE ERα-GROWTH RESPONSE

4.1 The co-expression of ERα, FOXA1 and GATA3 is required for estrogen-responsive growth

It is known that ERα is a ligand-activated TF that mediates the proliferative effects of E_2 in breast cancer cells. Garcia *et al.* (Garcia *et al*, 1992) showed inhibited growth in MDA-MB-231 cells with forced expression of ERα upon E_2 treatment. The rationale for these different outcomes has remained elusive. We hypothesize that the absence of critical coregulators such as FOXA1 and GATA3 are responsible for the ERα response cassette.

To test this hypothesis, we stably transfected the MDA-MB-231 cells with individual ERα, FOXA1, GATA3 or in combination. The induction of ERα, FOXA1 and GATA3 expressions following transfections were verified (Figure 49).

We assessed the cell proliferation in response to E_2 stimulation using two assays: WST-1 and cell count using Hoechst stain. In parallel to the reports by Garcia *et al.* (Garcia *et al*, 1992) and Wolf *et al*. (Wolf *et al*, 2007), we observed marginally inhibited growth in cells with forced expression of ERα and a greater inhibitory effect with forced expression of FOXA1. There was unaltered growth in cells with expression of GATA3. Coexpression of ERα and FOXA1; ERα and GATA3 exhibited inhibition of cell proliferation as compared to control cells. However, the coexpression of ERα together

with FOXA1 and GATA3 resulted in marked induction of cell proliferation under E_2 stimulation as assessed by either growth detection assays (Figure 50-51).

We have recapitulated this cellular reprogramming in another ERα-negative breast cancer cell line, BT-549 and observed similar growth inhibition in BT-549 cells expressing ERα and FOXA1 individually (Figure 52-53). We found minor induction of growth in GATA3-expressing BT-549 cells, however this growth was independent of E_2 stimulation. However, like MDA-MB-231 cells, we were able to induce E_2-dependent growth in ERα+FOXA1+GATA3-expressing BT-549 cells (Figure 54).

This suggests that only with the full activation of conjoint binding sites by the three TFs will the proliferative phenotype associated with ligand induced ERα be manifest. This further suggests that like induced pluripotent stem cells (iPS cells) only the combination of multiple factors (in this case, ERα, FOXA1 and GATA3) can transcriptionally reprogramme MDA-MB-231 and BT-549 cells to be estrogen responsive for growth. Our observation on the induced growth observed at 10 days after estrogen stimulation was akin to the iPS reprogramming, where the presence of reprogramming TFs and the time in culture has led to the accumulation and selection of novel genomic aberrations which were quantitatively of the same magnitude as those inflicted during the iPS reprogramming (Blasco MA *et al*, 2011).
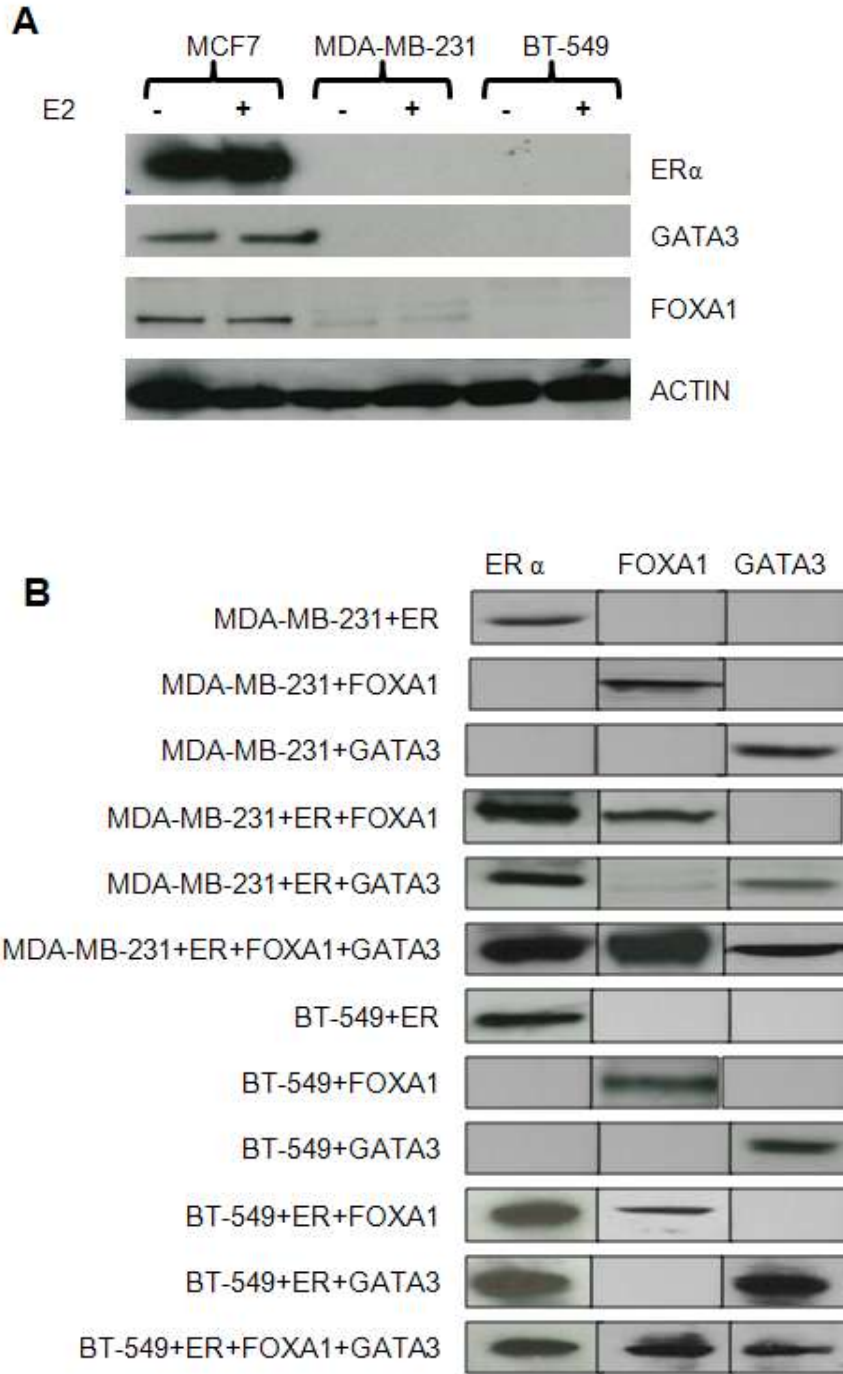
Figure 49. The expression of ERα, GATA3 and FOXA1 in ERα-positive MCF-7 cells and ERα-negative MDA-MB-231 and BT-549 cells prior to transfection experiment. (B) The expression of ERα, GATA3 and FOXA1 in the transfected MDA-MB-231 and BT-549 cells with various TF combinations.

128
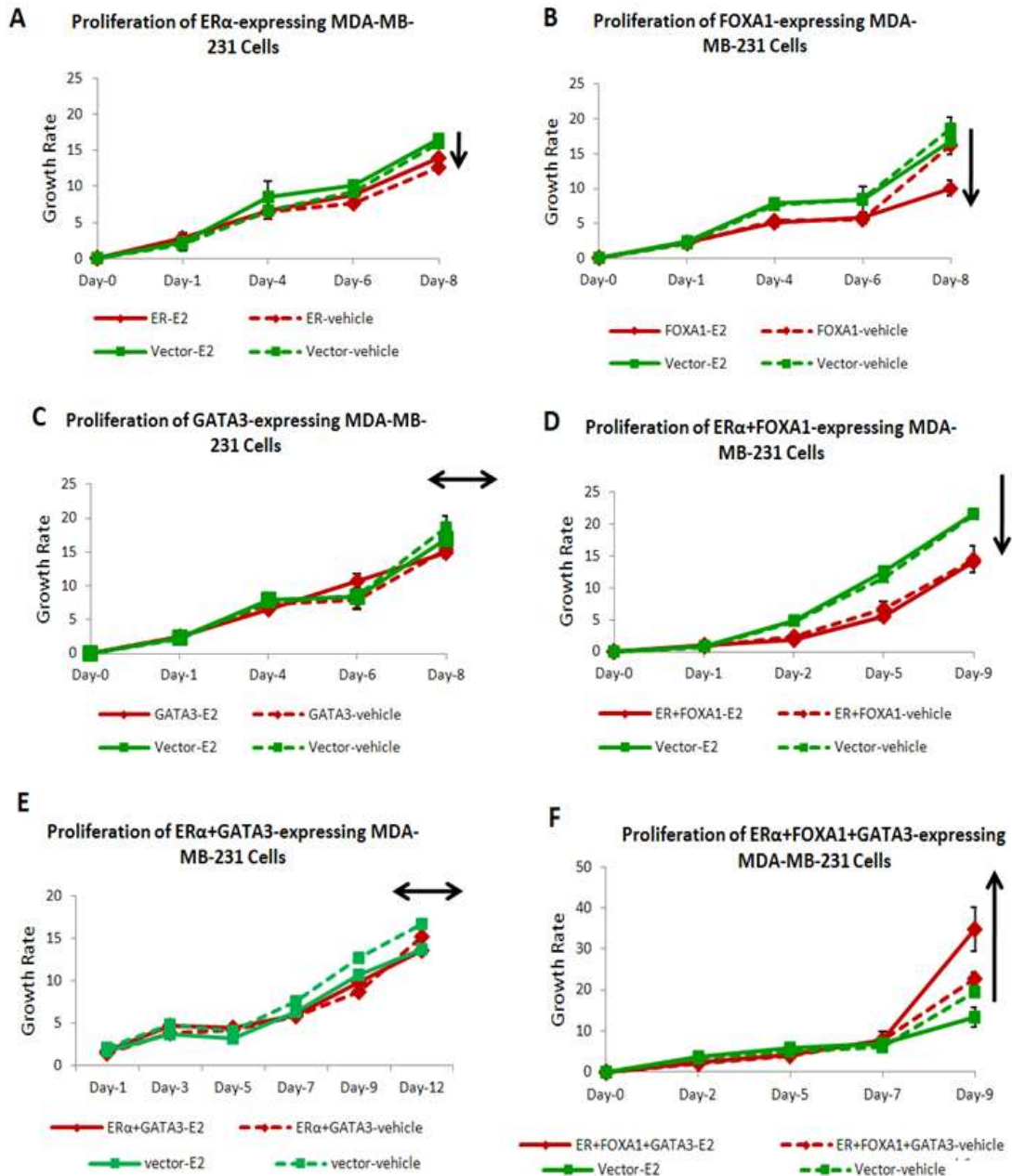
Figure 50. The proliferation of reprogrammed MDA-MB-231 cells assayed by WST-1. (A) The subtle inhibited growth of MDA-MB-231 cells with the transfection of ERα. (B) The inhibited growth of MDA-MB-231 cells with the transfection of FOXA1. (C) The unaltered growth of MDA-MB-231 cells with the transfection of GATA3. (D) The inhibited growth of MDA-MB-231 cells with co-transfection of ERα+FOXA1. (E) The unaltered growth of MDA-MB-231 cells with co-transfection of ERα+GATA3. (F) The induced cell proliferation in response to E2 stimulation in the MDA-MB-231 cells with the co-transfection of ERα, FOXA1 and GATA3 in combination. For every sub-figure, means and standard errors of three independent experiments are shown.

129

A Proliferation of ER-expressing MDA-MB-231 Cells
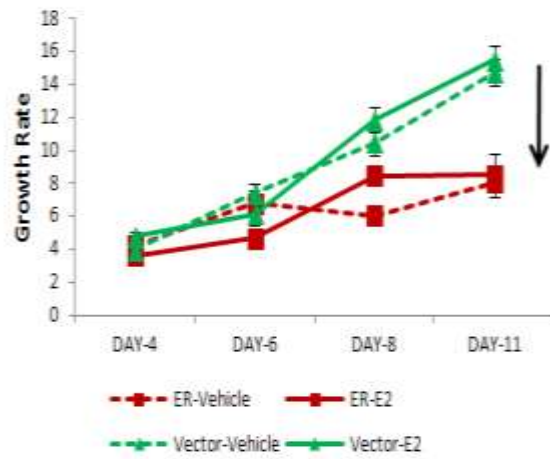
B Proliferation of FOXA1-expressing MDA-MB-231 Cells
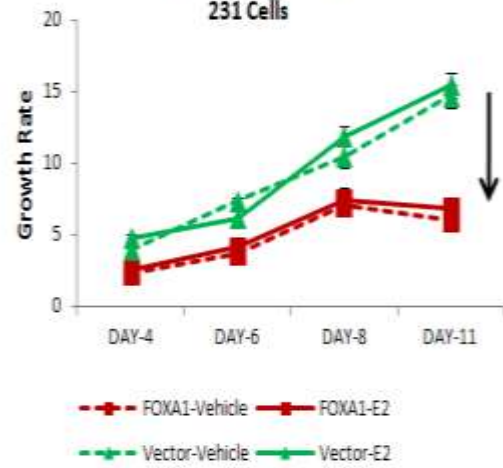
C Proliferation of GATA3-expressing MDA-MB-231 Cells
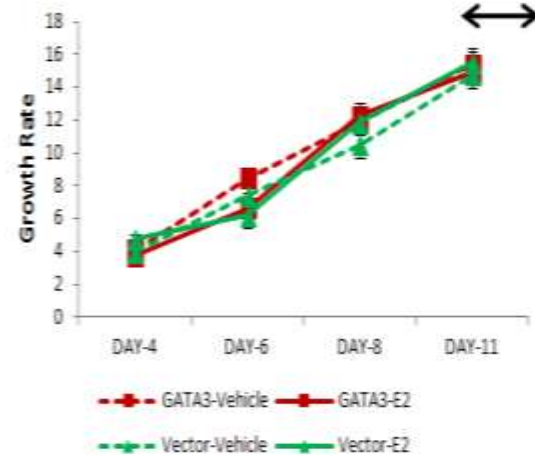
D Proliferation of ER+FOXA1-expressing MDA-MB-231 Cells

