THE UNIVERSITY OF CHICAGO

GENOMIC AND FUNCTIONAL ANALYSIS OF ESTROGEN RESPONSIVE CIS-REGULATORY ELEMENTS IN BREAST CANCER CELLS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON GENETICS

BY

SHAN YU

CHICAGO, ILLINOIS

DECEMBER 2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to express my greatest gratitude to Dr. Kevin White, my PhD advisor, for his advice on my thesis study in the past five years and for the wonderful research environment he created for me. Kevin is an extraordinary scientist who always has outstanding vision and sets high standards to the research in his lab. He influenced me a lot on how to bring up scientific questions, how to analyze and solve problems, and how to think, write and present as a scientist. Beyond these, he taught me how to become an independent scientist, which will be a big benefit for me through my entire future career.

I would like to thank the rest of my thesis committee, Dr. Andrey Rzhetsky, Dr. Geoffrey Greene and Dr. Kay Macleod, for their continuous guidance and encouragement. They provided me with insightful suggestions without reservation to help me structure my thesis work. They always asked sharp questions to push me to think harder and guide me to conduct my research towards right direction. I would not accomplish my research without their support.

I feel fortunate to have worked with my fellows in White lab and have interacted with so many talented people in IGSB. In particular, I would like to thank Dr. Yuwen Liu, Dr. Jie Zhou, Dr. Michael Bolt, Dr. Lijia Ma, Dr. Vineet Dhiman, Dr. Ruby Dhar, Dr. Mohana Ray, Dr. Megan McNerney, Dr. Xiaoyue Wang, Dr. Chai Bandlamudi, Dr. Aashish Jha, Jason Pitt, Amber Thomas, Dr. Mark Ciaccio and Eugene Leypunskiy for their valuable scientific input and technical assistance in my thesis study. I also want to thank directors and staffs of all IGSB core facilities for their continuous technical support. I am very grateful to Dr. Barry Aprison and Dr. Michelle Wright for their help with my scientific writing. Besides, I would like to thank Liza Holtzman, Heather Scott

and Temi Okubadejo. Because of their thoughtful administrative assistance, I was able to focus on my research without being distracted.

I also want to thank my graduate program advisors Dr. Richard Hudson and Dr. Yoav Gilad. As program chairs, they spent valuable time communicating with me to help me solve all kinds of problems I encountered during my graduate life. And I sincerely appreciate the patient and professional administrative support from Susan Levison.

Lastly, I am extremely grateful for having the endless support from my loving parents. They sacrificed a lot in order to create the best opportunities for me to achieve my career success. I feel guilty that I did not get many chances to visit my parents and was not able to be accompanied with them in China in the past six years, but they showed their biggest understanding and strongest support of my graduate study. They are always the source of my courage and power.

ABSTRACT

Understanding of estrogen signaling governed gene transcriptional regulation is fundamental to elucidating the biology of hormone-related diseases such as breast cancer and has wide clinical implications. In this thesis, we present a systematic analysis of estrogen-dependent genome-wide interplay between cis-acting chromatin and transcription factors (TFs) in estrogen-responsive breast cancer cell line MCF7. In Chapter 1, we give an overarching introduction of the molecular basis of breast cancer heterogeneity, focusing on the biology of estrogen receptor (ER)-mediated estrogen action in ER positive breast cancer. We also give a review on gene transcriptional regulation and the development of recent technology in studying cis-regulatory elements. In Chapter 2, via profiling estrogen-triggered dynamic changes of chromatin accessibility landscape and functional assessment of regulatory activity of massive DNase I hypersensitivity sites (DHSs) in high-throughput manner, we uncover unique chromatin signatures associated with estrogen-responsive DHSs and experimentally identify large amounts of enhancers including estrogen-responsive enhancers. From genomic analysis, we reveal intensive crosstalk and coordinated action of multiple TFs on estrogen-responsive enhancers. In Chapter 3, we focus on a specific ER-responsive TF, KLF4, and functionally characterize the role of KLF4 in estrogen-governed transcriptional regulation. We demonstrate estrogen-dependent interaction between KLF4 and cis-regulatory elements and discover widespread crosstalk between KLF4 and ER at estrogen-responsive DHSs. Finally, we identify KLF4 as a novel estrogen early effector and cooperative TF of ER in regulating genes involved in cell cycle regulation, which is important for maintaining estrogen-dependent cell proliferation of MCF7 cells.

CHAPTER 1

Introduction

1.1 Breast cancer heterogeneity

Breast cancer is an invasive cancer originating from epithelial cells in breast lobules or milk ducts, and sometimes from stromal tissues surrounding the mammary gland. It is the most common cancer in women across the world. As of 2016, breast cancer contributes to 29% (246,660 cases) of the estimated new cancer cases and 14% (40,450 cases) of the estimated cancer deaths in US women (Siegel, Miller, and Jemal 2016). Breast cancer is the outcome of the accumulation of a series of genetic aberrations in tumor cells that collectively alter cell physiology. Therefore, it is a heterogeneous disease with high inter- and intra-tumor diversity. Understanding subtype- and individual-specific molecular basis of the disease is crucial for improving breast cancer prevention, diagnosis, therapy and prognosis.

1.1.1 Breast cancer subtypes

Breast cancer is among the very few cancers in which different disease subtypes are systematically classified and characterized. Clinically, breast cancer patients are stratified based on the expression of three pathological markers - estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) – that are predictive of tumor progression and factored into treatment decisions and prognosis. Due to the development of high-throughput

gene expression profiling technologies such as gene expression microarray and RNA-seq in the past 15 years, more comprehensive classification system called "intrinsic subtype" has been established in order to better represent the molecular portraits of breast cancer (Perou et al. 2000; Sorlie et al. 2001).

The six intrinsic subtypes of breast cancer are Luminal A, Luminal B, HER2-enriched, Basal-like, Normal breast-like and Claudin-low. Their frequencies in patient population, expression of classical pathological markers, histologic grades, disease aggressiveness and prognosis are reviewed in **Table 1**.

Intrinsic subtype	Frequency	ER/PR/HER2 expression	Histologic grade	Prognosis
Luminal A	50-60%	ER+, PR+, HER2-	Low	Excellent
Luminal B	10-20%	ER+, PR+/-, HER2+/-	High	Intermediate/ Poor
HER2-enriched	10-15%	ER-, PR-, HER2+	High	Poor
Basal-like	10-20%	ER-, PR-, HER2-	High	Poor
Normal breast-like	5-10%	ER+/-, HER2-	Low	Intermediate
Claudin-low	12-14%	ER-, PR-, HER2-	High	Poor

Table 1. Characteristics of different breast cancer intrinsic subtypes
SOURCE: Eroles et al. 2012

Luminal A and Luminal B are ER-positive cancers that account for about 75% of all breast cancers (Eroles et al. 2012). Luminal A breast cancers have high levels of ER and PR expression, but relatively low levels of pro-proliferative genes and HER2 expression. Patients belonging to these subtypes generally have good response to endocrine therapy targeting estrogen pathway such as hormonal aromatase inhibitors (AIs)(Smith and Dowsett 2003) and selective estrogen receptor modulators (SERMs)(Guarneri and Conte 2009), but tend to be less sensitive to chemotherapy (Clarke, Tyson, and Dixon 2015). They have less aggressive tumor progression, low histologic grades and good prognosis (~27.8% relapse rate) (Kennecke et al. 2010). Compared to Luminal A, Luminal B breast cancers have more aggressive phenotypes, higher histologic grades, poorer prognosis and higher rate of resistance to endocrine therapy. Many Luminal B breast cancers have higher expression of pro-proliferative genes than Luminal A breast cancers, including MKI67, cyclin B1, EGFR and HER2. However, the mechanisms leading to the more aggressive disease progression and worse endocrine therapy response of Luminal B are not well understood. HER2-enriched breast cancers highly express HER2 and HER2 pathway-related genes, and have frequent HER2 amplification on chromosome 17q12. Though anti-HER2 treatment such as trastuzumab has been proven to significantly improve patient survival in more than 33% of the cases (Slamon et al. 2001; Piccart-Gebhart et al. 2005), this breast cancer subtype still has poor prognosis and high recurrence rate (Ross et al. 2009). Breast cancers with HER2 overexpression can be further categorized into additional subtypes based on the expression of 158 additional marker genes (Staaf et al. 2010). Basal-like, Normal breast-like and Claudin-low breast cancers all lack ER, PR and HER2 expression, so as to be called triple-negative breast cancers (TNBCs)(Foulkes, Smith, and Reis-Filho 2010). Basal-like breast cancers usually express genes present in normal breast myoepithelial cells. Normal breast-like subtype shows characteristic expression of genes present in adipose tissue. Claudin-low subtype has low expression of genes related to tight junctions and intercellular adhesion. Patients belonging to

Basal-like and Claudin-low subtypes generally have worst prognosis among all breast cancer subtypes. Due to the lack of knowledge of the molecular basis and dominant signaling pathways in TNBCs, there is no well-defined therapeutic target or treatment strategy (Foulkes, Smith, and Reis-Filho 2010). Many of the patients from TNBC subtypes are associated with BRCA1 mutations that lead to deficiency in double-stranded DNA repair and increased genome instability (Reis-Filho and Tutt 2008). Therefore, a promising therapeutic strategy is to use poly (ADP-ribose) polymerases (PARPs) inhibitors to block single-stranded DNA repair so as to fully inactivate all DNA repair machinery in the cells, which results in accumulation of DNA breakage and cancer call death (Farmer et al. 2005; Fong et al. 2009). Several PARP inhibitors are being tested in different stages of clinical trials used as monotherapy (Kaufman et al. 2015; Sandhu et al. 2013) or in combination with other drugs (Lee et al. 2014; Isakoff et al. 2010), and some clinical studies have shown promising response rates and acceptable toxicity (Livraghi and Garber 2015).

1.1.2 Endocrine therapy for estrogen receptor (ER) positive breast cancer

For the majority of breast cancers that are ER positive, tumor growth heavily relies on estrogen, a steroid hormone that is important for maintaining female reproductive function (Further elucidation of estrogen signaling will be provided in details in Section 1.3 of this chapter.). Endocrine therapy targeting estrogen pathway is widely applied clinically and have been proven to be quite effective (Early Breast Cancer Trialists' Collaborative 1998; Goss et al. 2003; Howell et al. 2002). The two major endocrine therapeutic strategies are to antagonize estrogen by competing it to bind to ER and to deprive estrogen or ER level. SERMs and ovarian ablation/suppression are standard therapies for premenopausal women (Osborne and Schiff 2011). SERMs such as tamoxifen act as ER antagonists and inhibit ER-mediated transcriptional regulation on specific target genes

(Dutertre and Smith 2000). In postmenopausal women, AIs are used to inhibit estrogen biosynthesis (Simpson et al. 1994) and selective estrogen receptor degraders (SERDs) such as fulvestrant are applied to down-regulate ER level (Wakeling and Bowler 1992).

Despite of the various options of endocrine therapy in neoadjuvant (before surgery) and adjuvant (post surgery) settings, resistance to endocrine therapy is observed at all cancer progression and treatment stages with different frequencies. At the neoadjuvant stage of the treatment, only 50-70% of the patients are sensitive to endocrine therapy (Miller et al. 2009). For adjuvant treatment, 10-15% of the early-stage patients have tumor recurrence within 5 years after surgery, and the recurrence rate increases to 30% within 15 years after surgery (Dowsett et al. 2010; Early Breast Cancer Trialists' Collaborative 2005). The majority of advanced- and metastatic-stage patients eventually show resistance to endocrine therapy within a short frame of time (Clarke, Tyson, and Dixon 2015). Clinical observations suggest the existence of multiple mechanisms contributing to resistance to endocrine therapy, but without more in-depth studies it is hard to predict patient response based on current diagnosis. Endocrine therapy resistance can be either de novo or acquired during treatment, and loss of estrogen dependency is not necessarily correlated with loss of ER expression (Clarke, Tyson, and Dixon 2015; Osborne and Schiff 2011).

A number of mechanisms are implicated in the resistance to endocrine therapy. First of all, factors modulating ER (including ER variants) and estrogen signaling are closely associated with endocrine therapy resistance. Changes in expression of ER and ER isoforms can directly affect hormone sensitivity (Encarnacion et al. 1993; Shi et al. 2009). Levels of co-activators, co-repressors and interacting transcription factors (TFs) of ER could affect ER-mediated transcriptional regulation (Osborne et al. 2003; Lavinsky et al. 1998; Zhou et al. 2007). In addition, alterations of various intracellular signaling pathways could change the activity of both ER and ER coregulators via post-translational modifications (PTMs) (Musgrove and Sutherland 2009). Secondly, cell growth may

bypass the dependence on estrogen signaling through the activation of alternative growth factor receptor pathways. Many tumors with loss of ER expression show elevated expression of HER2, suggesting a switch of the driving pathway for cell survival a7nd proliferation from ER to HER2 (Osborne and Schiff 2011). Other pathways like insulin, fibroblast growth factor (FGF), vascular endothelial growth factor (VEGF), Src and AKT are also implicated in the resistance to endocrine therapy (Arpino et al. 2008; Chakraborty, Welsh, and Digiovanna 2010; Morgan et al. 2009). Third, some resistance to endocrine therapy is related to the expression of cell cycle genes, for example, overexpression of positive regulators of cell cycle and negative regulators of apoptosis (Span et al. 2003; Butt et al. 2005; Chu, Hengst, and Slingerland 2008; Perez-Tenorio et al. 2006). Fourth, tumor microenvironment may be an important regulator of cell sensitivity to hormone therapy, such as stromal cells, extracellular matrix (ECM), growth factors, cytokines, hypoxia and acidity (Generali et al. 2006; Helleman et al. 2008).

To summarize, breast cancers are highly heterogeneous across different subtypes in terms of tumor origin, tumor progression, histologic grade, hormone dependency and prognosis, which is largely related to subtype-specific gene regulation and driving signaling pathways. Even for the majority of the breast cancers that are estrogen-dependent, patients show heterogeneous response to endocrine therapy, which is largely due to dysregulation of estrogen signaling at different levels. Hence, it is crucial to fully understand how estrogen signaling regulates gene expression and cell growth, and how different factors influence estrogen action in cancer cells in order to elucidate breast cancer heterogeneity.

1.2 Nuclear receptors (NRs) in health and disease

Estrogen is a pivotal hormone regulating the growth of majority of breast cancers, and it exerts its function mainly through estrogen receptors (ERs), which belong to nuclear receptor superfamily. In order to better understand estrogen-dependent cellular response, here, a brief introduction to NRs including their function, category, structure and mode of gene regulation is given in this section.