130
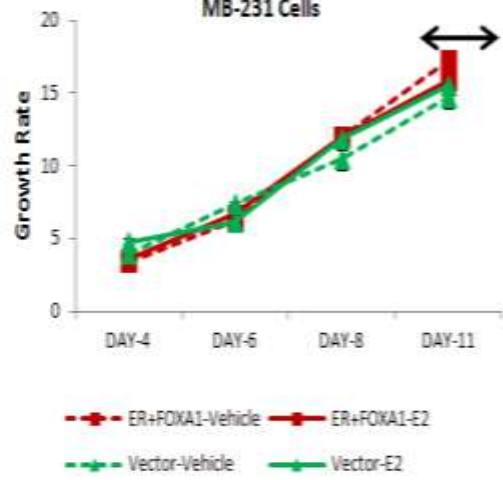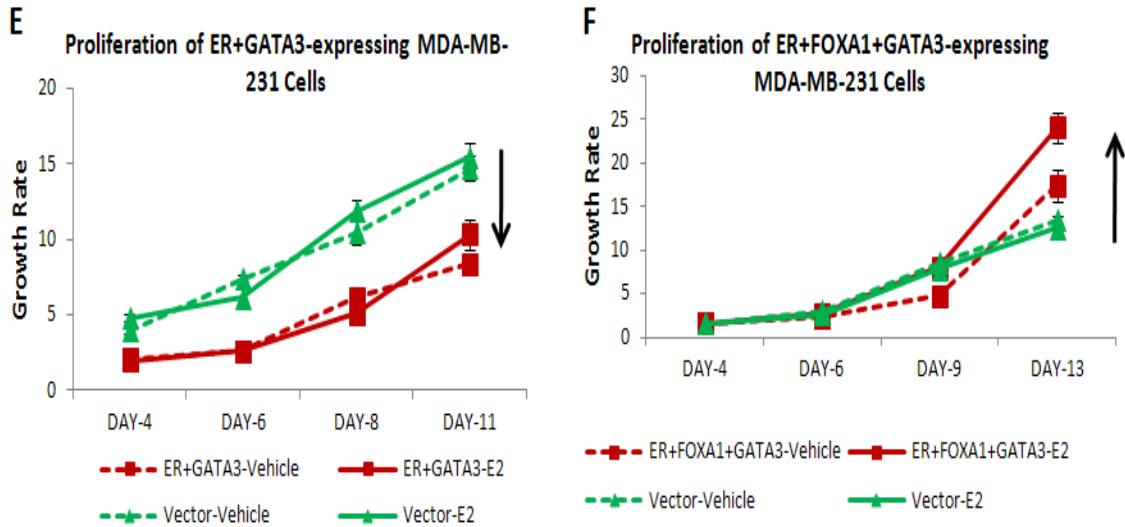
Figure 51. The proliferation of reprogrammed MDA-MB-231 cells assessed by cell number count using Hoechst stain. (A) The subtle inhibited growth of MDA-MB-231 cells with the transfection of ERα. (B) The inhibited growth of MDA-MB-231 cells with the transfection of FOXA1. (C) The unaltered growth of MDA-MB-231 cells with the transfection of GATA3. (D) The unaltered growth of MDA-MB-231 cells with co-transfection of ERα+FOXA1. (E) The inhibited growth of MDA-MB-231 cells with co-transfection of ERα+GATA3. (F) The induced cell proliferation in response to E2 stimulation in the MDA-MB-231 cells with the co-transfection of ERα, FOXA1 and GATA3 in combination. For every sub-figure, means and standard errors of three independent experiments are shown.

A Proliferation of ERα-expressing BT-549 cells

B Proliferation of FOXA1-expressing BT-549 Cells

C Proliferation of GATA3-expressing BT-549 Cells

D Proliferation of ERα+FOXA1-expressing BT-549 Cells

Figure 52. Recapitulating the reprogramming work in another ERα-negative BT-549 cell line assayed by WST-1. (A) The inhibited growth of BT-549 cells with the transfection of ERα. (B) The inhibited growth of BT-549 cells with the transfection of FOXA1. (C) The subtle induced growth of BT-549 cells with the transfection of GATA3. (D) The unaltered growth of BT-549 cells with co-transfection of ERα+FOXA1. (E) The unaltered growth of BT-549 cells with co-transfection of ERα+GATA3. (F) The induced cell proliferation in response to E2 stimulation in the BT-549 cells with the co-transfection of ERα, FOXA1 and GATA3 in combination. For every sub-figure, means and standard errors of three independent experiments are shown.

A The proliferation of ERα-expressing BT-549 Cells

- - ER-vehicle — ER-E2
- - VECTOR-vehicle — VECTOR-E2

B The Proliferation of FOXA1-expressing BT-549 Cells

- - FOXA1-vehicle — FOXA1-E2
- - VECTOR-vehicle — VECTOR-E2

C The Proliferation of GATA3-expressing BT-549 Cells

- - GATA3-vehicle — GATA3-E2
- - VECTOR-vehicle — VECTOR-E2

D The Proliferation of ERα+FOXA1-expressing BT-549 Cells

- - ER+FOXA1-vehicle — ER+FOXA1-E2
- - VECTOR-vehicle — VECTOR-E2

134
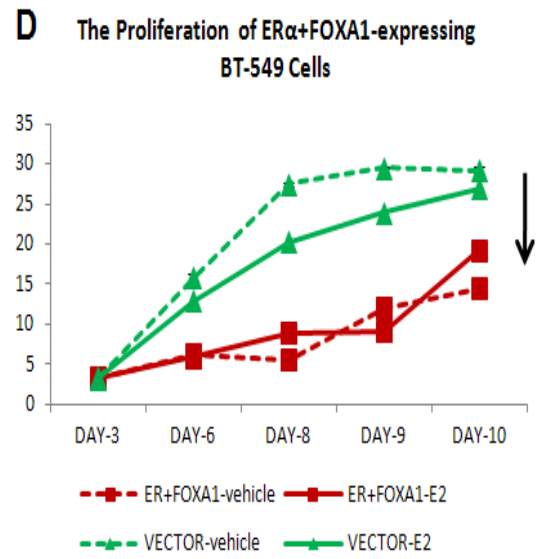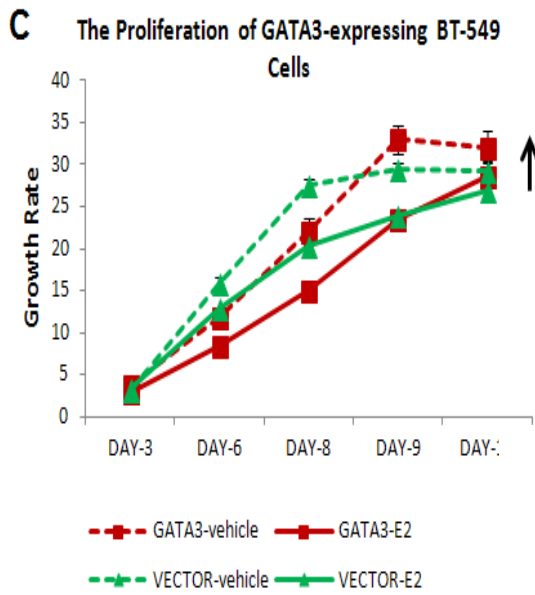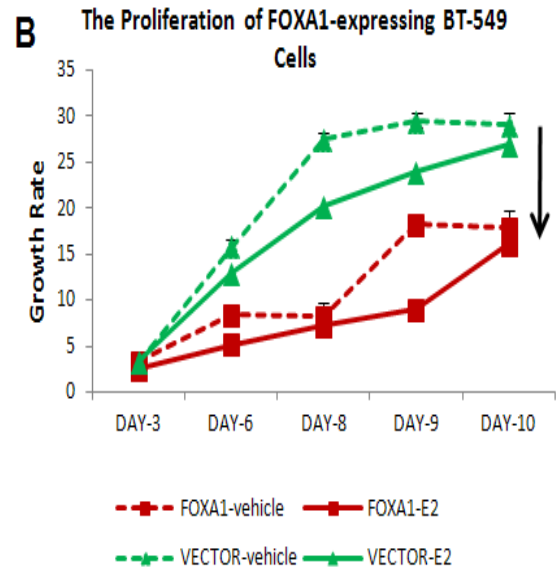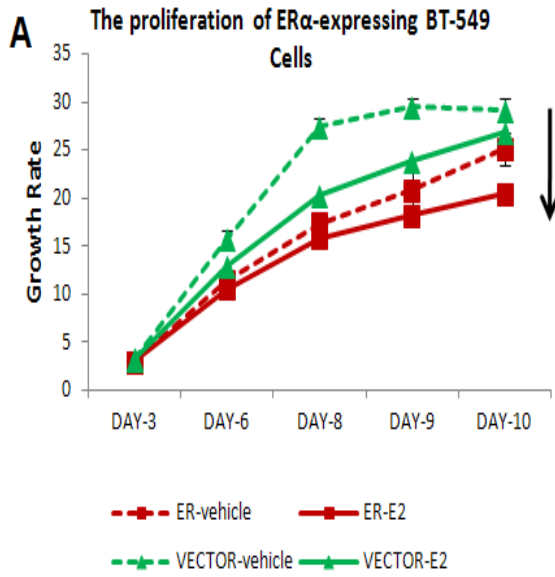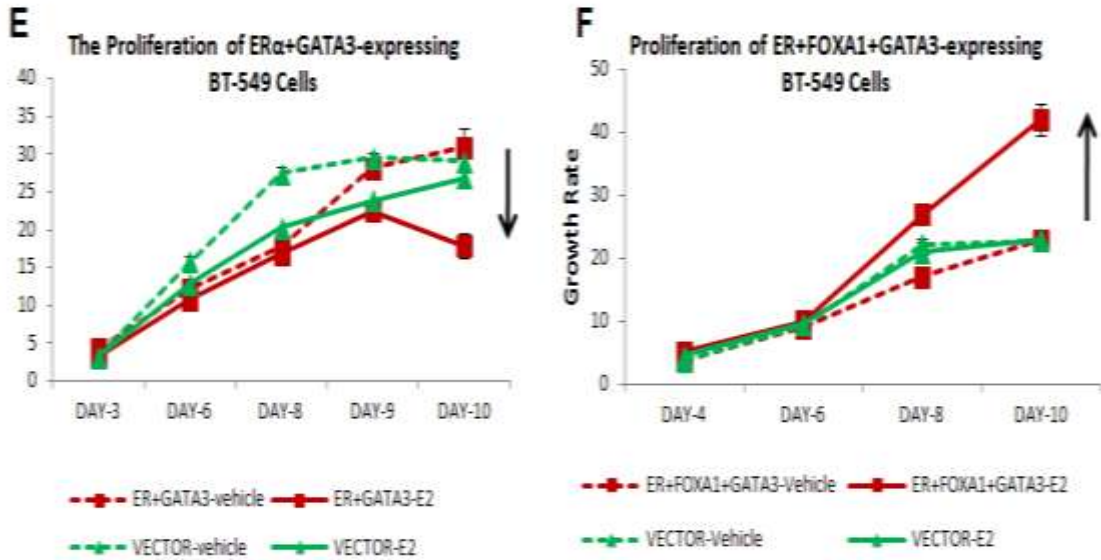
Figure 53. The proliferation of reprogrammed BT-549 cells assessed by cell number count using Hoechst stain. (A) The subtle inhibited growth of BT-549 cells with the transfection of ERα. (B) The inhibited growth of BT-549 cells with the transfection of FOXA1. (C) The marginal induced growth of BT-549 cells with the transfection of GATA3. (D) The inhibited growth of BT-549 cells with co-transfection of ERα+FOXA1. (E) The inhibited growth of BT-549 cells with co-transfection of ERα+GATA3. (F) The induced cell proliferation in response to E2 stimulation in the BT-549 cells with the co-transfection of ERα, FOXA1 and GATA3 in combination. For every sub-figure, means and standard errors of three independent experiments are shown.
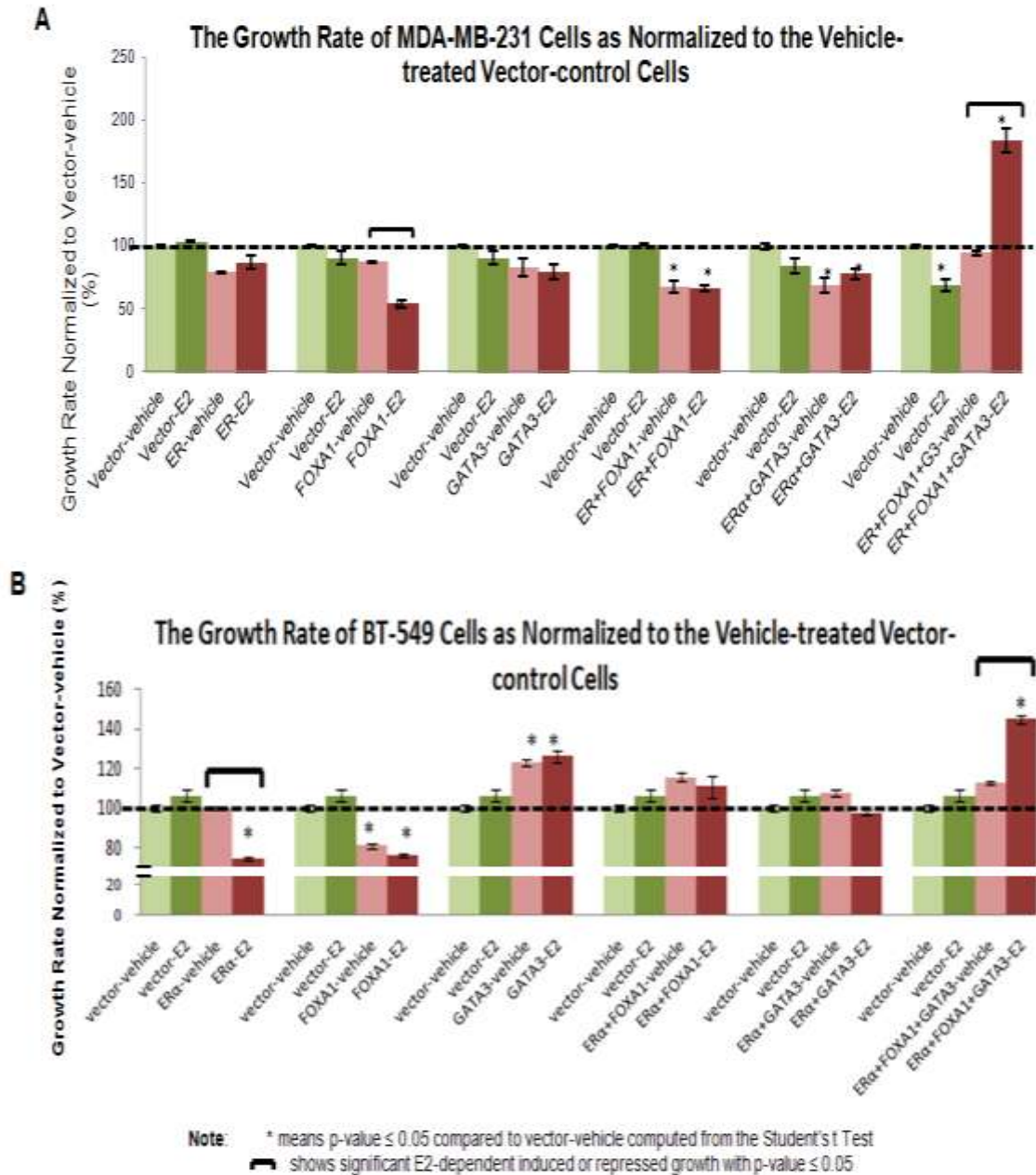
Figure 54. FOXA1 and GATA3 are essential components of E_2-induced ERα-response casette. (A) The growth of MDA-MB-231 cells transfected with different combinations of TFs relative to the vehicle-treated MDA-MB-231 vector control cells at the final day of WST-1 measurement. (B) The recapitulation of reprogramming work in another ERα-negative BT-549 cells. The growth of BT-549 cells tranfected with different combinations of TFs relative to the vehicle-treated BT-549 vector control cells at the final day of WST-1 assay.

136

4.2 The co-existence of ERα, FOXA1 and GATA3 in modulating the luminal and basal cassettes of the reprogrammed cells

Next, we asked if the reprogrammed MDA-MB-231 and BT-549 cells have acquired luminal cell characteristics. We investigated the expression of luminal and basal markers genes defined by Kao et al. (Kao J *et al*, 2009) in the transfected MDA-MB-231 and BT-549 cells. The analysis revealed a modest but discernible induction of luminal markers genes and suppression of basal marker genes in the ERα+FOXA1+GATA3-expressing MDA-MB-231 and BT-549 cells as compared to the ERα-only or vector control cells (Figure 55-56). Therefore, the enhanceosome consists of ERα, FOXA1 and GATA3 not only 'step on the accelerator' to induce a new gene expression program, but also 'put on the brakes' to inactivate the regulators of existing cell type, leading to extinction of markers characteristic of the old phenotype.

Moreover, we found that 63% of the luminal genes are associated with conjoint binding of the three transcription factors ERα+FOXA1+ GATA3 within 20 kb of the TSS, and only 13% of these luminal genes showing no proximity binding of any of these 3 TFs. On the other hand, 24% of the basal genes are associated with proximate conjoint ERα+FOXA1+ GATA3 binding with 40% of these genes are not associated with any ERα, FOXA1 or GATA3 binding (Figure 57).
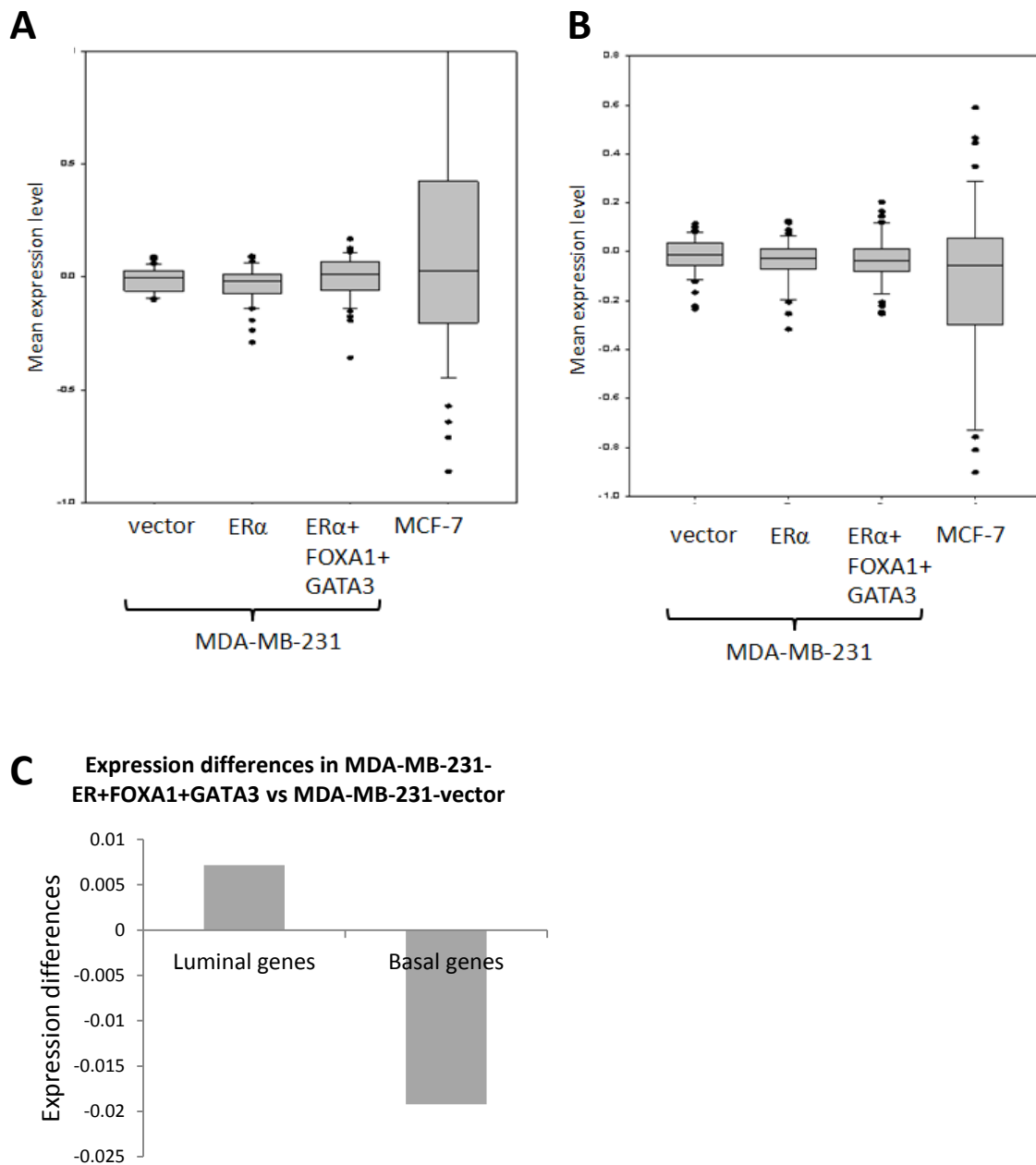
Figure 55. We compared the expression profile of luminal and basal marker genes defined by Kao J *et al.* in the transfected MDA-MB-231 cells. (A) We observed that there is induced expression of luminal marker genes in the ERα+FOXA1+GATA3-expressing MDA-MB-231 cells as compared to ERα-only or vector-control MDA-MB-231 cells. We included the expression of MCF-7 cells that is defined as luminal subtype as comparison. (B) There is reduced expression of basal marker genes in the ERα+FOXA1+GATA3-expressing MDA-MB-231 cells as compared to ERα-only or vector-control MDA-MB-231 cells. (C) This bar plot represented the average expression difference of luminal and basal marker genes in ERα+FOXA1+GATA3-expressing MDA-MB-231 cells as compared to vector-control MDA-MB-231 cells.

138

A

B

C The expression differences in BT-549-ER+FOXA1+GATA3 vs BT-549-vector
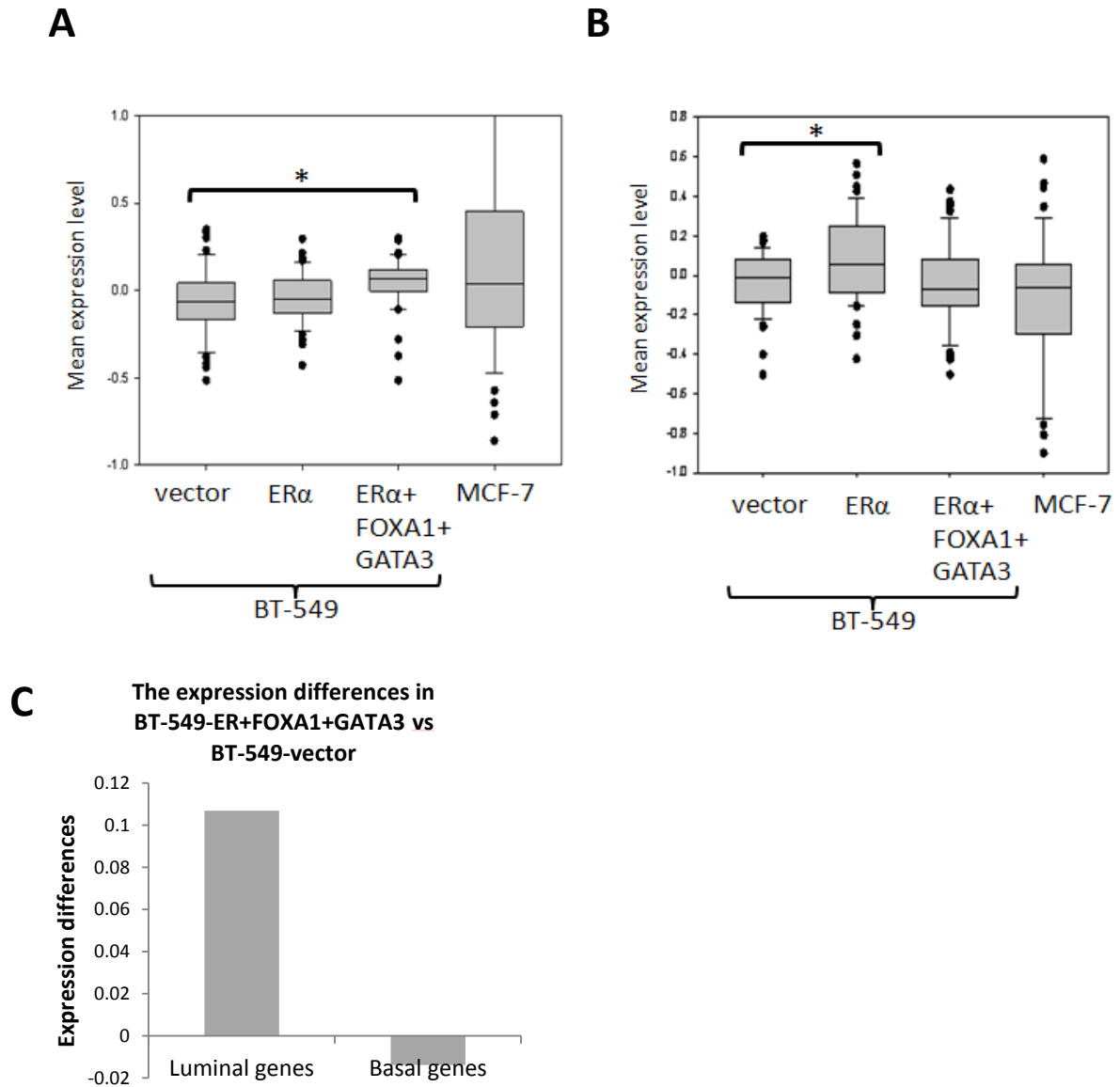
Figure 56. We compared the expression profile of luminal and basal marker genes defined by Kao J *et al.* in the transfected BT-549 cells. (A) We observed that there is induced expression of luminal marker genes in the ERα+FOXA1+GATA3-expressing BT-549 cells as compared to ERα-only or vector-control BT-549 cells. We included the expression of MCF-7 cells that is defined as luminal subtype as comparison. (B) There is reduced expression of basal marker genes in the ERα+FOXA1+GATA3-expressing BT-549 cells as compared to ERα-only or vector-control BT-549 cells. (C) This bar plot represented the average expression difference of luminal and basal marker genes in ERα+FOXA1+GATA3-expressing BT-549 cells as compared to vector-control MDA-MB-231 cells (* denotes p-value ≤ 0.01 compared to the vector-control cells computed from Student's t Test).