Lipophilic hormones are important regulators involved in many biological processes, such as embryonic development, homeostasis, cell differentiation, metabolism, reproduction and organ physiology (Mangelsdorf et al. 1995; Novac and Heinzel 2004; Jagannathan and Robinson-Rechavi 2011; Sonoda, Pei, and Evans 2008; Xu 2015). Dysregulation of lipophilic hormone signaling pathways is closely associated with the pathology of hormone-related diseases including hormone-related cancers (e.g. breast, prostate and ovary), diabetes, obesity, rheumatoid arthritis and asthma (Tenbaum and Baniahmad 1997). The intracellular action of lipophilic hormones is mainly mediated by nuclear receptors that form a superfamily of specialized transcription factors (TFs) sensing signal of lipophilic ligands, spanning the evolution of metazoan species (Mangelsdorf et al. 1995). Not only do NRs act as regular TFs that are able to recognize and bind to specific regulatory DNA sequences to activate/repress target gene transcription, but their activity is also directly modulated by the specific interaction with lipophilic ligands, making them important mediators linking extracellular stimuli to transcriptome response.

There are 48 NRs identified in human, among which, 24 NRs have known endogenous ligands. Human NRs can be classified into three major subtypes based on ligand type and physiological function (**Figure 1**)(Sonoda, Pei, and Evans 2008). The first subtype "endocrine

receptors" contains well-characterized NRs with high affinity to steroid hormones, thyroid hormones, vitamin A and D. They are closely related to various signaling pathways in homeostasis and cancer, which makes them one of the favorite targets in drug discovery. For example, the first human NR to be cloned, glucocorticoid receptor (GR), is ubiquitously expressed in all tissue types and has an anti-inflammatory, anti-proliferative, pro-apoptotic and anti-angiogenic role (Weinberger et al. 1985). Synthetic glucocorticoids are widely applied to treat asthma, skin infection, cancer and immunosuppression in organ transplantation (Kadmiel and Cidlowski 2013). The second subtype "adopted orphan receptors" typically has low affinity to dietary lipids and xenobiotics. They are promising drug targets in metabolic diseases. For instance, serving as oxysterol sensors, liver X receptors (LXRs) are crucial for cholesterol homeostasis (Tontonoz and Mangelsdorf 2003). In contrast, Farnesoid X receptors (FXRs), sensors of bile acids that are the end products of hepatic cholesterol catabolism, counteract the function of LXRs in cholesterol and triglyceride metabolism (Kalaany and Mangelsdorf 2006). The third group "orphan receptors" was identified by sequence homology, but no ligand has been found for their ligand-binding domain (LBD). Structural analysis suggest the pocket-like structure of their LBD may be too small to host lipophilic ligands (Ingraham and Redinbo 2005). Their activity, instead, relies on coregulator availability, expression level and post-translational modifications. They also play important roles in metabolic regulation based on genetic studies. One example is HNF4α. HNF4α is essential for mouse embryonic development. Conditional knockout of HNF4α in mouse pancreatic β-cells results in insulin resistance (Miura et al. 2006). A number of studies have suggested that HNF4α is a key regulator for hepatic gene expression including genes involved in glucose, cholesterol and fatty acid metabolism (Miura et al. 2006; Iwayanagi, Takada, and Suzuki 2008; Jump 2004). Since endocrine receptors and adopted orphan receptors have provided many opportunities in drug discovery for a number of hormone-

related diseases, there is an increasing need for deepening our knowledge of all members of the NR superfamily especially orphan receptors.

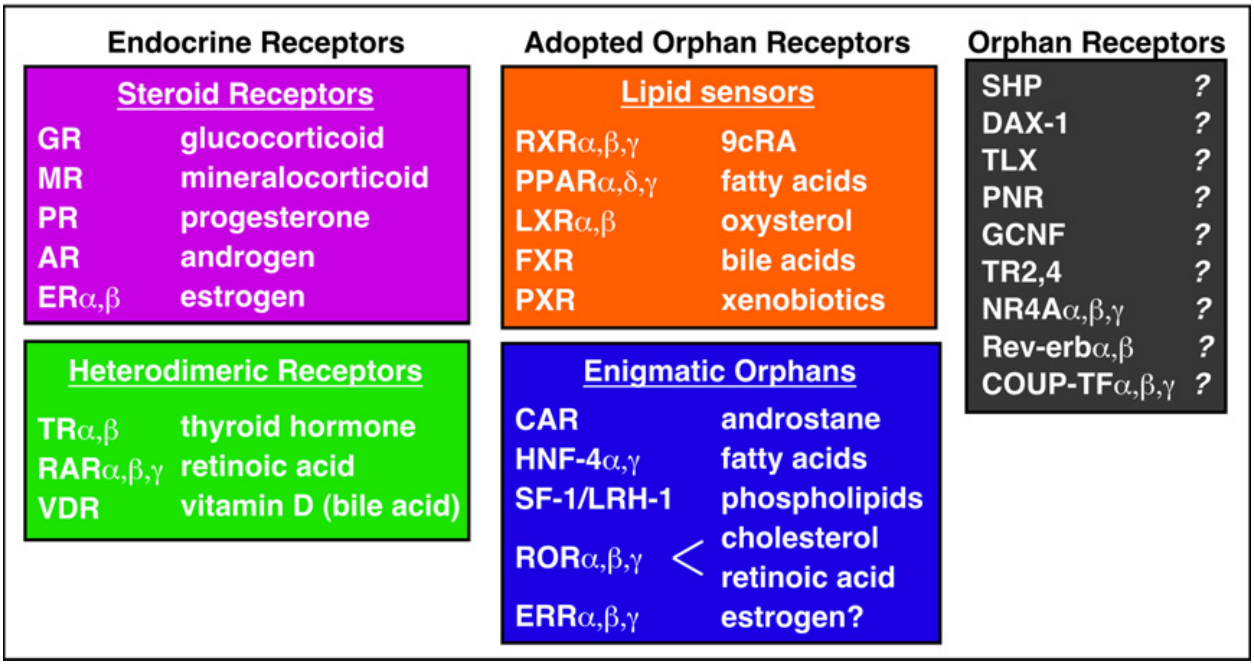

Figure 1. Human nuclear receptor superfamily
SOURCE: Sonoda, Pei, and Evans 2008. Representative natural ligands are shown at right. Purple: homodimerizing steroid receptors. Green: RXR heterodimers. Orange: adopted orphan receptors sensing lipids and xenobiotics. They were originally identified using sequence homology as orphan receptors, but were "de-orphanized" later because of identification of corresponding ligands. Blue: enigmatic adopted orphans. The ligand-dependent regulation in physiology has not been established for this group of NRs. Grey: orphan receptors.

Nuclear receptor family members share common core structural domains (**Figure 2**). From N-terminal to C-terminal, there are ligand-independent activation function 1 (AF1, or A/B), DNA-binding domain (DBD), hinge region, and ligand-binding domain (LBD) that includes ligand-dependent activation function 2 (AF2) (Gronemeyer, Gustafsson, and Laudet 2004). The interaction between NRs and regulatory DNA sequences (also called response elements) is achieved via the evolutionarily highly conserved DBD. The DBD consists of two zinc-finger motifs commonly shared throughout the entire family except for two members, and each zinc finger recognizes response elements derived from a canonical RGGTCA (R represents purine) sequence motif that can be arranged in different ways (e.g. direct repeat, inverted repeat and different spacing). The ligand regulation on NR activity is conducted through the less conserved LBD yet the overall 3D structure is shared by all NRs (Wurtz et al. 1996). The function of the LBD is not limited to ligand binding. In addition, it mediates dimerization between NRs (Tamrazi et al. 2002), interaction with heat shock proteins (Smith 1993) and ligand-dependent transactivation activity (Wurtz et al. 1996). The signature feature of LBD is the pocket-like structure formed by 12 helices, with 11 of them forming a compact cavity for accommodating the ligand and the most C-terminal helix (AF2 motif) acting as a lid over the pocket. The allosteric orientation of the most C-terminal helix is crucial for the transactivation function and is regulated through ligand-dependent conformational change. Such conformational change is not only different between unliganded and liganded states, but also different between agonist and antagonist binding, which is critical for ligand-induced NR activation and dimerization (Gronemeyer, Gustafsson, and Laudet 2004). Unlike the DBD and LBD, AF1 is the least conserved domain across NR family members both in length and in sequence composition (Dahlman-Wright and McEwan 1996). It interacts with coregulators and other TFs in ligand-independent manner, and cooperatively works with AF2 to achieve transcriptional regulation. The activity of AF1 can be regulated by multiple signaling pathways through PTMs (Warnmark et al.

2003; Onate et al. 1998). Last but not the least, the hinge region between DBD and LBD for many NRs has nuclear localization signal that is important for nuclear translocation of NRs (Gronemeyer, Gustafsson, and Laudet 2004).

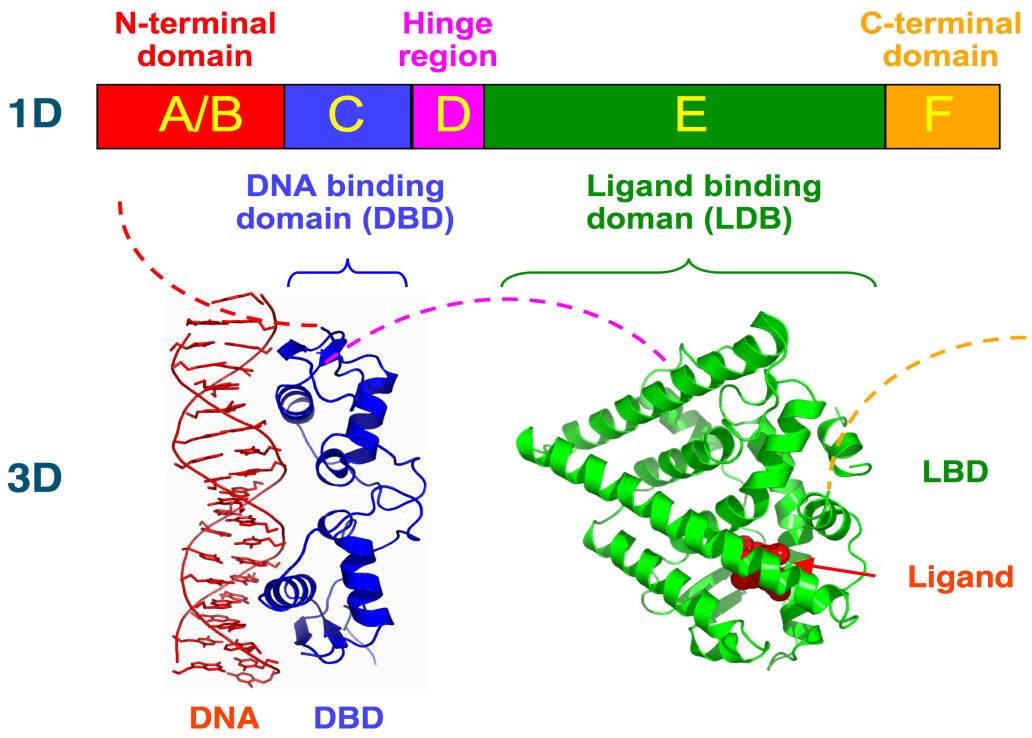

Figure 2. Common structure of nuclear receptor
SOURCE: Wikipedia "Nuclear Receptor" page.

Depending on ligand binding, response element and dimerization mode, NRs can also be classified into four categories (Mangelsdorf et al. 1995; Jagannathan and Robinson-Rechavi 2011). Type I NRs are steroid hormone receptors, which form homodimers and bind to response elements consisting of two half-sites arranged as inverted repeats. Representative members in this category are GRs, ERs, androgen receptors (ARs) and PRs. Type II NRs function by heterodimerizing with retinoid X receptor (RXR) and binding to response elements mostly arranged as direct repeats (Stunnenberg 1993). Retinoic acid receptors (RARs), thyroid hormone receptors (TRs) and RXR are all in this category. Similar to Type I receptors, Type III receptors form homodimers, but they primarily bind to response elements in direct repeats. Unlike the other types of NRs, Type IV receptors bind as monomers to half-site of response elements. Most orphan receptors are in Type III and Type IV categories. In the absence of ligands, many NRs except steroid hormone receptors reside on condensed chromatin by binding to co-repressors such as NCOR1 and SMRT that further recruit histone deacetylatases (HDACs) to achieve gene silencing (Chen and Evans 1995; Horlein et al. 1995; Kurokawa et al. 1995). Once lipophilic ligands diffuse into cells, NRs dissociate from co-repressors, and then bind to co-activators such as p160 that recruits histone acetyl transferases (HATs) and chromatin remodeling complexes to activate target gene transcription (Nagy and Schwabe 2004; Li, Lambert, and Xu 2003). Unliganded steroid hormone receptors are often inactive and coupled with chaperones like heat shock proteins (Pratt and Toft 1997). Upon activated by ligand binding, they release chaperones and translocate into nucleus, then dimerize and recruit coregulators to response elements to activate/repress target gene transcription (Gronemeyer, Gustafsson, and Laudet 2004; Novac and Heinzel 2004).

1.3 Estrogen (E2) signaling

1.3.1 Overview of ERs: ligands, structures and variants

Estrogen is a steroid hormone that primarily regulates the female reproductive system and secondary sex characteristics, while also playing a role in certain male reproductive functions, such as sperm maturation. Endogenous estrogen is synthesized from cholesterol in all vertebrates via steroidogenesis process that is regulated by numerous aromatases. There are three major natural endogenous estrogens in women: estrone (E1), estradiol (E2) and estriol (E3) (Additionally, estetrol (E4) is only produced during pregnancy in fetal liver.). E2 is the most abundant and predominant estrogen in non-pregnant premenopausal women, produced from ovary. In postmenopausal women, E2 is synthesized in non-gonadal sites such as breast, brain and adipose tissue converted from testosterone and androstenedione (Gruber et al. 2002). The cellular effects of estrogen are mainly mediated by two estrogen receptors: ERα (encoded from ESR1 gene)(Walter et al. 1985) and ERβ (encoded from ESR2 gene)(Mosselman, Polman, and Dijkema 1996). Since they are expressed in various tissues including uterus, ovary, breast, kidney, bone, adipose tissue, liver, central nervous system (CNS), cardiovascular system, lung, colon and immune system, estrogen is highly relevant to many hormone-related cancers (e.g. breast, ovary, prostate and colon) and metabolic diseases (e.g. obesity, diabetes, insulin resistance, adipose inflammation and dysfunction of glucose homeostasis)(Heldring et al. 2007).