A

The Presence of ERα, FOXA1 and GATA3 Bindings within 20kb of the TSS of Luminal Marker Genes

- ERα+FOXA1+GATA3 peaks
- ERα+FOXA1 peaks
- ERα+GATA3 peaks
- FOXA1+GATA3 peaks
- ERα peaks
- FOXA1 peaks
- GATA3 peaks
- No peaks

B

The Presence of ERα, FOXA1 and GATA3 Binding within 20kb of the TSS of Basal Marker Genes

- ERα+FOXA1+GATA3 peaks
- ERα+FOXA1 peaks
- ERα+GATA3 peaks
- FOXA1+GATA3 peaks
- ERα peaks
- FOXA1 peaks
- GATA3 peaks
- No peaks

Figure 57. The presence of ERα, FOXA1 and GATA3 bindings at 20kb around the TSS of luminal and basal marker genes defined by Kao *et al.* (A) We demonstrated that ERα, FOXA1 and GATA3 bindings present in 63% of the luminal marker genes, only 13% of these luminal marker genes have no association with either TF binding. (B) The ERα, FOXA1 and GATA3 bindings are only present in 23% of the basal marker genes while there are as many as 40% of these genes have no association with either of these 3 TFs bindings.

140

This suggests that ERα+FOXA1+GATA3 binding exerts greater impact on regulating the transcription of luminal marker genes as compared to the basal marker genes. Taken together, we demonstrate that the co-expression of the ERα enhanceosome components, namely ERα, FOXA1 and GATA3, is required to approximate the appropriate luminal expression cassette.

CHAPTER 5: THE CORE ESTROGEN-RESPONSIVE CASSETTE THAT DRIVES THE GROWTH AND PROLIFERATION OF BREAST CANCER CELLS

5.1 The presence of ERα, FOXA1 and GATA3 has reprogrammed the transcriptome of ERα-negative cells to acquire the transcriptome of ERα-positive cells

Though the estrogen-driven growth and proliferation is a well-defined signature for ERα-positive breast cancer cells, significant gap of knowledge still exists especially in the understanding on how this hormone controls the growth and proliferation of neoplastic breast epithelium. To address this, we wish to investigate the estrogen-mediated gene regulation in the breast cancer and the manner in which these changes in gene expression affect breast cancer proliferation.

In order to assess the nature of the transcriptional reprogramming, we asked the question if the reprogrammed MDA-MB-231 and BT-549 cells display any similarity in the expression profile of the ERα-positive breast cancer cell line, MCF-7. Gene expression profiling was performed on the ERα-only and ERα+FOXA1+GATA3 transfectant cells. We compared the E_2-regulated genes from these differently transfected MDA-MB-231/ BT-549 cells with MCF-7 cells. Strikingly, we found that the expression profile of E_2 induced ERα+FOXA1+GATA3 expressing MDA-MB-231 cells display a good positive correlation (R=0.42) with the E_2 induced expression profile of MCF-7. By contrast, we

observed a negative correlation between the expression profiles of MDA-MB-231 transfected with ERα only (R= -0.21) (Figure 58). We observed the similar negative correlation (R= -0.12) in the ERα-only BT-549 cells and positive correlation (R = 0.16) in ERα+FOXA1+GATA3 expressing BT-549 cells with MCF-7 though in a weaker correlation strength. Though the temporal expression regulated by the three TFs in MDA-MB-231 and BT-549 cells does not mimic the temporal expression patterns of the MCF-7 cells precisely, it resulting in the same outcome, suggesting that these transcriptional networks are highly robust.

5.2 The reprogrammed cells regulate the cell cycle and proliferation genes in a similar manner with the ERα-positive MCF-7 cells

Using the Ingenuity Pathway Analysis (IPA), we observed that the estrogen responsive genes regulated in MDA-MB-231and BT-549 transfectant cells were significantly associated with cell cycle, cellular proliferation and DNA replication functionalities (p-value = 7.27E-12 – 1.28E-04; 5.59E-07-1.82E-02, Table 10). In addition, when only cell cycle and proliferation genes were examined, again, there was positive correlation between MDA-MB-231 and BT-549 transfected with ERα+FOXA1+GATA3 and MCF-7 but no correlation between ERα-only MDA-MB-231 and BT-549 cells with MCF-7 (Figure 59).
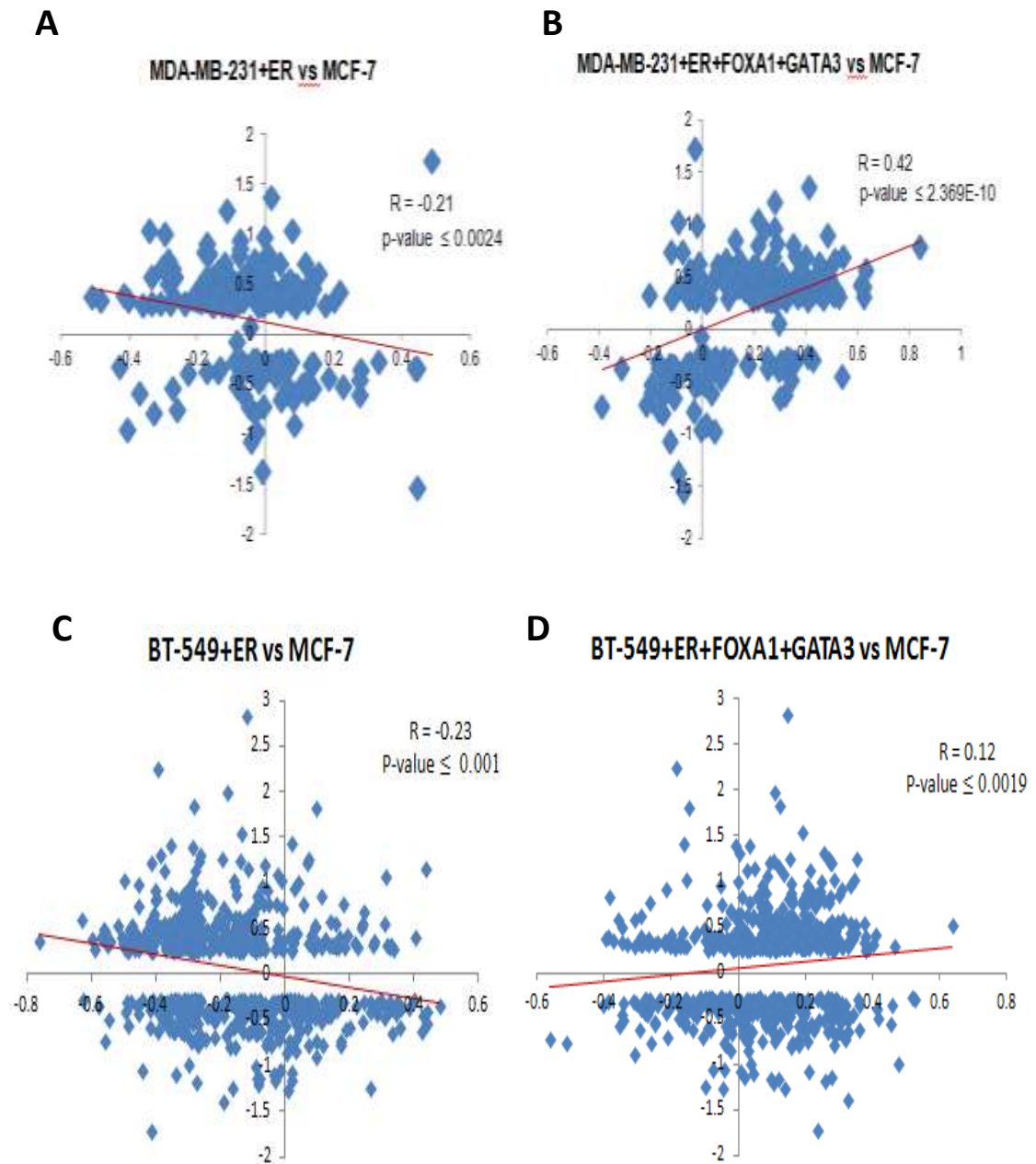
Figure 58. The comparison of gene profilings in reprogrammed MDA-MB-231 and BT-549 with MCF-7 cells. (A) The gene profilings of ERα-only MDA-MB-231 cells show weak correlation with the expression profiles of MCF-7 cells.(B) The ERα+FOXA1+GATA3-expressing MDA-MB-231 cells display good correlation with the expression profile of ERα-positive MCF-7 cells. (C) The gene profilings of ERα-only BT-549 cells show negative correlation with the expression profiles of MCF-7 cells.(D) The ERα+FOXA1+GATA3-expressing BT-549 cells display positive correlation with the expression profile of ERα-positive MCF-7 cells.

144
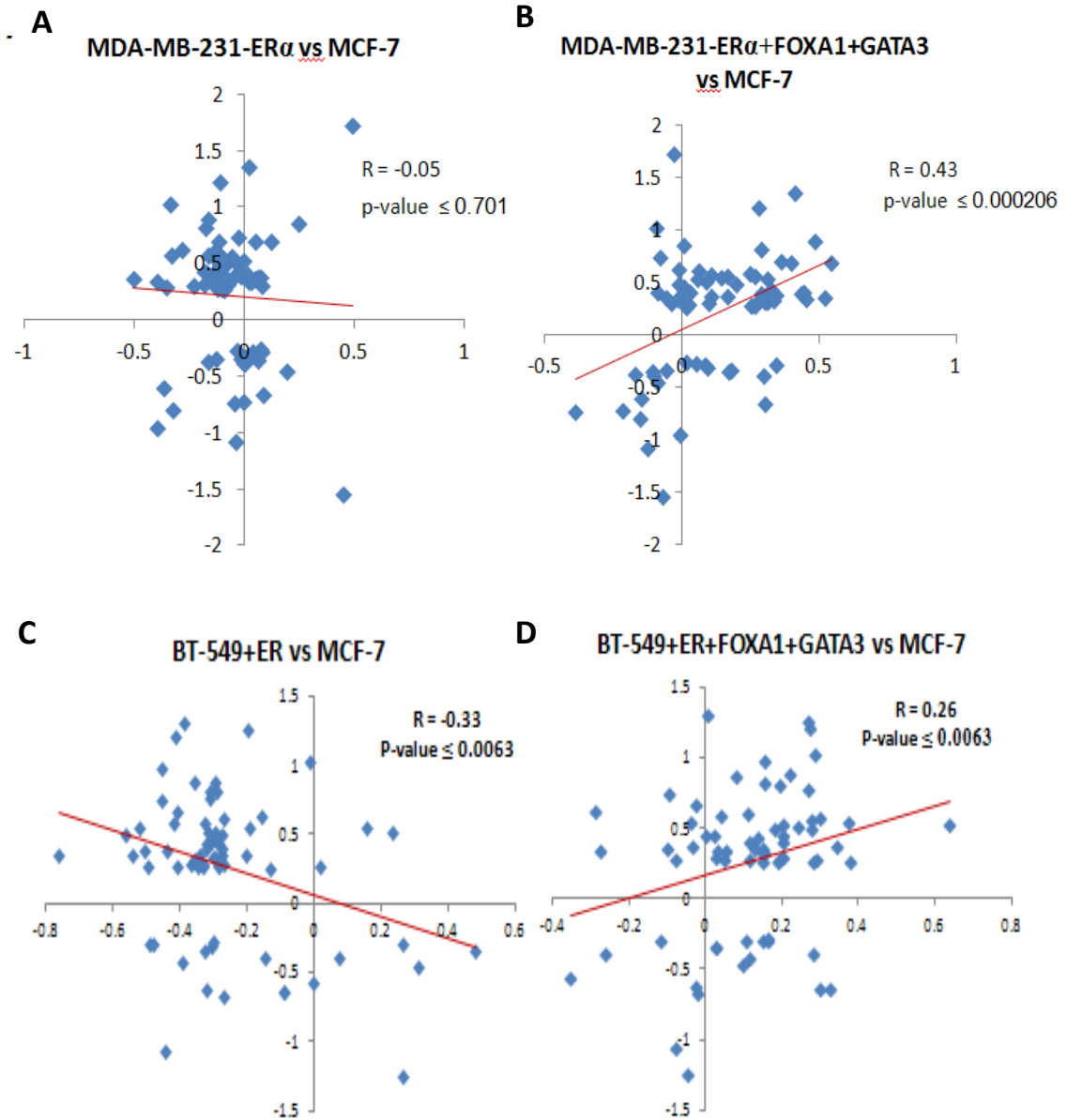
Figure 59. The correlation of the expression profile of cell cycle and cellular proliferation genes between the MCF-7 cells, and the MDA-MB-231 cells transfected with (A) ERα-only, and (B) ERα+FOXA1+GATA3; and BT-549 cells transfected with (C) ERα-only, and (D) ERα+FOXA1+GATA3 . We observed that there was positive correlation between MDA-MB-231 and BT-549 cells transfected with ERα+FOXA1+GATA3 and MCF-7 cells; and negative correlation between the ER-only MDA-MB-231 and BT-549 cells with MCF-7 cells.

Category	p-value for MDA-MB-231 transfectant cells	p-value for BT-549 transfectant cells
Cell Cycle	7.27E-12-4.69E-04	5.59E-07-1.94E-02
Cellular Growth and Proliferation	1.84E-08-4.74E-04	6.52E-06-1.82E-02
DNA Replication, Recombination, and Repair	8.96E-08-1.28E-04	1.22E-06-1.83E-02

Table 10. The Ingenuity Pathway Analysis (IPA) was used to access the functionality of estrogen responsive genes identified in MDA-MB-231 and BT-549 transfectant cells. The estrogen regulated genes identified in the transfected MDA-MB-231 and BT-549 sublines are significantly associated with the cell cycle, cellular proliferation and DNA replication functionalities.

Moreover, we found that there is up-regulation of pro-proliferative cell cycle genes in the ERα+FOXA1+GATA3-expressing MDA-MB-231 and BT-549 cells compared to ERα-only cells (Figure 60). We assessed specific genes previously known to be E_2-regulated in ERα responsive cell lines such as *CCND1, STC2, ABCA3* and *DUSP3 (Frasor J. et al, 2003; Lin C. Y. et al, 2004),* we found that these genes were also regulated in the same direction by ligand in the triple factor transfected MDA-MB-231 or BT-549 cells.

Next, we performed ChIP-seq in the reprogrammed MDA-MB-231 and BT-549 cells to investigate the ERα binding profiles. Intriguingly, we found there are consistently more ERα binding events in the ERα+FOXA1+GATA3-expressing cells compared to the ERα-only cells (Table 11). Most importantly, we observed that there are more ERα binding events at day 10, suggesting that there is progressive recruitment of ERα in the ERα+FOXA1+GATA3-transfectant cells at day 10.

A

B
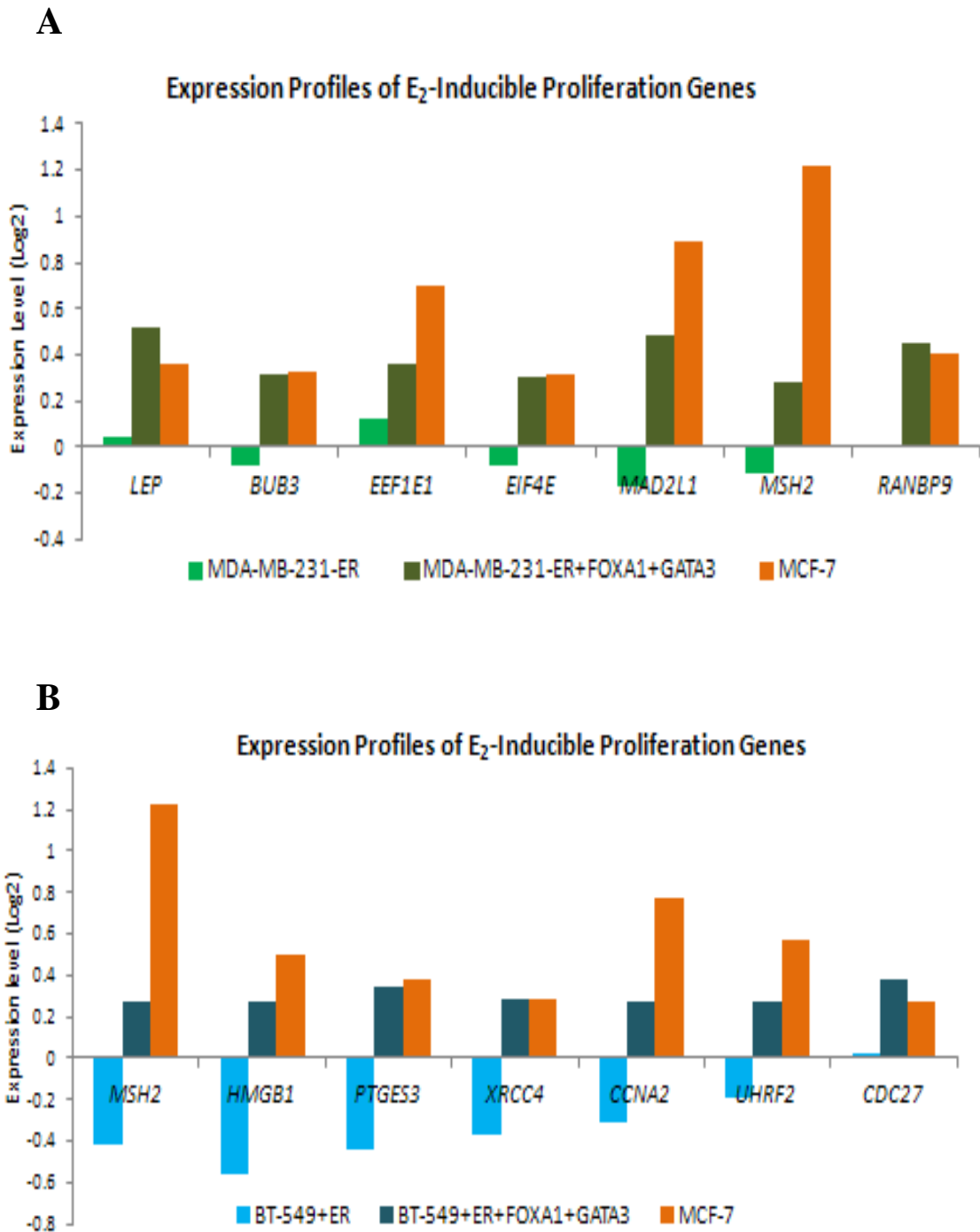
Figure 60. The differential regulation of cell cycle and proliferation genes in the (A) MDA-MB-231 transfectants; (B) BT-549 transfectants and MCF-7cells in response to estrogen stimulation.

	ERα+FOXA1+ GATA3, 45 min	ERα+FOXA1+ GATA3, Day 10	ERα, 45min	ERα, Day 10
MDA-MB-231	2666	3777	934	940
BT-549	2962	4284	2465	1544

Table 11. The ERα binding events in the reprogrammed MDA-MB-231 and BT-549 cells at 45 minutes and Day 10 upon estrogenic stimulation.

We then segregated the ERα binding sites to different categories unique to early (45 minutes) or late (day 10) binding events in ERα+FOXA1+GATA3 or ERα-only cells respectively. We annotated the genes to nearest ERα binding site within ±50kb of TSS, we observed that genes annotated to those unique ERα binding sites in ERα+FOXA1+GATA3 cells were significantly associated with gene expression, cellular assembly and organization, growth and proliferation as well as cell signaling which is absent in the ERα-only cells (Figure 61-62).

Taken together, these results suggest that the presence of ERα, FOXA1 and GATA3 has transcriptionally reprogrammed the ERα-negative MDA-MB-231 as well as the BT-549 cells to resemble the ERα-positive MCF-7 cells by recapitulating the estrogen responsive cassette and manifesting the proliferative phenotype. This suggests that modulation of signaling pathways is able to induce transdetermination (conversion between two closely related progenitor cells that share a direct common progenitor) for the closely related lineages.
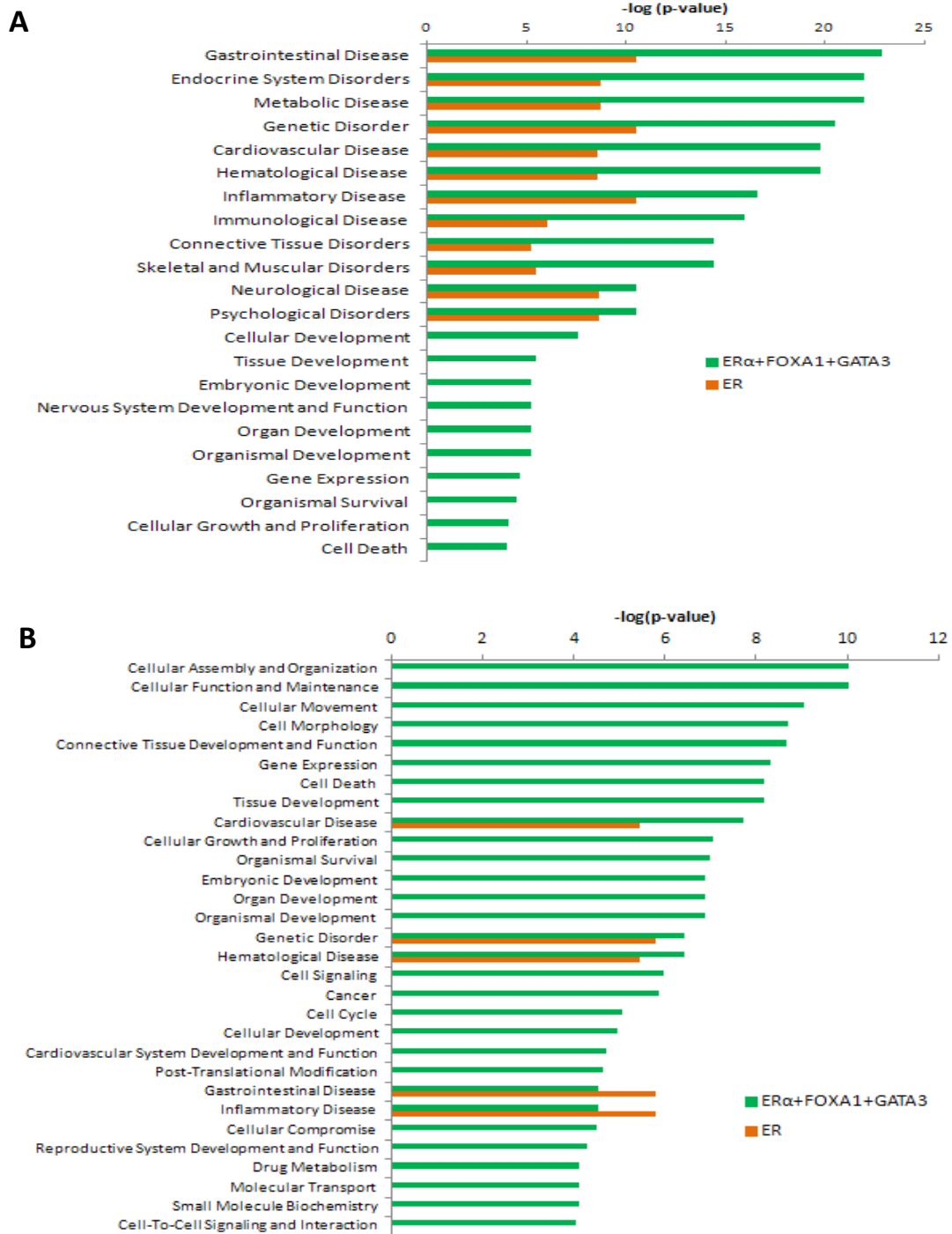
Figure 61. IPA was performed on genes annotated to unique (A) early and (B) late ERα binding events in reprogrammed ERα+FOXA1+GATA3-expressing and ERα-only MDA-MB-231 cells. (p-value ≤ 1E-04 is considered significant)
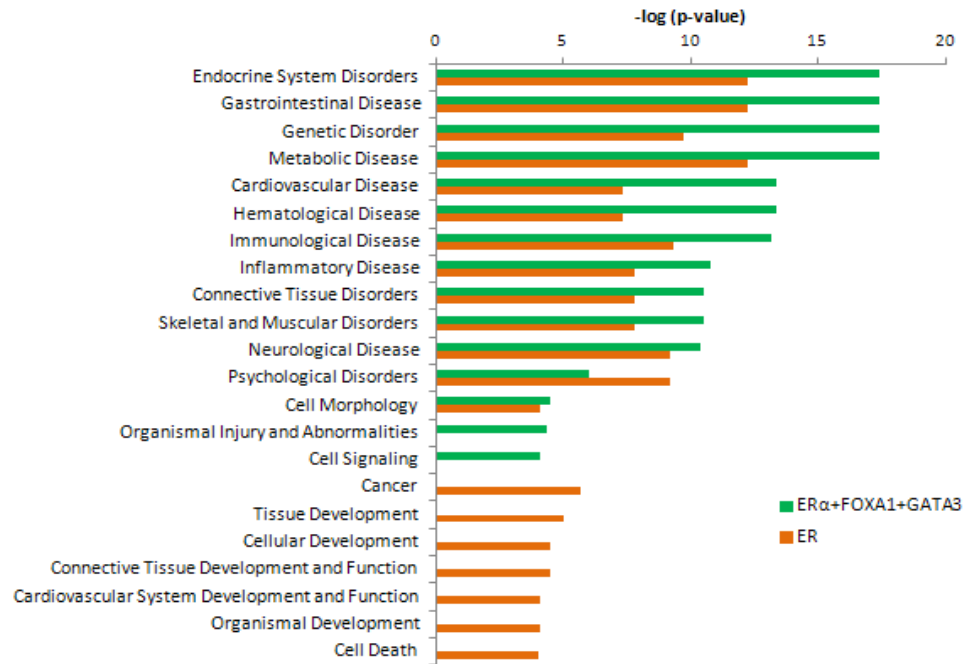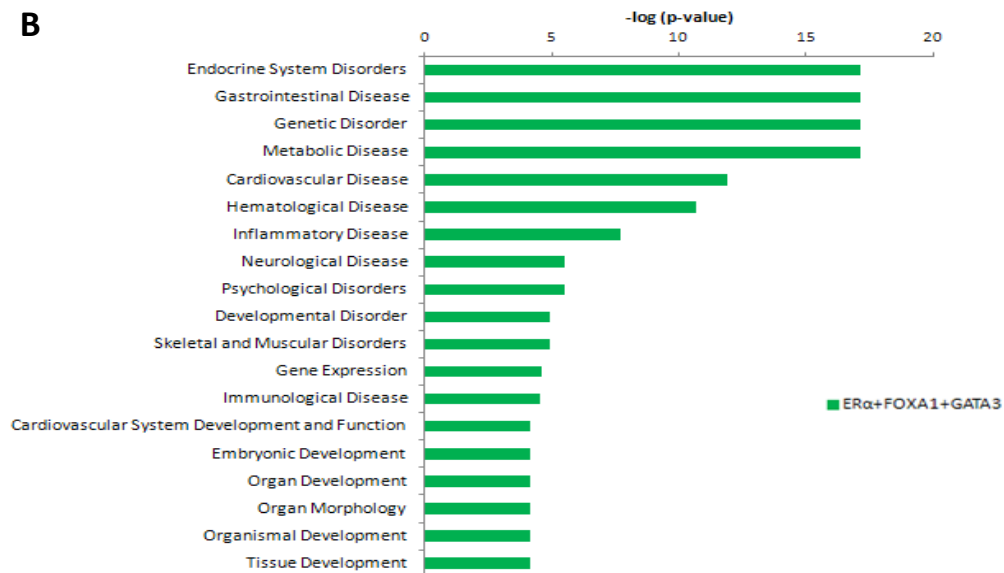
Figure 62. (A) IPA was performed on genes annotated to unique early ERα binding events in reprogrammed ERα+FOXA1+GATA3-expressing and ERα-only BT-549 cells. (B) Since there is no unique late ERα binding event in the ERα-only cells, only the genes associated with late ERα binding events in ERα+FOXA1+GATA3-expressing cell were shown. (p-value ≤ 1E-04 is considered significant)