Though genes encoding ERα and ERβ are located on chromosomes 6q25.1 and 14q23.2, respectively, ERs share similar functional domains in structure like other NRs (**Figure 2**). They firstly share a highly conserved DBD (97% amino acid similarity) so that they both recognize

estrogen response elements (EREs). There is medium level of conservation in LBD (59% amino acid similarity) between ERα and ERβ, leading to subtype-specific ligand binding capacity despite the fact that they both bind to E2. ERα and ERβ only share 16% amino acid similarity in AF1 domain, which accounts for the major functional difference of ER subtypes in gene transcriptional regulation (Thomas and Gustafsson 2011). In fact, it has been shown that ERα and ERβ tend to have opposite effects on gene regulation, and ERβ is significantly decreased in breast cancer cells compared to adjacent normal cells. ERβ has repressive effects on expression of 70% of ERα regulated genes involved in proliferation and metabolism in ER-positive T47D cells (Williams et al. 2008). Both ERα and ERβ have several isoforms, and they have also been shown to play important roles in estrogen-related diseases. For instance, ERαΔ3 does not have DBD due to lack of the third exon. It acts as a dominant negative regulator of ERα, which is usually decreased in breast cancer (Wang and Miksicek 1991; Iwase et al. 1998). ERα36 lacks both AF1 and AF2 domains but binds to regulatory DNA sequences similar to wild type ERα, functioning as an ERα competitor and correlating with breast cancer carcinogenesis (Wang et al. 2005; Zhang and Wang 2013). ERα46, which does not have AF1 domain, serves as a negative regulator of breast cancer by inhibiting cell proliferation (Flouriot et al. 2000; Klinge et al. 2010). Since ERα and E2 play predominant roles in the majority of breast cancers, the later discussion on ER and estrogen all refers to ERα and E2, respectively.

1.3.2 Estrogen action in the cell: genomic and nongenomic pathways

Our gained knowledge of estrogen signaling reveals complex and extensive intracellular response triggered by estrogen, in addition to the canonical model in which estrogen directly

regulates transactivation of ER on estrogen response elements (EREs) to activate or repress gene transcription. ER not only directly affects transcriptional regulation machinery, but also has reciprocal signaling crosstalk with many other intracellular signaling pathways through PTMs. ERs can function in both ligand-dependent and ligand-independent manner. In addition, besides ER, another non-NR family member, GPR30, has been identified to be a receptor for estrogen and involved in mediating estrogen signaling in the cells (Carmeci et al. 1997).

As shown in **Figure 3**, it is known that ER can mediate cellular response in the following ways (Bjornstrom and Sjoberg 2005; Marino, Galluzzo, and Ascenzi 2006; Hayashi and Yamaguchi 2008; Hah and Kraus 2014; Thomas and Gustafsson 2011).

1) Ligand-dependent direct binding to response elements: unliganded ERs are often coupled with chaperones primarily distributed in nucleus but also localized in cytoplasm. Ligand induces conformational change of ERs so that ERs disassociate chaperones and translocate from cytoplasm to nucleus (Pratt and Toft 1997; Lu et al. 2002). In nucleus, ligand-activated ERs form homodimers and bind to estrogen response elements (EREs) (Kumar and Chambon 1988). ERs recruit co-activator (CoA) or co-repressor (CoR) complexes including other regulatory proteins, such as histone acetyl transferases (HATs) and ubiquitin ligase (UL), and interplay with chromatin remodeling machinery, to cooperatively activate or repress gene transcription.

2) Ligand-dependent indirect binding to cis-regulatory elements: in addition to direct binding to EREs, ERs can form heterodimers with other TFs to act on ER-half sites, or be tethered to other TF binding sites without presence of EREs, such as AP-1 and Sp1 sites (Kushner et al. 2000; Safe 2001).

3) Ligand-independent ER activation: growth factor (GR) signaling such as MAPK and PI3K pathways are able to phosphorylate and activate ERs without estrogen stimulation (Levin

2003; Driggers and Segars 2002). Phosphorylated ERs exert downstream transcriptional regulation in ways similar to the above described.

4) Membrane-associated ERs triggered nongenomic signaling: some ERs are localized on cell membrane coupled with SRC, PI3K and G proteins (GPs). Estrogen can trigger PTM-based signal transduction through membrane ER associated signaling proteins to activate ERK and AKT, which further activate their own downstream TFs and influence transcription readout. Such nongenomic action happens very rapidly upon estrogen stimulation within a few minutes (Migliaccio et al. 1996; Castoria et al. 2001).

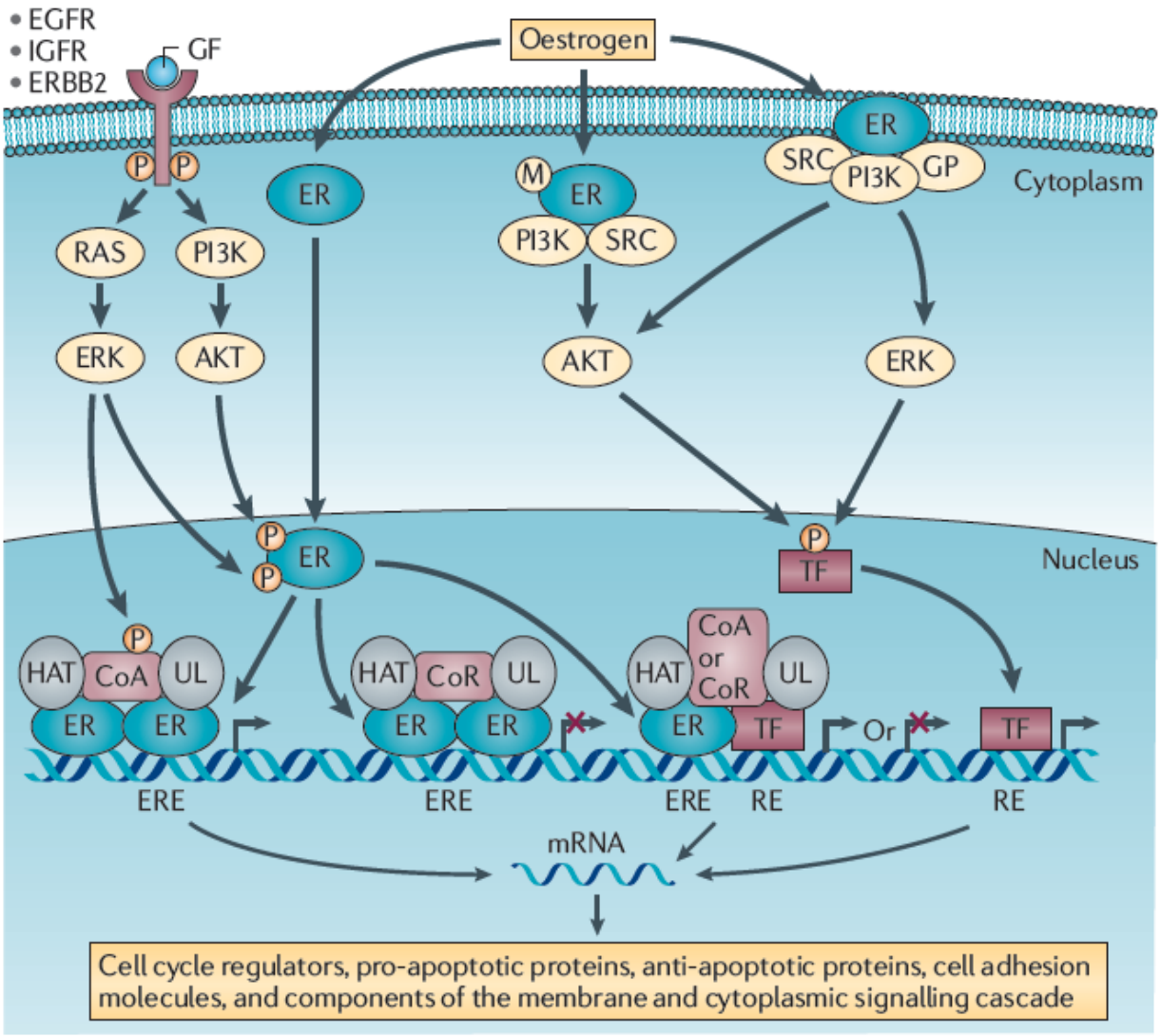

Figure 3. Estrogen signaling through ERs

SOURCE: Thomas and Gustafsson 2011. EREs: estrogen response elements. RE: response elements. HAT: histone acetyl transferase. UL: ubiquitin ligase. CoA: co-activator. CoR: co-repressor. P: phosphorylated. M: methylated. GPs: G proteins. EGFR: epidermal growth factor receptor. ERBB2: HER2. IGFR: insulin-like growth factor receptor. GFs: growth factors.

Besides ER-mediated estrogen signaling, it has been demonstrated that a seven-transmembrane (7TM) G-protein coupled receptor (GPCR) family member, GPR30, is sufficient to mediate rapid estrogen signaling and initiate a series of downstream cellular and transcriptional response (Prossnitz, Arterburn, et al. 2008). GPR30 is primarily located on endoplasmic reticulum membrane (Revankar et al. 2005). In ER negative cells such as SKBr3 cells, estrogen is found to be able to trigger cellular response through GPR30 (Filardo et al. 2000). The estrogen binding affinity for GPR30 (Kd ~ 3-6 nM) has been found to be much higher than ERα (Kd ~ 0.5 nM) and also specific (Revankar et al. 2005; Thomas et al. 2005). Classical ER antagonists such as tamoxifen and ICI182,780 act as agonists for GPR30, indicating that GPR30 may be involved in endocrine therapy resistance (Filardo et al. 2000; Filardo et al. 2002). Activation of GPR30 can activate downstream MAPK and PI3K signaling pathways to further modulate gene transcription (Filardo et al. 2000; Filardo et al. 2002) distinct from ER-mediated gene regulation (Albanito et al. 2007; Vivacqua et al. 2009). In breast cancer patients, GPR30 expression is positively associated with HER2, tumor size and metastasis (Filardo et al. 2006), suggesting its potential to become a novel therapeutic target in estrogen responsive breast cancer (Prossnitz, Sklar, et al. 2008).

1.3.3 Estrogen modulated transcriptional regulation

1.3.3.1 ER coregulators

There is no precise definition of coregulator. Widely speaking, coregulators are accessory molecules (protein, RNA, etc.) interacting with TFs to facilitate transcriptional regulation. They are involved in all biological processes from transcription to translation (e.g. chromatin remodeling, transcription (initiation, elongation, splicing, termination), RNA processing, translation, etc.). Since

the first coregulator SRC1 (NCOA1) was identified in 1995 (Onate et al. 1995), there are now over 350 NR coregulators recorded in the literature, and this number may still be an underestimate (Lonard and O'Malley 2012). Coregulators often exhibit enzymatic activities or are able to recruit various enzymes involved in acetylation, methylation, phosphorylation and ubiquitination. They are also subjected to post-translational modifications governed by upstream signaling cascades, which makes them integrators of intracellular signaling pathways. Based on the final outcome of gene expression, coregulators can be divided into co-activators (CoAs) and co-repressors (CoRs). Dysfunction of ER coregulators has been suggested to be associated with breast cancer evolution and patient therapeutic response (Lonard and O'Malley 2012; Smith and O'Malley 2004; Manavathi, Samanthapudi, and Gajulapalli 2014).

The majority of ER CoAs bind to ER LBD at LXXLL (X: any amino acid, L: leucine) motif in a ligand-dependent manner (Heery et al. 1997). In general, ER CoAs themselves have histone acetyl transferase (HAT) activity that dissociates histone-DNA interaction to help open up chromatin for trans-acting factors to interact with regulatory DNA (Klinge 2000). There are many different classes of ER CoAs. SRC/p160 family members SRC1, SRC2 and SRC3 directly interact with ER and further recruit secondary CoAs including p300 and CBP to enhancer transcription via chromatin modifications (Chen, Huang, and Stallcup 2000). SRC1 has also been shown to mediate the agonistic function of tamoxifen by interacting with AF1 domain of ER (Webb et al. 1998). Beside p300 and CBP, p160 family members could also recruit ATP-dependent chromatin remodeling complexes such as SWI/SNF in an estrogen-dependent manner (Belandia et al. 2002). Ubiquitin-protein ligases are another important family of ER CoAs. For instance, E6-associated protein (E6-AP) acts as ER CoA in several ways: it regulates the degradation of transcriptional complexes to fine tune ER gene regulation dynamics (Ramamoorthy and Nawaz 2008); it is also able to recruit p300 to ER target gene promoter (Catoe and Nawaz 2011). On the other hand, as a

ubiquitin-protein ligase, E6-AP directly regulates ER protein turnover, which suggests that there is a balance between the CoA and protein degrader roles of E6-AP (Gao et al. 2005). ER CoAs can be not only proteins but also RNA molecules. One typical example is steroid RNA activator (SRA), which interacts with AF1 domain and SRC1 to confer functional selectivity of CoA complexes to liganded-ER (Lanz et al. 1999).

CoRs generally exert their function by recruiting histone deacetylases (HDACs) to condense chromatin so as to inhibit gene transcription (Manavathi, Samanthapudi, and Gajulapalli 2014). NRs like TRs and RARs bind to DNA in heterodimer with RXR in absence of ligands, and recruit CoRs to maintain gene silencing. Ligand-activated NRs release CoRs and recruit CoAs to initiate transcription (Lavinsky et al. 1998). However, this is not the general way of how ER and CoR interact because ER is considered to only bind to DNA when activated. There are many fewer identified ER CoRs than CoAs, and their mechanisms of action are not clearly elucidated. However, it is known that HDAC complex components NCOR1, SMRT and MTA1 can repress ER transactivation in the presence of tamoxifen (Liu and Bagchi 2004). Competition with CoA to bind to ER is another way of blocking transactivation. Examples are REA competing with SRC1 at the ER LBD (Delage-Mourroux et al. 2000) and RIP40 not only competing for CoAs but also recruiting HDACs (Castet et al. 2004). Other factors showing ER CoR function but without clear mechanisms include BRCA1 (Fan et al. 2001, 2002) and prohibitin (PHB) (He et al. 2008).

1.3.3.2 ER genomic interplay with other transcription factors (TFs)

In addition to the interaction with numerous coregulators, ER has widespread genomic crosstalk with other TFs and NRs at transcriptional regulation level, which is usually ligand- and cell type- specific.