CHAPTER 6: DISCUSSION

Estrogen receptor, as a prototype of a nuclear hormone receptor, mediates a broad range of cellular and physiologic functions with organ and context specificity. The most proximate form of regulatory control resides in the protein-DNA interaction of TF binding to their cognate recognition motifs and modified by co-factors. However, genome-wide studies of ERα binding show a dispersed occupancy pattern at binding sites bearing heterogeneous recognition motifs that are, at the sequence level, also not well conserved in evolution (Kunarso G *et al*, 2010). This binding site heterogeneity is normalized by chromatin looping to bring these distant and distributed enhancers in proximity to the regulated TSS (Fullwood *et al*, 2009). Herein, we show that FOXA1 and GATA3 are essential for optimal ERα binding to DNA, that FOXA1 and GATA3 are recruited as a complex to the most functional ERα binding sites after ligand activation, and that the binding of this tripartite enhanceosome complex of ERα, FOXA1, and GATA3 is necessary for optimal transcriptional activation in reporter gene assays. The enhanceosome assembly is recruited to sites bearing the three recognition motifs suggests that this complex formation is "hard-wired" in the human genome and provides an evolutionary advantage. This notion is supported by the fact that the co-localization of the motifs for these TFs was found in 23,090 sites in the reference human genome, but in only 360 sites in a random nucleotide sequences for a 64 fold enrichment. This compares to a ~18 fold enrichment for the ERE alone suggesting a strong evolutionary selection for the three TFs to be co-localized (Figure 63).

Figure 63. The presence of ERα, FOXA1 and GATA3 recognition motifs. The position weight matrix from de novo motif finding by the Pomoda (Peak Oriented Motif Discovery Algorithm) was used to screen for the co-existing of the TF's motif in the human genome as compared to the random sequences of 200Mb with the cutoff of 1E-4. There is a 64-fold enrichment for the co-localization of ERα+FOXA1+GATA3 motif in the reference human genome as compare to the random nucleotide sequences.

An enhanceosome has been defined as protein complex composed of a repertoire of TFs that binds to the "enhancer" region of a gene and sequentially recruits components of the transcriptional machinery such as RNA polymerase to initiate the gene's transcription. Synergistic interplay among the members within the enhanceosome complex result in providing some functional specificity, and a multiple gene "fail-safe" mechanism for controlling gene expression (Robert and Tom, 1994). It is suggested that an enhanceosome may provide functional redundancy that minimizes the chances that a gene may be switched off due to mutation, or permit activation of a gene by orchestrating multiple different signaling cascades (Farnham, 2009). The importance of enhanceosome formation is evidenced by the virus-inducible transcriptional activation of the human interferon-β (IFN-β) gene by the assembly of transcriptional activator (p50/ p65), IRF-1, ATF-2, c-Jun and high mobility group protein HMG I to the basal transcription complex (Maniatis, 1995). Chen X *et al*. (Chen *et al*, 2008) showed that TFs coordinately expressed in embryonic stem cell differentiation form specific enhanceosomes adjacent to cassettes of genes that demarcate different developmental functions. The present study provides evidence of how the ERα, FOXA1 and GATA3 enhanceosome regulate this multifaceted transcriptional network operative in reproduction and cancer. Furthermore, we show that this ERα+FOXA1+GATA3 enhanceosome recruits distinct components of active transcription regulatory machinery, namely RNA Pol II and p300, an acetyltransferase associated with enhancer activity as well as chromatin opening. Interestingly, the ERα, FOXA1 and GATA3 binding were also coincided with retinoic acid receptor binding (RAR) though the overlap is less frequent, suggesting that FOXA1

and GATA3 could have a broader "universal" co-regulator function for nuclear hormone binding (Hua S. *et al*, 2009).

It is known that FOXA1 and GATA3 are important regulatory proteins in their own right. FOXA1 has winged helix domains that can structurally mimic histone H1 and H5, and thus permits its interaction with histone H3 and H4. This unique feature of FOXA1 allows it to bind to the specific DNA sequences on the nucleosome core and displace the linker histones, leading to de-compaction of chromatin and to facilitate the binding of other TFs (Cirillo *et al*, 1998; Clark *et al*, 1993; Kaestner, 2000). It is suggested that ERα+FOXA1 regulated network establishes a 'one-step forward' (through cyclin D1 induction) and 'one-step backward' (through p27^{KIP1} induction) manner to control cell cycle progression in breast cancer cells (Nakshatri and Badve, 2009). Recent work by Lupien *et al.* (Lupien *et al*, 2008) revealed that there was significant overlap of FOXA1 occupied sites on ERα cistrome, hence suggesting that FOXA1 contributed in the control of E$_2$ signaling in breast cancer cells.

GATA3 plays essential roles in the mammary gland morphogenesis and lactogenesis. Inactivation of GATA3 resulted in diminished mammary epithelial structure, severely impaired lactogenesis and disrupted differentiation of luminal progenitor cells into ductal and alveolar cells (Asselin-Labat *et al*, 2007). Moreover, GATA3 is also involved in the positive cross-regulatory loop with ERα in breast cancer cells in mediating the E$_2$ signaling (Eeckhoute *et al*, 2007) . Clinically, both FOXA1 and GATA3 are known to be co-expressed in ER positive breast cancers. In addition, Mehra *et al.* (Mehra *et al*, 2005) reported that low levels of GATA3 was strongly associated with larger tumor size,

156

positive lymph node status, higher histology grade, ERα-negative status, Her2-neu overexpression as well as increased risk for recurrence and metastasis. Taken together, we posit that such a complex regulatory and functional interaction of three TFs each subserving important functions is another evolutionary strategy to ensure the balanced co-regulation of gene networks important in mammalian reproduction.

Here, we have shown that the effects of the ERα+FOXA1+GATA3 enhanceosome expression is the regulation of the major important E_2-responsive genes associated with various signaling pathways, biology processes and molecular functions previously ascribed to ERα alone.

Though the presence of an ERα is necessary for E_2 induced growth in responsive cells, its presence is not sufficient for cellular proliferation, and in fact, the introduction of ERα or FOXA1 into ERα negative cell lines such as MDA-MB-231 and BT-549 leads to cell cycle arrest. Importantly, we show that transfection of the three TFs into the ERα negative cell lines, MDA-MB-231 and BT-549, could reprogram the cell to be estrogen responsive for cell proliferation, counteracting the growth inhibitory action of unaided FOXA1 or ERα. A possible explanation for this phenomenon is that the high levels of exogenous TFs simply saturate the transcriptional and translational machinery of the cell and thus actively outcompete the original transcriptional program. This cellular reprogramming is correlated with reconstruction of the approximate transcriptional cassette of the modified MDA-MB-231 and BT-549 to partially resemble that of E_2 stimulated MCF-7 cells. Thus, it appears that the primary role of FOXA1 and ERα alone

in breast cancer cells is as a growth or tumor inhibitor, but that the conditional expression of ERα, FOXA1, and GATA3 reverses this state to that of growth induction. We have revealed that FOXA1 and GATA3 are essential components in maintaining the dependence of ERα-positive breast cancer cells to estrogen for proliferation. Hence, the presence of FOXA1 and GATA3 may help to prevent the breast cancer cells from progressing from a good prognosis group that responds to endocrine treatment to a bad prognosis group that gained anti-estrogen resistance or estrogen independent growth.

Intriguingly, enforced expression of the triple factors, ERα, FOXA1, and GATA3, also induced a modest basal to luminal expression cassette change by reducing the basal signature and increasing the luminal signature in MDA-MB-231 and BT-549 cells not seen in the ERα-alone transfected clone. Our results suggest that the conjoint effects by the three TFs could formulate a luminal cassette and then manifest the proliferative phenotype in response to estrogen stimulation.

Our work also sheds some light on the functional role of FOXA1, which is thought to be a pioneering factor for nuclear hormone receptors such as the estrogen receptor and androgen receptor (Carroll *et al*, 2005). As a pioneering factor, FOXA1 may function to open chromatin structures so as to facilitate ERα binding to its cognate response elements. Indeed, our chromatin model predictive of ERα binding includes FOXA1 occupancy in the preligand (before E_2 exposure) state (Joseph *et al*, 2010). However, these studies did not examine the dynamic relationship of ERα and FOXA1 occupancy before and after ligand exposure. Our results suggest that ERα is as likely to be a pioneering factor to recruit FOXA1 as the converse.

Recently, Eeckhoute *et al.* (Eeckhoute *et al*, 2009) reported that a significant fraction of FOXA1-bound sites have a relatively closed chromatin conformation that is unrelated to gene expression suggesting that FOXA1 may require a repertoire of collaborating TFs to promote chromatin opening. Our findings suggest that ERα is one such collaborating transcription factor with GATA3 playing a more minor role.

Each transcriptional activation comprises a series of events, such as DNA methylation, histone modification, cofactor modulation, each of which alone might not be sufficient for full activation, but together converge to define a specific 'combinatorial code'. Our work has provided strong evidence to illustrate this combinatorial code orchestrated by ERα, FOXA1 and GATA3 enhanceosome that exerts great impact on the transcriptome and cellular growth of breast cancer cells. In contrast, in the collaborative work with Pervaiz and co-workers (Zhou *et al*, in preparation), we revealed that there was competition between activated ERα and Peroxisome-Proliferator-Activated Receptor-γ (PPAR-γ), leading to suppressed recruitment of these TFs to the cis-regulatory regions and consequently down-regulate the targeted genes. Interestingly, the work by Katzenellenbogen and co-workers has uncovered the mutual restriction and competitive selection models whereby the dominant ERα is capable of displacing ERβ binding events when both ERα and ERβ were co-expressed and activated (Charn *et al*, 2010). Cumulatively, these findings have further elaborated the various transcriptional models involving different TF players that synergistically orchestrate the transcription networks of breast cancer cells (Figure 64).

The various models transcriptional activation by different transcription factors
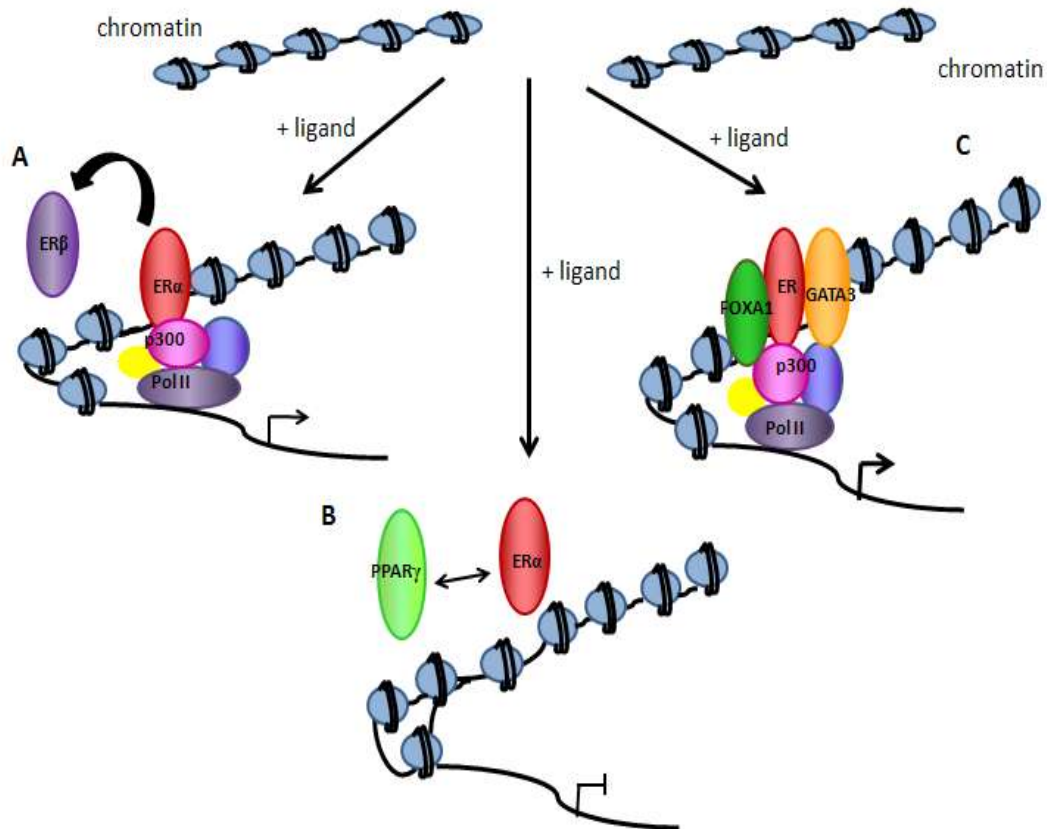
Figure 64. The various transcription activation models involving different players to orchestrate the transcription machinery of breast cancer cells. (A) The mutual restriction and competitive selection models reveal that the dominant ERα is capable in displacing ERβ binding events when both ERα and ERβ were co-expressed and activated. (B) The competitive model between two transcription factors that prevents their recruitment to cis-regulatory region, leading to inhibited transcription. (C) The conjoint action by the enhanceosome consists of ERα, FOXA1 and GATA3 is capable of engaging the optimal transcription machinery.

CHAPTER 7: CONCLUSION

Though numerous expression profiling studies on breast tumors have implicated the co-expression of ERα, FOXA1 and GATA3 in the luminal subtype patient, the nature of their coordinated interaction at the genome level and their biological consequences remain poorly understood.

My work provided evidence on how the ERα, FOXA1 and GATA3 converge to form the enhanceosome that exerts a specific combinatorial code to regulate the multifaceted transcriptional network operative in reproduction and cancer. Furthermore, we show that this ERα+FOXA1+GATA3 enhanceosome recruits distinct components of active transcription regulatory machinery, namely RNA Pol II and p300, an acetyltransferase associated with enhancer activity as well as chromatin opening. The enhanceosome also exerts the greatest impacts in regulating the estrogen-responsive genes and activate the various signaling pathways, biological processes and molecular functions previously ascribed to ERα alone.

The chromatin landscape is organized into higher-order folding and three-dimensional structures that influence the transcriptional circulatory. In our work, we revealed that the ERα+FOXA1+GATA3 enhanceosome is associated with vast majority of the complex ERα long-range interactome, suggesting that the enhanceosome components are essential in orchestrating the three-dimensional long-range interactome of higher complexity (Figure 65).

Despite the role of the ERα pathway as a key growth driver for breast cells, the phenotypic consequence of exogenous introduction of ERα into ERα-negative cells paradoxically has been growth inhibition. Importantly, our study has resolved this physiologic and cellular contradiction by demonstrating the competence of these three TFs to reprogramme the ERα-negative breast cancer cells to acquire the appropriate transcriptome profile and cellular growth resemble to an ERα-positive breast cancer line.

The present study also uncovered potential mechanistic insights on how the ERα+FOXA1+GATA3-expressing cells can positively regulate the growth genes as opposed to the ERα-only cells.

This study contributes to the accumulating body of data which will ultimately provide a systems perspective of the gene regulatory networks in hormonal carcinogenesis and should be of interest to basic and clinical researchers involved in studies of transcriptional regulatory mechanisms and their roles in cancer cell proliferation.

Taken together, we have uncovered the functional importance of an enhanceosome comprising ERα, FOXA1 and GATA3 in the estrogen responsiveness of ERα positive breast cancer cells. The findings described herein have further refined our understanding on the combinatorial control of the enhanceosome-driven transcriptional activation and illuminate a novel role of this enhanceosome component in regulating the growth and proliferation of ERα positive breast cancer cells.

In our future study, we are expanding the work to uncover the core cassette of genes that drive the proliferation of breast cancer cells with the ultimate aim to identify new candidates for treating breast cancer.
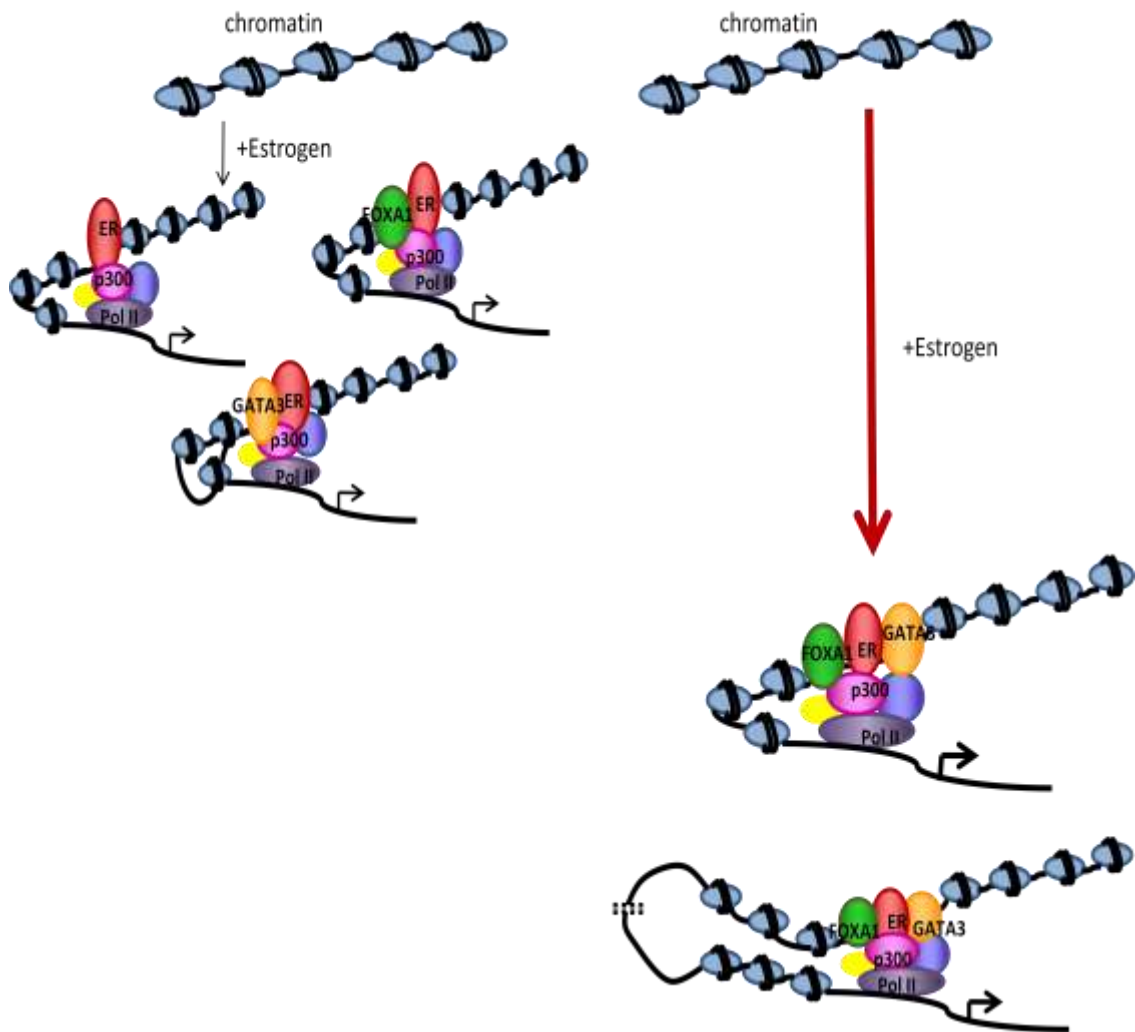
Figure 65. The model of enhanceosome-mediated transcriptional activation in breast cancer cells. Upon estrogen stimulation, the ERα will bind to the cis-regulatory regions individually or in co-operation with FOXA1 or GATA3. Importantly, we revealed that in the ligand-stimulated condition, ERα, FOXA1 and GATA3 will converge to form the enhanceosome that exerts greatest transcriptional and physiological impacts in breast cancer cells. This enhanceosome is essential in regulating the estrogen-responsive genes, activating the various signaling pathways and possibly mediating the ERα long-range interactome. Importantly, we demonstrated that the enhanceosome is competent to reprogramme the ERα-negative cells to acquire the appropriate transcription profile and estrogen-dependent cellular growth resembling to the ERα-positive cells.

163

REFERENCES

Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, Hartley L, Robb L, Grosveld FG, van der Wees J, Lindeman GJ, Visvader JE (2007) Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* **9:** 201-209.

Badve S, Turbin D, Thorat MA, Morimiya A, Nielsen TO, Perou CM, Dunn S, Huntsman DG, Nakshatri H (2007) FOXA1 expression in breast cancer--correlation with luminal subtype A and survival. *Clin Cancer Res* **13:** 4415-4421.

Bernado GM, Lozada K. L., Miedler J. D., Harburg G., Hewitt S. C., Mosley J. D., Godwin A. K., Korach K. S., Visvader JE, Kaestner KH, Abdul-Karim F. W., Montana M. M., A. KR (2010) FOXA1 is an essential determinant of ERa expression and mammary ductal morphogenesis. *Development* **137:** 2045-2054.

Bernado GM, Lozada K. L., Miedler J. D., Harburg G., Hewitt S. C., Mosley J. D., Godwin A. K., Korach K. S., Visvader JE, Kaestner KH, Abdul-Karim F. W., Montana M. M., A. KR (2010) FOXA1 is an essential determinant of ERa expression and mammary ductal morphogenesis. *Development* **137:** 2045-2054.

Bernstein L (2002) Epidemiology of endocrine-related risk factors for breast cancer. *J Mammary Gland Biol Neoplasia* **7:** 3-15.

Bhat-Nakshatri P., Wang G., Appaiah H., Luktuke N., Carroll JS, Geistlinger TR, Brown M, Badve S, Liu Y., Nakshatri H (2008) AKT alters genome-wide estrogen receptor a binding and impacts estrogen signaling in breast cancer. *Molecular and Cellular biology* **28:** 7487-7503.

Blasco MA, Serrano M, O. F-C (2011) Genomic instability in iPS: time for a break. *EMBO J* **30:** 991-993.

Boyle AP., Davis S., Shulha HP., Meltzer P., Marguilies EH., Weng Z, Furey TS., Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132:** 311-322.

Carey M. (1998) The enhanceosome and transcriptional synergy. *Cell* **92:** 5-8.

Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122:** 33-43.

Carroll JS, M. B (2006a) Estrogen receptor target gene: an evolving concept. *Mol Endocrinol* **20:** 1707-1714.

Carroll JS, Meyer CA, Song J., Li W., Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck K. C., Hall G. F., Wang Q., Bekiranov S., Sementchenko V., Fox E. A., Silver PA, Gingeras T. R., Liu X. S., M. B (2006b) Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38:** 1289-1297.

Charn TH, Liu E.T., Chang E. C., Lee Y. K., Katzenellenbogen JA, Katzenellenbogen BS (2010) Genome-wide dynamics of chromatin binding of estrogen receptors alpha and beta: mutual restriction and competitive site selection. *Mol Endocrinol* **24:** 47-59.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133:** 1106-1117.

Cheung E, Schwabish MA., W. LK (2003) Chromatin exposes intrinsic differences in the transcriptional activities of estrogen receptors alpha and beta. *EMBO J* **22:** 600-611.

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9:** 279-289.

Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, Clark KL, Burley SK, Zaret KS (1998) Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J* **17:** 244-254.

Clark KL, Halay ED, Lai E, Burley SK (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364:** 412-420.

Clarke RB., Howell A., Potten CS, E. A (1997) Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Res* **57:** 4987-4991.

Cosma MP (2002) Ordered recruitment: gene-specific mechanism of transcription activation. *Mol Cell* **10:** 227-236.

Couse JF., S. KK (1999) Estrogen receptor null mice: what have we learned and where will they lead us? *Endocr Rev* **20:** 358-417.

DeNardo DG, Cuba VL, Kim HT, Wu K, Lee AV, PH. B (2007) Estrogen receptor DNA binding is not required for estrogen-induced breast cell growth. *Molecular and Cellular Endocrinology* **277:** 13-25.

Eeckhoute J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M (2007) Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Res* **67:** 6477-6483.

Eeckhoute J, Lupien M, Meyer CA, Verzi MP, Shivdasani RA, Liu XS, Brown M (2009) Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res* **19:** 372-380.

Farnham PJ (2009) Insights from genomic profiling of transcription factors. *Nat Rev Genet* **10:** 605-616.

Frasor J., Danes J. M., Komm B., Chang K. C. N., Lyttle R., Katzenellenbogen BS (2003) Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology* **144:** 4562-4574.

Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li H, Desai KV, Thomsen JS, Lee YK, Karuturi RK, Herve T, Bourque G, Stunnenberg HG, Ruan X, Cacheux-Rataboul V, Sung WK, Liu ET, Wei CL, Cheung E, Ruan Y (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462:** 58-64.

Garcia M, Derocq D, Freiss G, Rochefort H (1992) Activation of estrogen receptor transfected into a receptor-negative breast cancer cell line decreases the metastatic and invasive potential of the cells. *Proc Natl Acad Sci U S A* **89:** 11538-11542.