It is known that many estrogen-regulated genes fail to be directly linked to EREs and a substantial fraction of estrogen-responsive regulatory elements do not harbor EREs (Kininis and Kraus 2008). So, it is believed that ER interacts with a wide range of TFs/NRs to affect target gene expression. Estrogen and some SERMs are able to activate genes with AP-1 promoters such as IGF-1 (Umayahara et al. 1994) and cyclin D1 (Sabbah et al. 1999) by two mechanisms: 1) liganded-ER indirectly interacts with AP-1 TFs jun/fos through their CoA complexes, in which ER directly interacts with p160 protein while jun/fos directly contact with CBP; 2) SERM-ER-NCoR complex sequester HDACs away from AP-1 sites so that jun/fos can recruit CoAs to activate transcription (Kushner et al. 2000). Similar to AP-1 promoters, GC-rich promoters could also be activated by liganded-ER by interaction between ER and Sp1 protein. ER and Sp1 physically interact at Sp1(N)xERE or Sp1(N)xERE half-site through "tethering" mode (Safe 2001) to regulate target genes such as c-fos (Duan, Porter, and Safe 1998) and cyclin D1 (Castro-Rivera, Samudio, and Safe 2001). Estrogen signaling has anti-inflammatory effect that is opposed to pro-inflammatory effect of NF-κB signaling. For example, estrogen down-regulates IL-6 that is NF-κB up-regulated gene (Boyce et al. 1999). It has been revealed that ER inhibits NF-κB activity through multiple mechanisms: affecting NF-κB upstream signaling complex IκB level, interrupting NF-κB DNA binding by physical interaction and competing/interfering NF-κB binding to its CoAs (Kalaitzidis and Gilmore 2005).

Another group of TFs, which are essential for ER genomic action, are called "pioneer factors". The concept is that pioneer factors are TFs residing on compacted chromatin to increase chromatin accessibility and make chromatin permissive for binding to other TFs (Zaret and Carroll 2011). Pioneer factors are essential for global ER binding or ER binding to particular genomic locations. It has been reported than FOXA1 (Carroll et al. 2005), GATA3 (Theodorou et al. 2013),

PBX1 (Magnani et al. 2011), AP2γ (Tan et al. 2011) and TLE1 (Holmes et al. 2012) can act as pioneer factors of ER. Both FOXA1 and GATA3 are positively correlated with ER expression in breast cancer. As the first identified ER pioneer factor, FOXA1 contains a winged helix domain that is structurally similar to histone H1 and H5, which allows it to have high affinity interaction with histone H3 and H4 to de-compact histone (Cirillo et al. 1998). More than half of ER binding sites coincide with FOXA1 binding sites in ER+ MCF7 cells, and FOXA1 knockdown abrogates ER binding, ER-regulated gene expression and E2-dependent cell proliferation (Hurtado et al. 2011). Similar to FOXA1, GATA3 and PBX1 are also important for ER binding, and they form big complexes like FOXA1-GATA3 (Kong et al. 2011) or FOXA1-PBX1 (Magnani et al. 2011) to co-bind with ER.

A number of genome-wide studies have shown widespread genomic crosstalk between ER and other NRs. In breast cancer, while estrogen serves as a major mitogenic hormone, retinoic acid (RA) is considered to exert anti-proliferative and pro-apoptotic functions. As mediators of estrogen and RA, ER and retinoic acid receptors (RARs) frequently co-localize and antagonistically bind to regulatory elements to regulate many breast cancer-related genes (Hua, Kittler, and White 2009). As a major biomarker for breast cancer, PR is not only an ER up-regulated gene, but also an ER-associated protein and able to recruit ER to particular sites, resulting in gene expression associated with good clinical outcome (Mohammed et al. 2015). Besides ER and PR, GR status in breast cancer is another important indicator of disease outcome. Co-activation of both ER and GR in mouse mammary cells lead to reprogramming of global chromatin landscape. Specifically, induction of GR facilitates ER access to particular sites, while induction of ER in turn results in a number of ER-dependent GR binding sites (Miranda et al. 2013).

To summarize, estrogen-governed gene regulation is achieved via highly complex mechanisms, which can be modulated at different levels and function in ligand- and cell type-specific manner relevant to breast cancer clinical outcome. As the major mediator for estrogen signaling, estrogen receptor widely interacts with coregulators, TFs and NRs at open chromatin regions harboring cis-regulatory elements to regulate gene transcription in an orchestrated way. Understanding how estrogen-triggered transcriptional regulation networks are formed and how different components in these networks are coordinated will deepen our knowledge of the molecular basis for different ER+ breast cancers and shed light upon novel therapeutic strategies.

1.4 Cis-regulatory elements in gene transcriptional regulation

Protein-coding gene expression in eukaryotic cells can be regulated at different levels, including transcriptional regulation, post-transcriptional regulation (e.g. mRNA processing, mRNA transport, etc.), translational regulation and post-translational modification. The regulation of gene transcription, which is the production of RNA from DNA, is a vital process in gene regulation relying on the interaction between trans-acting milieu and cis-acting chromatin using a variety of mechanisms in spatial- and temporal-specific manner. The trans-acting milieu refers to all kinds of diffusible factors (proteins and RNAs) that collectively interplay with specific regulatory regions of chromatin to influence gene transcription. Cis-acting chromatin harbors clusters of cis-regulatory elements that are sequence-specific DNA elements recognized and bound by TFs, and it consists of an assortment of chromatin marks that respond to trans-acting milieu by facilitating or suppressing the interaction between cis-regulatory elements and TFs. Dysregulation of gene transcription machinery is closely associated with various human diseases including cancers, immune diseases,

diabetes, cardiovascular diseases and many more (Lee and Young 2013). Due to the increased genome instability in tumors, mutations targeting key TFs, cofactors, regulatory RNAs and cis-regulatory elements can all contribute to tumorigenesis, tumor progression and patient response to therapy. In breast cancer, for example, a missense mutation of ER, Y537N, leads to hormone-independent constitutive activation of ER, which contributes to breast cancer progression and hormone resistance (Zhang et al. 1997). In recent years, a lot of DNA polymorphisms associated with breast cancer or mRNA levels (expression quantitative trait loci, referred as eQTLs) in breast cancer have been discovered through genome-wide association studies (GWASs) (Michailidou et al. 2013; Easton et al. 2007; Long et al. 2012; Han et al. 2016; Gold et al. 2008). Though their functional roles are largely under-investigated, many of them are non-coding variants and thought to be involved in transcriptional regulation by affecting cis-regulatory elements or regulatory function of noncoding RNAs (Freedman et al. 2011).

As important parts of cis-acting chromatin in eukaryotic cells, cis-regulatory elements are embedded throughout the genome and often organized in a modular fashion, allowing multiple TFs to bind and collectively regulate gene transcription (Yuh, Bolouri, and Davidson 1998). Cis-regulatory elements can be classified into promoters, enhancers, silencers and insulators according to the gene regulation fashion and outcome (Jeziorska, Jordan, and Vance 2009). Promoters are regulatory DNA essential for maintaining basal transcription, which are usually located immediately upstream of transcriptional start sites (TSSs) or proximal to TSSs. The core promoter (usually located between -40 bp to +40 bp of TSS) guides the initiation of protein-coding gene transcription by RNA polymerase II. Core promoters contain motifs that interact with basal TFs and form preinitiation complexes (PICs), including the initiator (Inr), TATA box, TFIIB recognition element (BRE), downstream core promoter element (DPE) and motif ten element (MTE) (Juven-Gershon and Kadonaga 2010). Promoter regions are typically high in G+C dinucleotide (referred as CpG

islands) (Gardiner-Garden and Frommer 1987), and it has been found that cytosine methylation in CpG islands is often correlated with transcriptional silencing (Iguchi-Ariga and Schaffner 1989), which serves as another layer of transcriptional regulation. Enhancers can activate or promote the transcription of target genes independent of orientation and location (can be located more than 100 kb away from target gene TSS), in spatial- and temporal-specific manner. Regulation of gene transcription in multi-cellular organisms is largely achieved through the intricate interaction between enhancers and various sequence-specific TFs. It is popularly proposed that enhancers affect gene transcription by looping to distal promoters so that enhancer-bound TFs and related coregulators are introduced proximal to TSSs to facilitate recruitment of RNA polymerase II (Maston, Evans, and Green 2006; Malik and Roeder 2010). The spatial-temporal regulatory fashion of enhancers not only is dependent on the expression of corresponding TFs, but also relies on whether or not the chromatin landscape is permissive, which in turn is associated with histone modifications and histone protein variants. It has been found that while many promoter regions are lowly occupied by nucleosomes, many functional enhancers have frequent nucleosome repositioning marked by H3.3/H2A.Z double variant (Jin et al. 2009). For example, H3K4me2-marked nucleosomes at a class of androgen-responsive enhancers are shifted upon androgen stimulation, due to displacement of the H3.3/H2A.Z containing central nucleosome by activated TFs (He et al. 2010). Besides the dynamics of nucleosomes, a number of studies mapping all kinds of histone modifications across genome have shown their various associations to enhancer activity and gene expression in cell-type specific and enhancer-specific way (Heintzman et al. 2007; Heintzman et al. 2009; Cui et al. 2009; Barski et al. 2007; Wang et al. 2008). As opposed to enhancers, silencers exert negative regulatory function on gene transcription via binding by transcriptional repressors. Silencers may repress gene transcription by either blocking the function of nearby transcriptional activators (Harris, Mostecki, and Rothman 2005) or by recruiting co-repressors to reprogram chromatin to a non-permissive state

for gene transcription (Privalsky 2004). Finally, insulators are boundary DNA elements that block communication between enhancers or silencers and promoters. They play important roles in transcriptional regulation by preventing regulatory elements from affecting unrelated promoters and maintaining global genome organization (Wallace and Felsenfeld 2007).

Because of the fundamental importance of cis-regulatory elements in gene regulation, the ENCODE Project and Epigenome Roadmap Consortiums have invested tremendous efforts into the generation of comprehensive predictive annotations of cis-regulatory elements located in non-coding regions of the DNA (Consortium 2012; Roadmap Epigenomics et al. 2015). These datasets include, but are not limited to, analyses of TF binding sites, histone modifications and chromatin accessibility regions in various human tissues and cell types. These data are crucial for the identification of cis-regulatory elements and for defining the epigenomic landscape in relation to functional impact on the regulatory programs that control development and disease. However, despite the improvement on data-driven predictive algorithms in linking cis-regulatory elements to their respective target genes, a proportionally small number of predictions have actually been tested. This lack of genome-wide prediction validation through functional activity assays can lead to many false positive predictions (Kwasnieski et al. 2014; Shlyueva, Stampfel, and Stark 2014; Muerdter, Boryn, and Arnold 2015). Furthermore, these datasets can only identify putative cis-regulatory elements based on TF and histone chromatin immunoprecipitation (ChIP-seq) data, and therefore do not necessarily exhaustively define functional cis-regulatory elements, leading to false negatives. With the discovery of thousands of nucleotide variants in these regulatory elements linked to potential phenotypic outcome, these limitations of the current predictive approaches highlight the importance of obtaining genome-wide functional validation of cis-regulatory elements.

Traditionally, regulatory activity of DNA has been quantified using reporter assays under the control of a minimal reporter to measure the regulatory activity of the cloned region of interest.

While this method of quantification is widely accepted and applied, it is primarily used to serially investigate individual candidate sequences, making comprehensive assays of large genomes untenable. Recent advances have improved on this technique by developing enhancer reporter assays that combine massively parallel assays coupled with cell sorting techniques (Gisselbrecht et al. 2013; Dickel et al. 2014; Murtha et al. 2014), molecular barcode sequencing (Patwardhan et al. 2012; Kheradpour et al. 2013; Smith et al. 2013), or by coupling the activity of a candidate enhancer to the abundance of its self-transcribed mRNA by placing candidate regions downstream of a minimal promoter and into the 3'UTR of a reporter gene (Arnold et al. 2013). The latter method, known as Self-Transcribing Active Regulatory Regions Sequencing (STARR-Seq), has allowed the assessment of enhancer activity on a genome-wide scale in drosophila genomes and in selected regions of the human genome (Arnold et al. 2013; Arnold et al. 2014; Shlyueva et al. 2014; Muerdter, Boryn, and Arnold 2015; Zabidi et al. 2015).

To summarize, cis-regulatory elements are important regulatory DNA sequences in chromatin controlling gene transcription by interacting with trans-acting factors. The spatial- and temporal-specific interaction between cis-regulatory elements and trans-acting factors is realized given the chromatin states defined by nucleosome position and epigenetic modifications. The development of technology brings to us more and more power to systematically identify and functionally characterize cis-regulatory elements, which allows us to better elucidate the genetic and molecular basis of diseases associated with functional noncoding regions of human genome.

CHAPTER 2

Genomic and functional analysis of estrogen responsive cis-regulatory elements reveals estrogen triggered coordinated interplay between various TFs and chromatin

2.1 Introduction

Estrogen (E2) signaling is one of the driving pathways for cell survival and proliferation in ER+ breast cancers, which account for ~75% of total breast cancer cases. Triggered by estrogen stimuli, estrogen receptors (ERs) undergo conformational change, and post-translational modification (PTM)-based intracellular signal transduction cascades are initiated involving ERs and many other signal transducing proteins (e.g. PI3K, AKT, ERK, etc.). Activated ERs are translocated into nucleus and interact with cofactors (e.g. SRC/p160 family members, p300/CBP, E6-AP, etc.), other TFs and chromatin remodeling machinery (e.g. SWI/SNF) at cis-regulatory elements (CREs) to coordinately regulate target gene transcription. Therefore, to decipher E2-governed transcriptional regulation network is not only important for illustrating E2-dependent tumor evolution, but also informative for understanding the underlying mechanisms of resistance to endocrine therapy and developing novel therapeutic strategy. In the past decade, a number of genomic studies have made many achievements in identifying genes regulated by E2 or E2-related ligands, genomic recruitment of ER and ER coregulators, and ER transcriptional regulation crosstalk with other TFs. Using ER+ breast cancer cell lines such as MCF7 and T47D as models,

thousands of E2-regulated genes have been identified (Carroll et al. 2005; Fullwood et al. 2009; Welboren et al. 2009; Hurtado et al. 2011; Theodorou et al. 2013). Though the numbers of E2-regulated genes from different studies or based on different methods vary a lot, and by using protein synthesis inhibitor it is indicated that only 20-30% of E2-regulated genes are E2 direct targets (Lin et al. 2004), it has been shown that E2 triggers very rapid transcriptional response within the time frame of a few minutes and has impact on ~25% of the transcriptome (Hah et al. 2011). A number of chromatin immunoprecipitation (ChIP)-based methods have successfully mapped genomic occupancy of liganded ER (Carroll et al. 2005; Fullwood et al. 2009; Welboren et al. 2009; Hurtado et al. 2011; Theodorou et al. 2013). Despite the differences in mapped ER binding sites among different studies, these datasets have revealed that many ER binding sites tend to be distal to nearest TSSs (>50 kb) but are still positively correlated with the signal of RNA polymerase II loading, suggesting that ER regulates target gene transcription through enhancer-promoter looping mechanism. ChIA-PET data has further supported the mechanism and discovered extensive cis- and trans-chromatin interaction networks mediated by ER (Fullwood et al. 2009). Besides, only a fraction of experimentally profiled ER binding sites contain canonical ERE motif (AGGTCAnnnTGACCT) or ERE half site, suggesting that ER can be recruited to target sites through alternative mechanisms mediated by other TFs or coregulators. A few TF binding motifs have been found to be enriched proximal to ER binding sites, including AP-1, Sp1, FOXA1, OCT1, MYC and so on, indicating widespread interplay between ER and these TFs at cis-chromatin. Loss-of-function experiments have further demonstrated the necessity of some TFs (e.g. FOXA1, GATA3, etc.) in guiding/facilitating ER to load onto cis-regulatory elements (Hurtado et al. 2011; Theodorou et al. 2013). Work led by former colleagues from our group has shown that ER acts as one of the central TFs in breast cancer-associated gene regulation networks and extensive crosstalk occurs between ER signaling and other TF signaling at transcriptional regulation level (Kittler et al.