Grodstein F, Manson JE., Colditz GA., Willert WC., Speizer FE., MJ S (2000) A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med* **133:** 933-941.

Gruber CJ, Tschugguel W., Schneeberger C., C. HJ (2002) Production and actions of estrogens. *N Engl J Med* **346:** 340-352.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39:** 311-318.

Hua S., Kittler R., K.P. W (2009) Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* **137:** 1259-1271.

Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, Bild A, Iversen ES, Liao M, Chen CM, West M, Nevins JR, AT. H (2003) Gene expression predictors of breast cancer outcomes. *Lancet* **361:** 1590.

Hurtado A, Holmes K.A., Ross-Innes C.S., Schmidt D., Carroll JS (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* **43:** 27-34.

Jensen EV, Jacobson HI (1960) Basic guides to the mechanism of estrogen action. *Recent Prog Horm Res* **18:** 387-414.

Jordan VC, Wolf MF., Mirecki DM, Whitfold DA, WV. W (1988) Hormone receptor assays: clinical usefulness in the management of carcinoma of the breast. *Crit Rev Clin Lab Sci* **26:** 97-152.

Joseph R, Orlov YL, Huss M, Sun W, Kong SL, Ukil L, Pan YF, Li G, Lim M, Thomsen JS, Ruan Y, Clarke ND, Prabhakar S, Cheung E, Liu ET (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor α. . *Mol System Biol* **6:456**.

Kaestner KH (2000) The hepatocyte nuclear factor 3 (HNF3 or FOXA) family in metabolism. *Trends Endocrinol Metab* **11:** 281-285.

Kao J, Salari K., Bocanegra M., Choi Y., Girard L., Gandhi J., Kwei K. A., Hernandez-Boussard T., Wang P., Gazdar A. F., Minna J. D., R. PJ (2009) Molecular Profiling of Breast Cancer Cell Lines Defines Relevant Tumor Models and Provides a Resource for Cancer Gene Discovery. *Plos one* **4:** e6146.

Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, Kwei KA, Hernandez-Boussard T, Wang P, Gazdar AF, Minna JD, Pollack JR (2009) Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* **4:** e6146.

Kouros-Mehr H, Slorach EM., Sternlicht MD., Z. W (2006) GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127:** 1041-1055.

Kunarso G, Na-Yu Chia, Justin Jeyakani, Catalina Hwang, Xinyi Lu, Yun-Shen Chan, Huck-Hui Ng, Bourque G (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics* **42:** 631-634.

Lacroix M, G. L (2004) About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. *Mol Cellular Endocrin* **219:** 1-7.

Li G, Fullwood MJ, Xu H, Mulawadi FH, Velkov S, Vega V, Ariyaratne PN, Mohamed YB, Ooi H-S, Tennakoon C, Wei C-L, Ruan Y, Sung W-K (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biology* **11:** R22.

Lin C. Y., Strom A., Vega VB, Kong SL, Yeo AL, Thomsen JS, Chan W. C., Doray B., Bangarusamy D. K., Ramasamy A., Vergara L. A., Tang S., Chong A., Bajic V. B., Miller L. D., Gustafsson JA., L. ET (2004) Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome Biol* **5: R66**.

Lin CY, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, Yeo A, George J, Kuznetsov VA, Lee YK, Charn TH, Palanisamy N, Miller LD, Cheung E, Katzenellenbogen BS, Ruan Y, Bourque G, Wei CL, Liu ET (2007) Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet* **3:** e87.

Lonning PE, Sorlie T, Perou CM, Brown PO, Botstein D, AL. B-D (2001) Microarrays in primary breast cancer--lessons from chemotherapy studies. *Endocr Relat Cancer* **8:** 259-263.

Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132:** 958-970.

Maniatis DTaT (1995) Virus Induction of Human IFN-β Gene Expression Requires the Assembly of an Enhanceosome. *Cell* **83:** 1091-1100.

Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* **31:** 374-378.

Mehra R, Varambally S, Ding L, Shen R, Sabel MS, Ghosh D, Chinnaiyan AM, Kleer CG (2005) Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* **65:** 11259-11264.

Mehta RJ, Jain RK., Leung S., Choo J., Nielsen T, Huntsman D, Nakshatri H, Badve S (2011) FOXA1 is an independent prognostic marker for ER-positive breast cancer. *Breast Cancer Res Treat*.

Metivier R, Penot G., Hubner M.R., Reid G., Brand H., Kos M., F. G (2003) Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115:** 751-763.

Naftolin F, Garcia-Segura LM., Keefe D., Leranth C., Maclusky NJ., JR B (1999) Role of astroglia in estrogen regulation of synaptic plasticity and brain repair. *J Neurosci* **40:** 574-584.

Nakshatri H, Badve S (2009) FOXA1 in breast cancer. *Expert Rev Mol Med* **11:** e8.

Nilsson S, Makela S., Treuter E., Tujague M., Thomsen JS, Andersson G., Enmark E., Pettersson K., Warner M., JA. G (2001) Mechanisms of estrogen action. *Physiol Rev* **81:** 1535-1565.

Oh DS, Troester MA., Usary J., Hu Z., He X., Fan C., Wu J., Carey LA., Perou CM (2006) Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* **24:** 1656-1664.

Paganini-Hill A, Dworsky R., M. KR (1996) Hormone replacement therapy, hormone levels, and lipoprotein cholesterol concentrations in elderly women. *Am J Obstet Gynecol* **174:** 897-902.

Pandolfi PP, Roth ME., Karis A., Leonard MW., Dzierzak E., Grosveld FG, Engel JD., MH L (1995) Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nat Genet* **11:** 40-44.

Parikh P, Palazzo JP, Rose LJ, Daskalakis C, RJ. W (2005) GATA-3 expression as a predictor of hormone response in breast cancer. *J Am Coll Surg* **200:** 705-710.

Perissi V, Rosenfeld MG (2005) Controlling nuclear receptors: the circular logic of cofactor cycles. **6:** 542-554.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Asklen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale A, Brown PO, Botstein D (2000) Molecular portraits of human breast tumors. *Nature* **406:** 747-752.

Pierrou S, Hellqvist M., Samuelsson L., Enerback S., P. C (1994) Cloning and characterization of seven human forkhead proteins: Binding site specificity and DNA bending. *EMBO J* **13:** 5002-5012.

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtypes of breast cancer. *Breast Cancer Res* **12**.

Robert T, Tom M (1994) Transcriptional Activation: A Complex Puzzle with Few Easy Pieces. *Cell* **77:** 5-8.

Ross-Innes C.S., Stark R., Teschendorff A.E., Holmes K.A., Ali H. R., Dunning M. J., Brown G. D., Gojis O., Ellis I. O., Green A. R., Ali S., Chin S., Palmieri C., Caldas C., Carroll JS (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481:** 389-393.

Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M (2000) Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103:** 843-852.

Shao W, Brown M (2004) Advances in estrogen receptor biology: prospects for improvements in targeted breast cancer therapy. *Breast Cancer Res* **6:** 39-52.

Shoemaker J, Saraiva M., A. OG (2006) GATA-3 directly remodels the IL-10 locus independently of IL-4 in CD4+ T cells. *J Immunol* **176:** 3470-3479.

Sorlie T, Tibshirani R., Parker J., Hastie T., Marron J. S., Nobel A., Deng S., Johnsen H., Pesich R., Geisler S., Demeter J., Perou CM, Lonning PE, Brown PO, Borresen-Dale A, D. B (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100:** 8418-8423.

Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, ET. L (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* **100:** 10393-10398.

Takahashi K, S. Y (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663-676.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, S. Y (2007) Induction of plutipotent stem cells from adult human fibroblasts by defined factors. *Cell* **131:** 861-872.

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13:** 2129-2141.

van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530-536.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, R. B (2002) A

gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347:** 1999-2009.

Walker P, Germond JE., Brown-Luedi M., Givel F., W. W (1984) Sequence homologics in the region preceeding the transcription initiation site of the liver estrogen-responsive vitellogenin and apo-VLDLII genes. *Nucleic Acids Res* **12:** 8611-8626.

Wall L., deBoer E., Grosveld F (1988) The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes & development* **2:** 1089-1100.

Webb P., Nguyen P., Valentine C., Lopez GN., Kwok GR., Mclnerney E., Katzenellenbogen BS, Enmark E., Gustafsson JA., Nilsson S., PJ. K (1999) The estrogen receptor enhances AP-1 activity by two distinct mechanisms with different requirements for receptro transactivation functions. *Mol Endocrinol* **13:** 1672-1685.

Wilson BJ, Giguere V (2008) Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. *Mol Cancer* **7:** 49.

Witkowska HE, Carlquist M., Engstrom O. (1997) Characterization of bacterially expressed rat estrogen receptor beta ligand binding domain by mass spectrometry: structural comparison with estrogen receptor alpha. *Steroids* **62:** 621-631.

Wolf I, Bose S, Williamson EA, Miller CW, Karlan BY, Koeffler HP (2007) FOXA1: Growth inhibitor and a favorable prognostic factor in human breast cancer. *Int J Cancer* **120:** 1013-1022.

Yu Q., He M., Lee N.M., E.T. L (2002) identification of myc-mediated death response pathways by microarray analysis. *journal of biological chemistry* **277:** 13059-13066.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137.

Zhang Z, Chang CW, Goh WL, Sung WK, Cheung E (2011) CENTDIST: discovery of co-associated factors by motif distribution. *Nucleic Acids Res*.

Zhang Z., Chang C. W., Goh W. L., Sung W. K., Cheung E (2011) CENTDIST: discovery of co-associated factors by motif distribution. *Nucleic Acids Res* **39:** W391-399.

Zhao HH, Herrera RE, Coronado-Heinsohn E, Yang MC, Ludes-Meyers JH, Seybold-Tilson KJ, Nawaz Z, Yee D, Barr FG, Diab SG, Brown PH, Fuqua SA, Osborne CK (2001) Forkhead homologue in rhabdomyosarcoma functions as a bifunctional nuclear

receptor-interacting protein with both coactivator and corepressor functions. *J Biol Chem* **276:** 27907-27912.

Publications

Goncalves VF, Tiwari AK, de Luca V, <u>Kong SL</u>, Zai C, Tampakeras M, Likodi O, Mackenzie B, Shaikh S, Sun L, Paterson A and Kennedy JL. DRD4 VNTR polymorphism and age at onset of severe mental illnesses. Neuroscience Letters 519(1): 9-13 (2012).

<u>Kong SL</u>, Li G, Loh SL, Sung WK and Liu ET. Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. Molecular Systems Biology 7: 526 (2011).

Joseph R, Orlov YL, Huss M, Sun W, <u>Kong SL</u>, Ukil L, Pan YF, Li G, Lim M, Thomsen JS, Ruan Y, Clarke ND, Prabhakar S, Cheung E and Liu ET. Integrative model of genomic factors for determining binding site selection by estrogen receptor-α. Molecular Systems Biology 6: 456 (2010).

<u>Kong SL</u>, Li G, Loh SL, Sung W and Liu ET. The core estrogen responsive cassette that drives the growth and proliferation of breast cancer cells. *(In preparation)*

Conference papers & oral presentation

<u>Kong SL.</u> An enhanceosome-mediated framework for the optimal transcriptional activation and cellular proliferation in breast cancer cells. The 3rd Ray Wu Symposium, Zhejiang University, 8-9th November 2012, Hangzhou, Republic of China.

<u>Kong SL</u>, Li G, Loh SL, Sung K, Liu ET. The synergistic cellular reprogramming by the enhanceosome consists of ERα, FOXA1 and GATA3 in the breast cancer cells. The 2nd Yong Loo Lin School of Medicine – Graduate Scientific Congress, National University of Singapore, 15th Feb 2012, Singapore.

<u>Kong SL</u>, Li G, Loh SL, Sung K, Liu ET. Delineate the enhanceosome-mediated transcriptional networks in breast cancer cells. The 1st Yong Loo Lin School of Medicine – Graduate Scientific Congress, National University of Singapore, 25th Jan 2011, Singapore.

<u>Kong SL</u>, Li G, Loh SL, Sung K, Liu ET. Deciphering the transcriptional networks regulated by estrogen receptor, FOXA1 and GATA3 in breast cancer cells. The EMBO meeting, 4th – 7th September 2010, Barcelona, Spain.

APPENDIX I

Awards

Received Top 10 Publication Award for the paper entitled "Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. Molecular Systems Biology 7: 526 (2011)" in the contest for Merck Millipore Young Scientist Award 2011, 19th November 2012, Singapore.

Received Ray Wu Prize 2012 awarded by Ray Wu Memorial Fund, 8-9th November 2012, Hangzhou, Republic of China.

Received 1st-runner up Best Oral Presentation Award in the 1st Yong Loo Lin School of Medicine – Graduate Scientific Congress, National University of Singapore, 25th Jan 2011, Singapore.

APPENDIX II

Supplemental information:

The raw ChIP-seq sequences and processed data can be accessed from NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE23701, GSE23893, GSE26831 and GSE29073. The gene expression data can be accessed from NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE30574.

APPENDIX III

Contributions:

Kong Say Li and Prof Edison T Liu conceptualized and designed the experimental strategy. Kong Say Li coordinated the study and performed all the experiments in this study. Mr Gary Loh Siang Lin provided technical assistance in cell culture and western blot experiments. Dr Li Guoliang performed the computational and statistical analyses described in section 2.26-2.29 and 2.32. Kong Say Li performed the computational and statistical analyses described in 2.30-2.31. Prof Sung Wing-Kin provided advice for the computational analyses.