2013). One example is the genomic antagonism between ER and retinoic acid receptors (RARs), which plays a pivotal role in balancing gene expression in breast cancer (Hua, Kittler, and White 2009).

The fact that ER tends to bind to cis-regulatory elements distal to promoters and the difficulty of distinguishing E2 primary target genes from secondary regulated genes give rise to the need to identify functional E2-responsive cis-regulatory elements. Most studies in the field primarily focus on computationally predicting functional ER binding sites by integrating a variety of data obtained from genome approaches (e.g. gene expression microarray, RNA-seq, ChIP-Chip, ChIP-seq, ChIA-PET, FAIRE-seq, DNase-seq, GRO-seq, etc.). For instance, by investigating open chromatin, histone modifications, FOXA1, AP-1 and RNA Pol II occupancy, ER binding sites have been accurately predicted (Joseph et al. 2010). Recently, transcription of E2-induced enhancer RNA (eRNA) arising from ER binding sites has been discovered and thought to be a novel indicative marker of E2-responsive cis-regulatory elements. It has been proposed that eRNAs may play important roles in strengthening ER-mediated enhancer-promoter looping complexes (Li, Notani, et al. 2013; Hah et al. 2013). However, among the thousands of predicted enhancers, only a tiny fraction have been functionally interrogated using classical methods like reporter assay, due to the limitation of experimental throughput and short of powerful in vivo approaches. Recently, STARR-seq (Arnold et al. 2013) has been developed allowing us to assess regulatory activity of candidate DNA fragments based on target-specific mRNA-seq. In its design, candidate DNA fragments are cloned into 3'UTR region of reporter constructs, so that the transcription of reporter gene would be coupled with the transcription of testing fragments. As a result, regulatory activity of the testing fragments is directly evaluated by the abundance of self-transcribed mRNA. CapSTARR-seq is a variant of STARR-seq, which combines STARR-seq with genomic DNA capturing technology to specifically study regions of interest in the genome in order to increase flexibility, reduce library

preparation labor and sequencing cost (Vanhille et al. 2015). Both STARR-seq and CapSTARR-seq have shown good correlation with conventional approaches and have the potential to conduct massively parallel studies on regulatory regions.

In this chapter, we have systematically profiled the dynamic changes of chromatin accessibility landscape and gene expression in response to E2 in MCF7 cells, and quantitatively assessed regulatory activity of E2-responsive DNase I hypersensitivity sites (DHSs) using high throughput method CapSTARR-seq. We have revealed distinct chromatin signatures and gene regulation dynamics associated with E2-responsive DHSs. Direct interrogation of regulatory activity of over 10,000 candidate DHSs has allowed us to functionally validate the regulatory activity of more than 3000 cis-regulatory elements and E2-responsive regulatory activity of more than 800 cis-regulatory elements. The unbiased identified E2-responsive cis-regulatory elements have uncovered unique chromatin signatures and coordinated action of various TFs participating in E2-governed transcriptional regulation.

2.2 Results

2.2.1 DNase I hypersensitivity sites are correlated with gene expression

We mapped a total of 77,167 highly reproducible accessible chromatin regions in MCF7 cells by profiling DNase I hypersensitivity sites (DHSs) during time-course E2 treatment (0, 45min, 4h, 12h and 24h, IDR < 0.01). Compared to genomic background based on hg19 RefSeq annotation, DHSs are preferentially distributed in genomic areas proximal to transcriptional start sites (TSSs) such as those within 5 kb upstream of TSS and 5'UTR (**Figure 4A**), where potential cis-regulatory elements are expected to be located. The number of nearby DHSs (within 50 kb of TSS) is positively correlated with gene expression level (**Figure 4B**), supporting the concept that DHSs harbor functional cis-regulatory elements modulating nearby gene transcription.

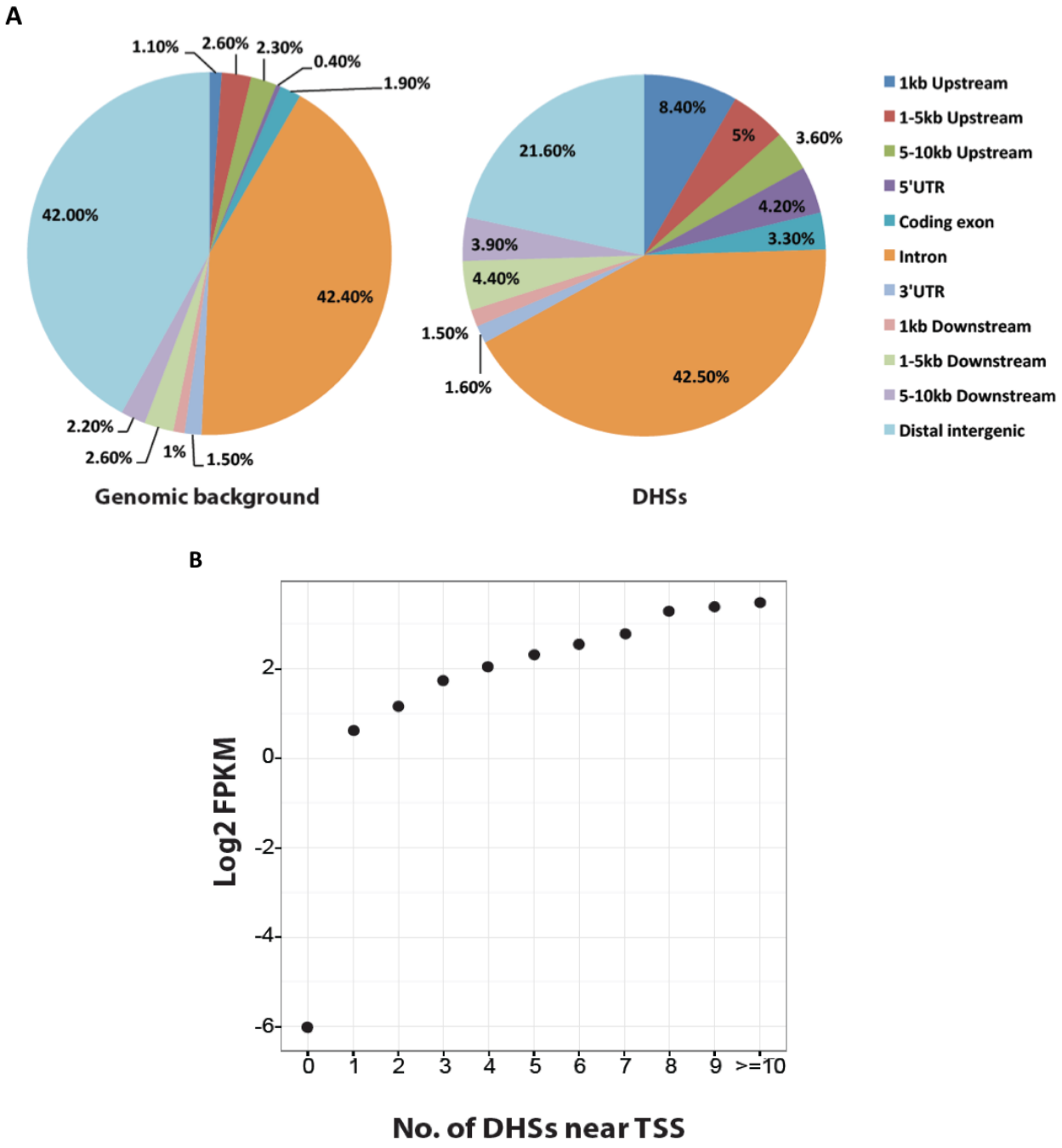

Figure 4. Association between DHSs and nearby genes

A) Genomic distribution of DHSs. Annotation of hg19 RefSeq genes was used as reference to estimate genomic background distribution. B) Correlation of DHSs and nearby gene expression level. The number of DHSs within 50 kb of TSS was counted for each gene. Genes were ranked and binned based on the number of nearby DHSs, and log2 (FPKM) of median ranked gene in each bin was plotted on y-axis.

2.2.2 Identification of E2-responsive DHSs in MCF7 cells

By either comparing E2 to vehicle treatment, or comparing E2 treatment between different time points, we have not observed dramatic changes of global DHS landscape, and we have not found chromatin regions changing from completely inaccessible to accessible, or vice versa, which is consistent with previous FAIRE-seq data (Joseph et al. 2010). Regardless, we detected thousands of E2-responsive DHSs whose chromatin accessibility was quantitatively altered upon E2 treatment. We have identified 2,258 DHSs showing increased chromatin accessibility and 1,193 DHSs showing decreased chromatin accessibility triggered by E2 across all time points (GLM likelihood ratio test on negative binomial model, $p < 0.001$), which we refer to as "E2 constantly induced DHSs" (**Figure 5A and 5D**) and "E2 constantly repressed DHSs" (**Figure 5B and 5D**), respectively. In addition, we identified 2,164 DHSs with increased chromatin accessibility and 2,032 DHSs with decreased chromatin accessibility in response to E2 for at least one time point ("E2 dynamically induced DHSs" and "E2 dynamically repressed DHSs", exact test on negative binomial model, $p < 0.001$, **Figure 5D**). We further classified E2 dynamically induced DHSs into four groups ("45 min max", "4 h max", "12 h max" and "24 h max") based on when the DHS was maximally induced (**Figure 5C**). As control, we randomly sampled 10,000 DHSs which chromatin accessibility remained unaffected by E2 at any time points ("random DHSs").

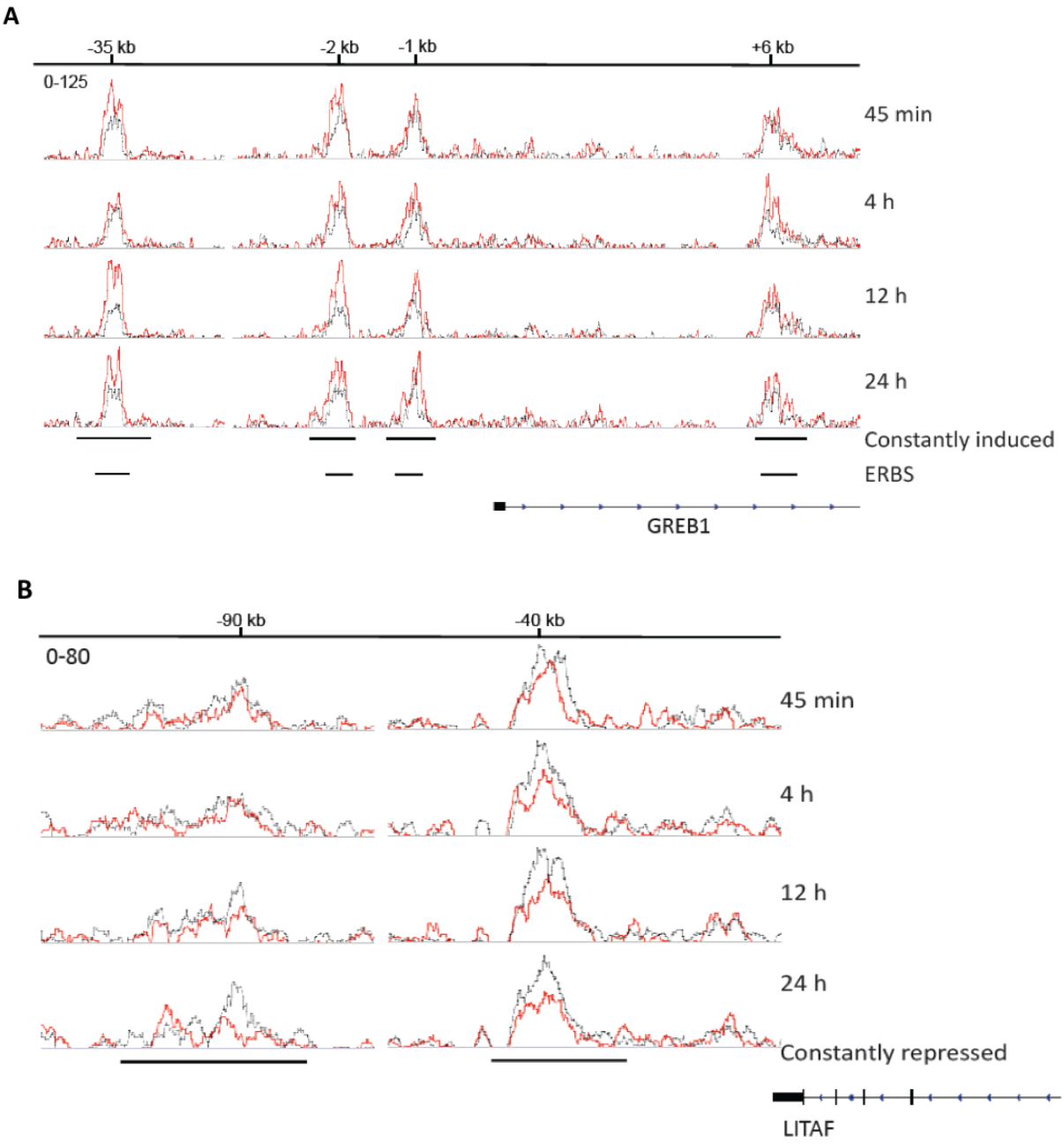

Figure 5. E2-responsive DHSs

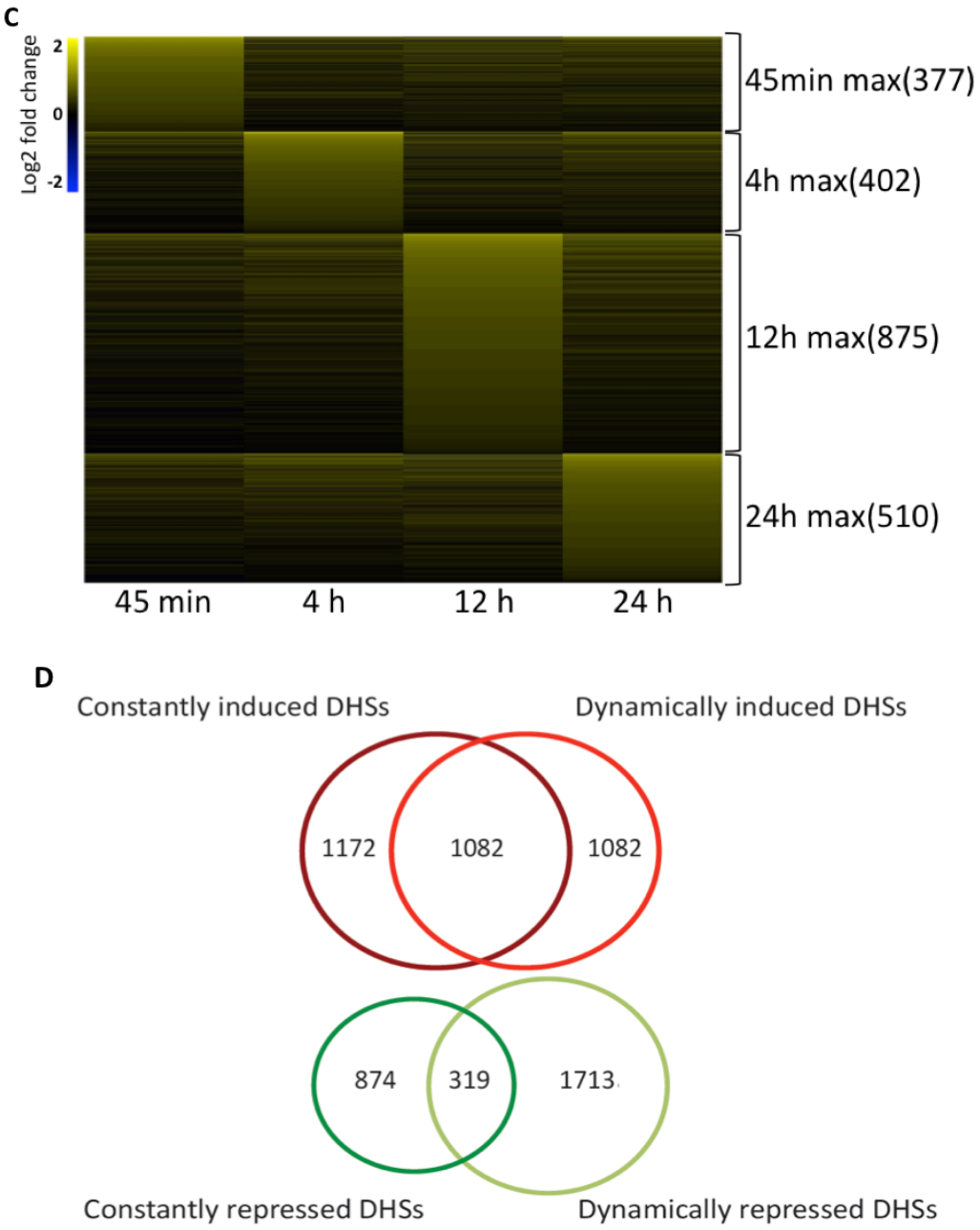

Figure 5. E2-responsive DHSs (continued)
Snapshots of DNase-seq signal on A) E2 constantly induced DHSs and B) E2 constantly repressed DHSs. Data was normalized by library size equivalent to 30 million reads. Red: E2. Grey: vehicle. C) Heat map of log2 fold change of DNase-seq signal in response to E2 on E2 dynamically induced DHSs. Numbers in the brackets denote the number of DHSs in each group. D) Venn diagram of the overlaps between different categories of E2-responsive DHSs.

2.2.3 Characteristics of E2-responsive DHSs are associated with E2-dependent cistrome

2.2.3.1 E2-induced DHSs are highly enriched for ER occupancy

Comparing to genomic data based on ER chromatin immunoprecipitation experiments such as ChIP-Chip, ChIP-seq and ChIA-PET, DHSs represent the global chromatin accessibility landscape and hence provide more unbiased information that allows us to search for E2-responsive cis-regulatory elements irrespective of ER binding. Nonetheless, we still detected a strong signature of ER occupancy in E2-induced DHSs. Compared to random DHSs, E2 constantly induced DHSs are highly enriched for ER binding sites (Fisher's exact $p < 2.2e\text{-}16$, **Figure 6A**), but this pattern is not observed for E2 constantly repressed DHSs (Fisher's exact $p = 0.9365$, **Figure 6A**). When we examined E2 dynamically induced DHSs we found that the enrichment of ER binding sites in DHSs maximally induced at 45min is far higher than those in DHSs maximally induced at later time points (Fisher's exact $p < 2.2e\text{-}16$). More than 70% of DHSs maximally induced at 45 min contain ER binding sites (**Figure 6B**). This is not likely due to the bias that ER ChIP-seq data was generated under the condition of E2 treatment for 45 min, because the same pattern is maintained by using ER ChIP-seq data collected with E2 treatment for 4 hours (**Figure 6C**, source of data: GSE54855). This result indicates that short-term E2 treatment induced chromatin accessibility changes are primarily associated with direct ER binding.

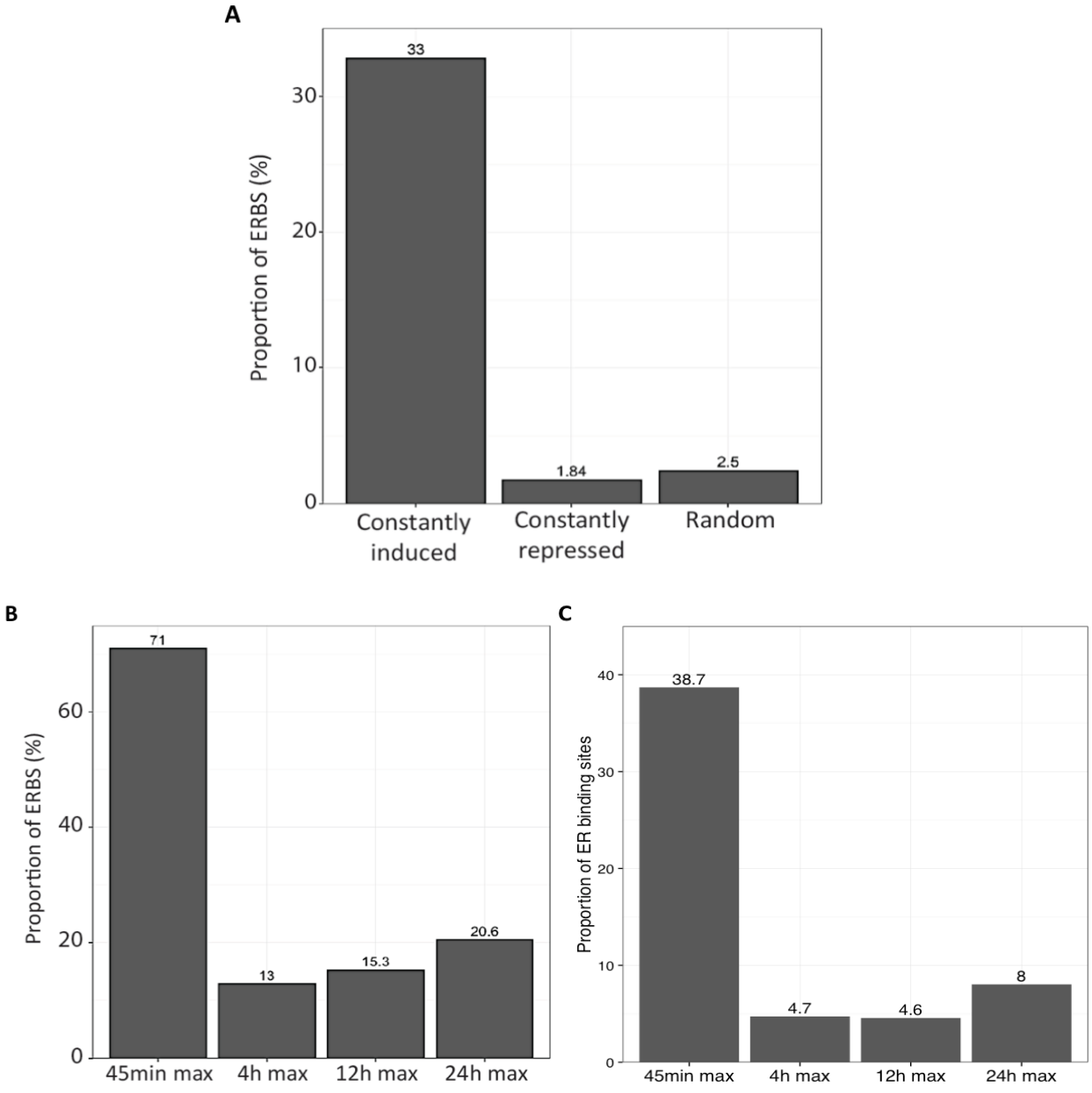

Figure 6. Fraction of ER binding sites in E2-responsive DHSs

A) Proportion of ER binding sites in E2 constantly responsive DHSs. B) Proportion of ER binding sites with E2 treatment for 45 min in E2 dynamically induced DHSs. C) Proportion of ER binding sites with E2 treatment for 4 hours in E2 dynamically induced DHSs.

Note: label on top of each bar represents the percentage of DHSs harboring ER binding sites in each category of DHSs.

2.2.3.2 Distinct RNA Pol II and histone modification signatures of E2-responsive DHSs

Next, we investigated the signal from several genetic and epigenetic marks on E2-responsive DHSs. Both with and without E2 treatment, E2 constantly induced DHSs have strongest signal from active chromatin marks compared to random DHSs and E2 constantly repressed DHSs (Wilcox $p < 2.2e\text{-}16$, **Figure 7A and 7B**). Though E2 constantly repressed DHSs show lower signal of RNA Pol II and H3K9ac compared to random DHSs, they have indistinguishable or even slightly higher signal for other active chromatin marks including H3K4me1, H3K14ac and H3K27ac compared to random DHSs (**Figure 7A and 7B**). Among these marks, RNA Pol II, H3K4me1, H3K9ac and H3K14ac signal intensities are significantly changed in response to E2 widely across all categories of DHSs, but their changes in E2 constantly induced DHSs are stronger than those in the other two categories of DHSs (**Figure 7A**). The signal from repressive marks H3K27me3 and H3K9me3 is low in all three DHS categories, but interestingly, we detected slightly enriched signal of both repressive marks at the center of E2 constantly repressed DHSs (**Figure 7C**). To summarize, E2 constantly induced DHSs are associated with strong basal and E2 responsive signal of active chromatin marks and RNA Pol II. E2 constantly repressed DHSs show signal of active marks at levels similar to random DHSs, while also show slightly enriched signal of repressive marks at the center of DHSs.

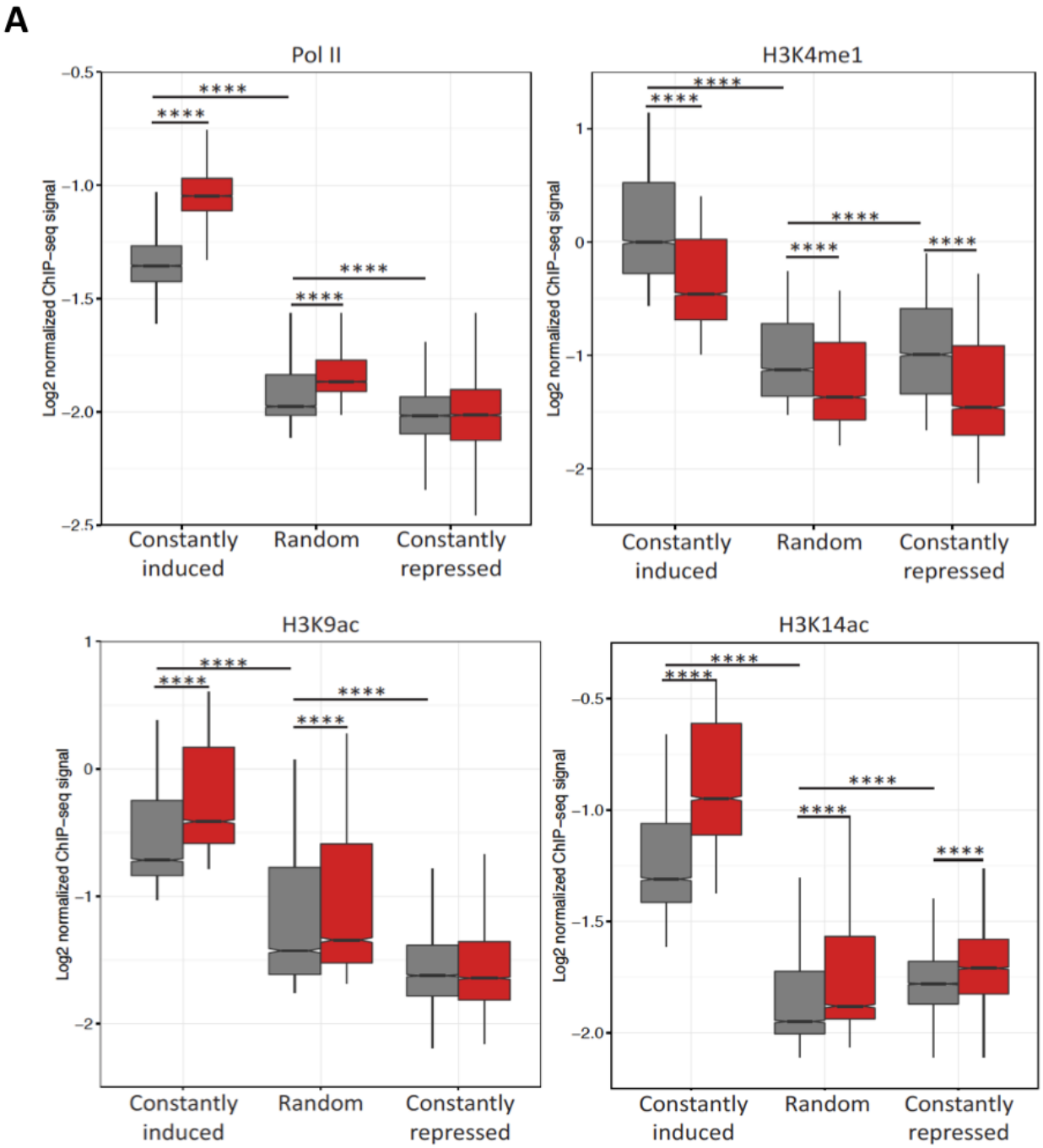

Figure 7. RNA Pol II and chromatin mark signal on E2-responsive DHSs

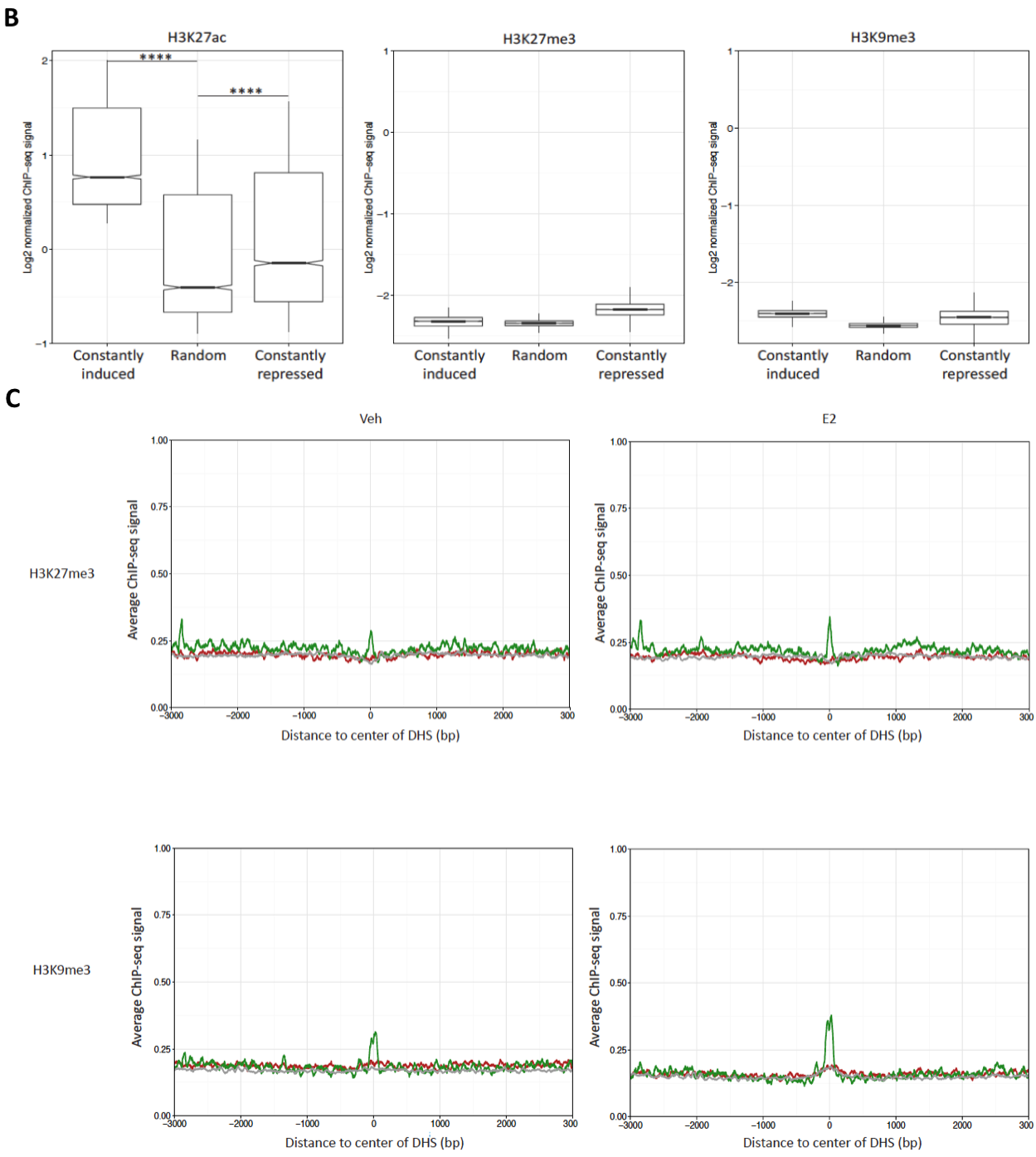

Figure 7. RNA Pol II and chromatin mark signal on E2-responsive DHSs (continued)
A) Box plot of average ChIP-seq signal of Pol II, H3K4me1, H3K9ac and H3K14ac in different categories of DHSs with E2 and vehicle treatment. Red: E2. Grey: vehicle. ****: p <0.0001. B) Box plot of average ChIP-seq signal of H3K27ac, H3K27me3 and H3K9me3 on different categories of DHSs with E2 treatment. ****: p <0.0001. C) Meta plot of H3K27me3 and H3K9me3 average ChIP-seq signal on different categories of DHSs. Red: E2 constantly induced DHSs. Green: E2

Figure 7. RNA Pol II and chromatin mark signal on E2-responsive DHSs (continued)
constantly repressed DHSs. Grey: random DHSs. Note: ChIP-seq tag count was summarized and averaged in +/- 3 kb flanking regions of the centers of DHSs, and then normalized by library size equivalent to 10 million reads.

2.2.3.3 E2-induced DHSs are enriched for HOT regions

Former colleagues in our group have built NR transcriptional regulation networks in breast cancer via mapping genomic binding sites of 24 NRs and 14 breast cancer-associated TFs in MCF7 cells, and defined HOT regions based on TF binding complexity (Kittler et al. 2013). In their study, they have showed that large numbers of TFs co-bind to highly occupied cis-regions. Based on the results of Kittler et al. 2013, regions bound by more than 8 different NR/TF bindings sites are defined as HOT regions. By overlapping HOT regions with E2-responsive DHSs, we found 1298/2258 (57.5%) of E2 constantly induced DHSs harbor at least one mapped binding site of these breast cancer-associated NRs/TFs, which is significantly higher than that of random DHSs (2427/10,000 = 24.3%, one-sided Fisher's exact $p < 2.2e-16$) and E2 constantly repressed DHSs (272/1193 = 22.8%, one-sided Fisher's exact $p < 2.2e-16$), but there is no difference between E2 constantly repressed DHSs and random DHSs (two-sided Fisher's exact $p = 0.2673$, **Figure 8A**). We found a significant higher fraction of HOT regions in E2 constantly induced DHSs (261/1,298 = 20.1%) than those in random DHSs (281/2,427 = 11.6%, one-sided Fisher's exact $p = 3.355e-12$) and E2 constantly repressed DHSs (24/272 = 8.8%, one-sided Fisher's exact $p = 2.199e-6$). Representing only 2.9% (2,258/77,167) of total mapped DHSs in MCF7 cells, E2 constantly induced DHSs harbor ~10% (261/2,628) of identified HOT regions in all DHSs, which demonstrates that E2 constantly induced DHSs are transcriptional regulation hotspots where extensive crosstalk among NRs and TFs occurs (**Figure 8B**).

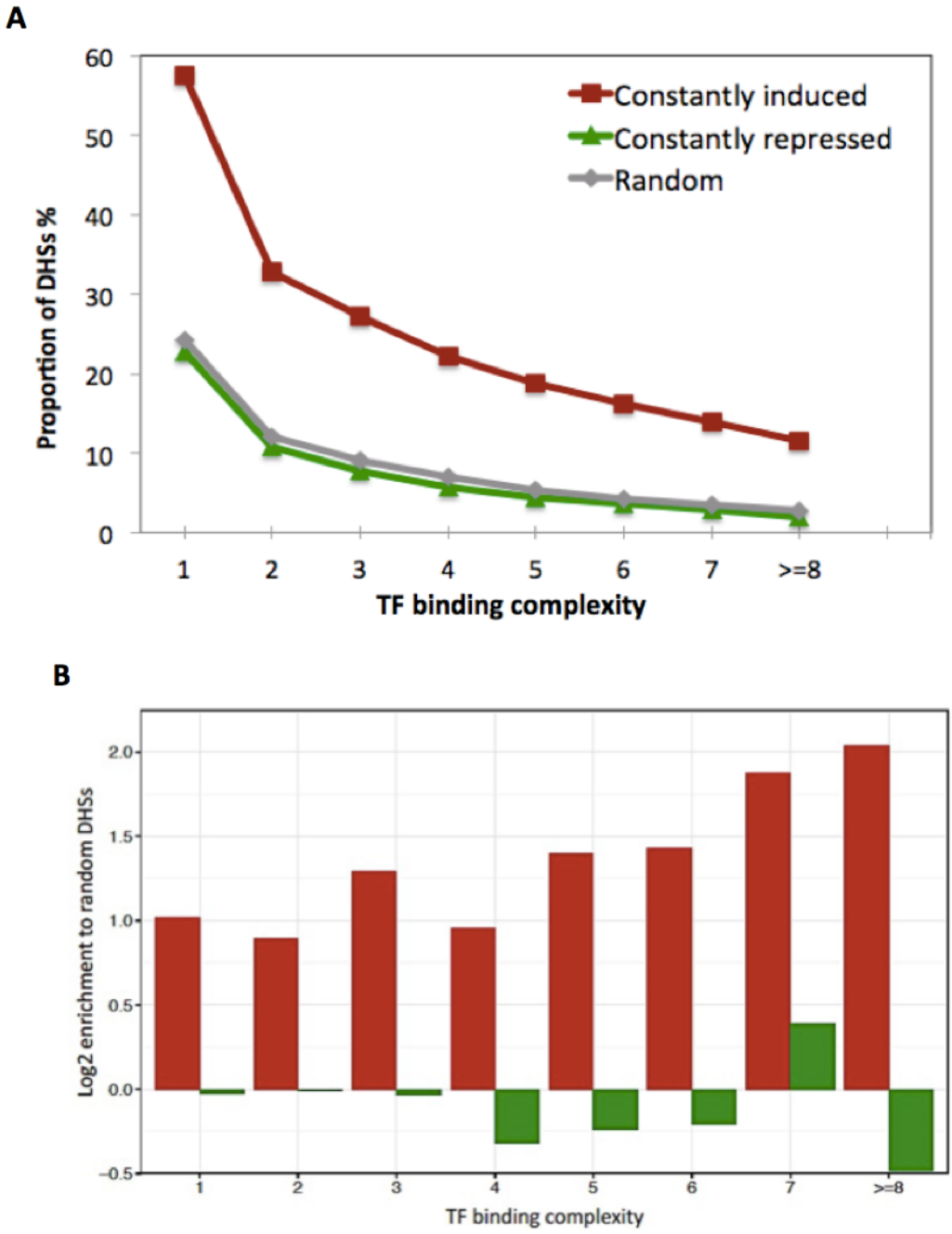

Figure 8. Enrichment of NR/TF binding sites in E2-responsive DHSs

A) Cumulative plot of percentage of DHSs with TF binding sites for different TF binding complexity. B) Enrichment of NR/TF binding sites in E2-responsive DHSs. Red: E2 constantly induced DHSs. Green: E2 constantly repressed DHSs. DHSs were grouped based on the number of harbored NR/TF binding sites ("TF binding complexity"). The fractions of DHSs with NR/TF binding sites in total number of DHSs were computed for all categories of DHSs. Log2 ratios of fractions of DHSs with different levels of TF binding complexity between E2-responsive DHSs and random DHSs were plotted on y-axis.

2.2.3.4 E2-responsive DHSs are closely associated with E2-regulated genes

From RNA-seq data, we identified 977 E2 up-regulated genes and 781 E2 down-regulated genes from all time points (exact test on negative binomial model, FDR < 0.01 and FPKM > 1). Regardless of time point, E2 up-regulated genes are strongly associated with E2 constantly induced DHSs (Fisher's exact p = 1.622e-12) and E2 down-regulated genes are negatively associated with E2 constantly induced DHSs (Fisher's exact p = 6.24e-8) compared to genes not regulated by E2 (**Figure 9A**). E2 constantly repressed DHSs tend to be associated with E2 down-regulated genes but this tendency is not statistically significant (Fisher's exact p = 0.1912, **Figure 9A**). We further classified E2-regulated genes into four clusters in order to reflect the trajectory of gene expression dynamics over time by k-means clustering (**Figure 9B**). "Cluster 1" genes are not only up-regulated by E2 at early time points (45min and 4h), but also remain up-regulated at high level over time (log2 fold change > 1.5). Genes in this category include many well-known E2-regulated genes including PGR, GREB1, XBP1, TMPRSS3 and PKIB. "Cluster 2" genes are gradually up-regulated by E2 over time and reach maximum up-regulation at 24h, such as CELSR2, DLG5, and PCNA. "Cluster 3" genes are E2 transiently up-regulated which show maximum up-regulation at 4h or 12h, including JARID2, PVT1 and NRIP1. "Cluster 4" genes represent E2 down-regulated genes such as CCNG2, IL1R1 and SALL4. Since we are interested in the association between the dynamics of DHSs and nearby gene expression, we generated an enrichment matrix representing the association between each group of E2 induced DHSs and each cluster of E2-regulated genes (**Figure 9C**). Cluster 1 genes are highly associated with all groups of dynamically induced DHSs compared to other gene clusters, especially for DHSs maximally induced at 45 min, indicating that they are E2 primary target genes and directly regulated by ER. Cluster 2 genes are more associated with late-induced DHSs (12h max and 24h max) than early ones (45min max and 4h max). In contrast, Cluster 3 genes are

more associated with early-induced DHSs (45min max and 4h max) than late ones (12h max and 24h max). Taken together, our data reveals that the dynamics of E2 response of DHSs are in phase with the dynamics of E2 response of gene transcription.

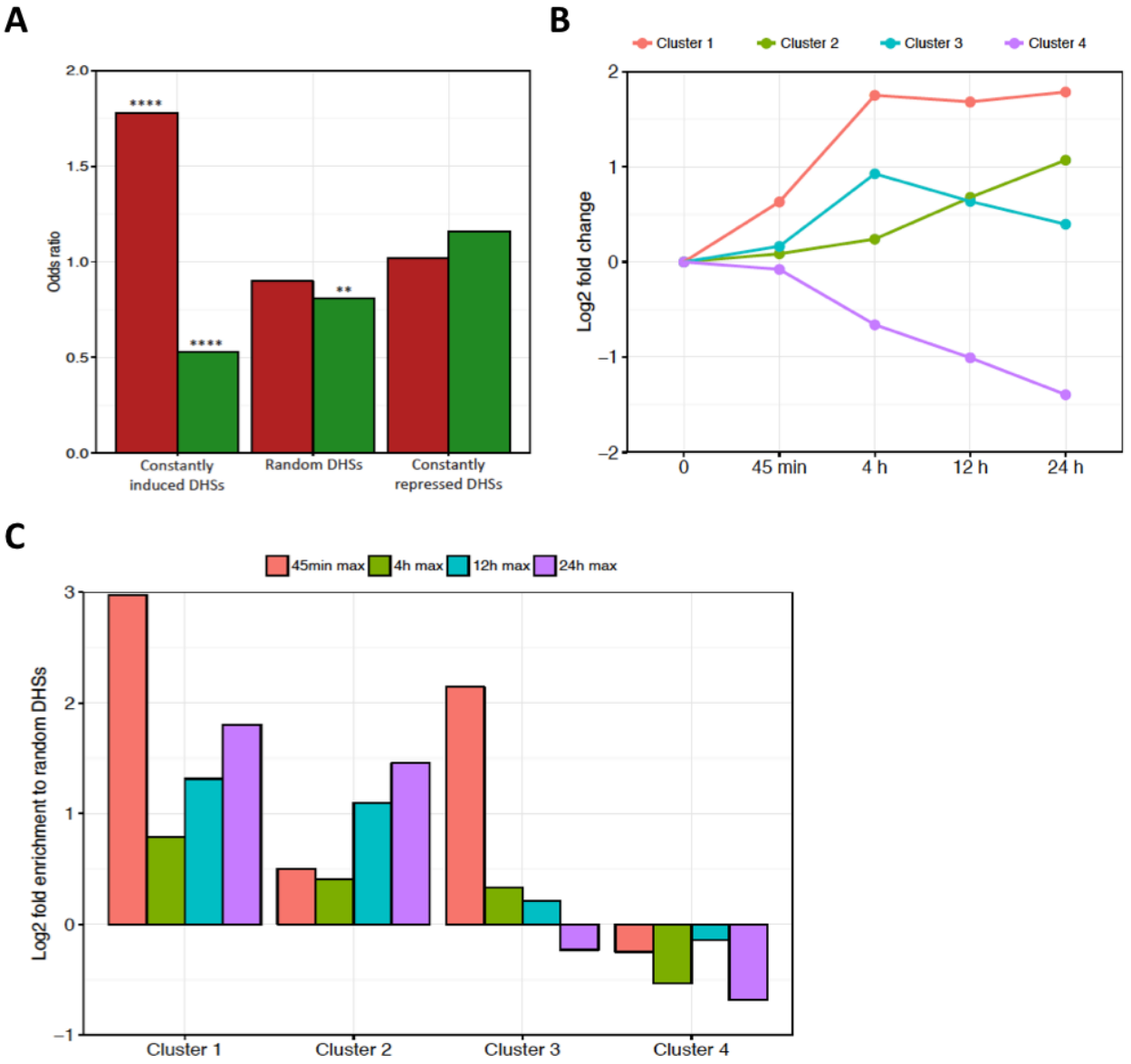

Figure 9. Association between E2-responsive DHSs and E2-regulated genes

A) Odds ratio of the association between E2-responsive DHSs and E2-regulated genes. Red: E2 up-regulated genes. Green: E2 down-regulated genes. Only genes with FPKM > 1 were included in the analysis. Odds ratio was computed as (# of regulated genes with >= 1 nearby DHSs / # of non-regulated genes with >= 1 nearby DHSs) / (# of regulated genes with 0 nearby DHS / # of non-regulated genes with 0 nearby DHS). One-sided Fisher's exact tests were performed to compare the enrichment of DHSs associated with regulated genes to DHSs associated with non-regulated genes. **: p <0.01. ****: p <0.0001. B) Dynamics of gene expression in response to E2 for different clusters of genes. Genes were grouped by k-means clustering (k = 4). The number of clusters was determined by comparing sum of squares within clusters and between clusters using different k. Data of centroid from each group was plotted on y-axis. C) Association of dynamics of E2 induced DHSs and gene expression. A DHS was assigned to a nearby gene if it is located within 50 kb of

Figure 9. (continued)
TSS of the gene. The enrichment of association between E2 dynamically induced DHSs and E2-regulated genes was calculated as (# of DHSs from given group associated with >= 1 genes from given gene cluster / total # of DHSs from given group) / (# of random DHSs associated with >= 1 genes from given gene cluster / total # of random DHSs), and plotted on y-axis in log2 scale.

2.2.4 Extensive crosstalk among various TFs on E2-responsive DHSs

2.2.4.1 TF motifs enriched in E2-responsive DHSs

In order to identify TFs that are involved in E2-modulated transcriptional regulation, we searched for known TF motifs from JASPAR database (Mathelier et al. 2016) enriched in E2-responsive DHSs compared to random DHSs. More than a hundred TF motifs are significantly enriched in E2 constantly induced DHSs, while only a handful of motifs belonging to GRHL and TEAD families are enriched in E2 constantly repressed DHSs (Fisher's exact test, FDR<0.01, **Figure 10A and Table 2**). Since analyses from earlier sections indicated that E2-induced DHSs are more closely linked to E2-triggered cistrome response and gene expression dynamics, we focused on motifs enriched in E2-induced DHSs. We noticed that key roles of some TFs from the list in E2-governed transcriptional regulation have been addressed in previous literature. To name a few, it is known that ER can be associated to enhancers/promoters through protein-protein interactions with numerous intermediary TFs including AP-1, Sp1, ATF, CREB and RUNX1 (O'Lone et al. 2004; Stender et al. 2010). Another group of TFs important for ER binding to cis-regions are pioneering factors, including FOXA1 (Hurtado et al. 2011), GATA3 (Theodorou et al. 2013) and PBX1 (Magnani et al. 2011). Besides the above-mentioned tethering factors and pioneering factors, it has been reported that ER genomic action widely interplays with genomic action of other TFs and NRs, like RARs (Hua, Kittler, and White 2009), NFκB (Frasor et al. 2009) and STAT3 (Wang et al. 2001). We compared TF binding motifs enriched in E2 constantly induced DHSs that harbor ER ChIP-seq peaks (742 DHSs) with those that do not harbor ER ChIP-seq peaks (1,516 DHSs). Irrespective of the presence of ER ChIP-seq peaks, the most significantly enriched motif families besides ERE motif are AP-1 motifs (e.g. JUN, FOS and AP-1 like motifs, Fisher's exact test, FDR = 8.77e-280),

SP/KLF motifs (e.g. KLF4, KLF5, EGR1 and SP1, Fisher's exact test, FDR = 7.01e-29) and AP-2 motifs (e.g. TFAP2C and TFAP2A, Fisher's exact test, FDR = 8.15e-28). For E2 constantly induced DHSs that harbor ER ChIP-seq peaks, in addition to the above mentioned motifs, there is also enrichment of many other NR motifs (PR, RARs, PPARG, etc.), and pioneer factor motifs such as GATA3, FOXA1 and PBX1, which agrees with previous studies. E2 constantly induced DHSs that do not harbor ER ChIP-seq peaks are particularly enriched for AP-1 and AP-1 like motifs (up to 80%), which is significantly higher than those harbor ER ChIP-seq peaks (67%, Fisher's exact test p = 7.326e-10). There is only very weak signal of ER binding in a few AP-1 motif containing DHSs of E2 constantly induced DHSs that do not harbor ER ChIP-seq peaks (**Figure 10B**), indicating that those AP-1 TFs predominantly occupied DHSs do not need to directly interact with ER.

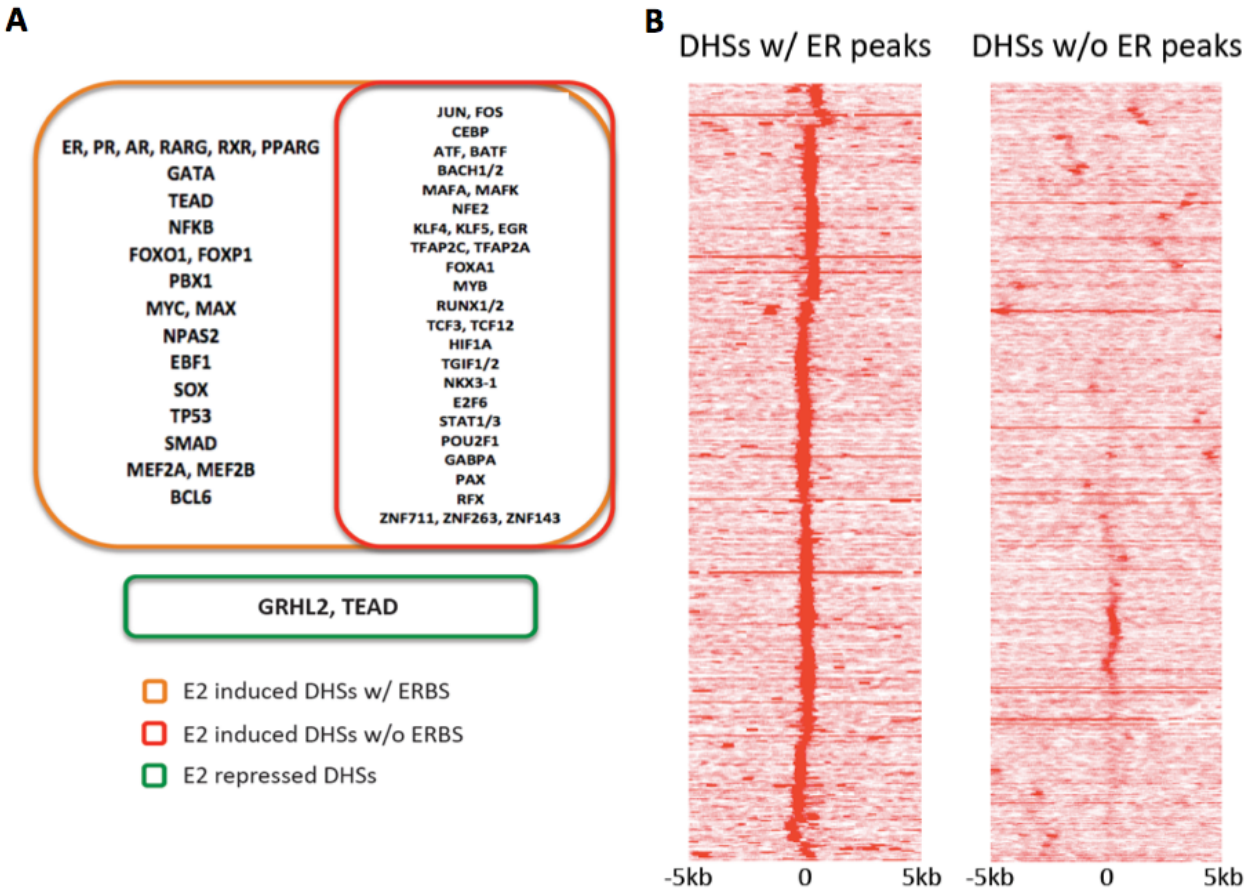

Figure 10. Known TF binding motifs enriched in E2-responsive DHSs
A) List of representative TF motifs enriched in different categories of DHSs. Motifs of TFs expressed MCF7 cells (FPKM cut-off is 1) were analyzed by FIMO using JASPAR database. Fisher's exact tests were performed using motif occurrence in random DHSs as control, and p values were adjusted for multiple testing. E2 constantly induced DHSs were divided into DHSs w/ ERBS and DHSs w/o ERBS depending on whether they contained MACS2 called ER ChIP-seq peaks. B) ER ChIP-seq signal on AP-1 motif containing E2 constantly induced DHSs. AP-1 and AP-1 like motif occurrence in E2 constantly induced DHSs was predicted by FIMO. ER ChIP-seq tag count was computed in every 50 bp window within +/- 5 kb of the center of DHSs, and normalized by library size equivalent to 10 million reads.

Motif ID	TF	Fraction %	P value	FDR	Fold enrichment
MA0490.1	JUNB	63.0	6.26E-279	1.60E-276	2.9
MA0491.1	JUND	60.9	3.44E-282	8.77E-280	2.9
MA0489.1	JUN	61.6	1.50E-249	3.83E-247	2.8
MA0476.1	FOS	61.3	2.33E-265	5.94E-263	2.8
MA0099.2	FOS::JUN	59.2	1.86E-187	4.74E-185	2.9
MA0478.1	FOSL2	64.3	5.53E-257	1.41E-254	2.3
MA0655.1	JDP2	47.8	4.45E-237	1.13E-234	2.8
MA0462.1	BATF::JUN	48.7	9.73E-116	2.48E-113	2.4
MA0258.2	ESR2	56.3	8.24E-138	2.10E-135	1.9
MA0150.2	Nfe2l2	45.3	1.30E-115	3.32E-113	1.9
MA0591.1	Bach1::Mafk	46.7	1.64E-116	4.18E-114	1.9
MA0112.3	ESR1	36.4	2.28E-110	5.81E-108	2.1
MA0599.1	KLF5	69.3	2.75E-31	7.01E-29	1.3
MA0746.1	SP3	59.4	5.22E-34	1.33E-31	1.3
MA0039.2	Klf4	59.2	6.96E-49	1.77E-46	1.3
MA0066.1	PPARG	32.2	1.37E-48	3.49E-46	1.6
MA0079.3	SP1	68.1	2.56E-23	6.53E-21	1.2
MA0516.1	SP2	67.8	9.96E-24	2.54E-21	1.2
MA0162.2	EGR1	61.3	2.08E-26	5.30E-24	1.3
MA0810.1	TFAP2A	47.7	1.12E-44	2.86E-42	1.4
MA0524.2	TFAP2C	43.6	4.68E-52	1.19E-49	1.4
MA0521.1	Tcf12	51.6	1.91E-27	4.87E-25	1.3
MA0741.1	KLF16	55.8	2.91E-29	7.42E-27	1.3
MA0824.1	ID4	48.7	2.66E-28	6.78E-26	1.3
MA0747.1	SP8	53.0	4.08E-26	1.04E-23	1.3
MA0522.2	TCF3	49.0	3.87E-26	9.87E-24	1.3
MA0065.2	Pparg::Rxra	53.2	2.82E-25	7.19E-23	1.2
MA0513.1	SMAD2::SMAD3::SMAD4	46.1	3.19E-32	8.13E-30	1.3
MA0816.1	Ascl2	43.0	5.29E-26	1.35E-23	1.3
MA0691.1	TFAP4	28.4	8.82E-31	2.25E-28	1.4
MA0774.1	MEIS2	30.8	1.07E-28	2.73E-26	1.4
MA0146.2	Zfx	51.1	3.03E-28	7.73E-26	1.2
MA0528.1	ZNF263	77.6	3.57E-13	9.10E-11	1.1
MA0002.2	RUNX1	41.6	7.30E-21	1.86E-18	1.3
MA0596.1	SREBF2	27.6	3.03E-31	7.73E-29	1.4
MA0775.1	MEIS3	26.0	3.72E-24	9.49E-22	1.4
MA0595.1	SREBF1	28.7	5.25E-30	1.34E-27	1.4
MA0504.1	NR2C2	47.3	1.32E-20	3.37E-18	1.2
MA0671.1	NFIX	30.2	2.32E-37	5.92E-35	1.3

Table 2. Complete list of known TF motifs enriched in E2 constantly induced DHSs

Table 2. Complete list of known TF motifs enriched in E2 constantly induced DHSs (continued)

MA0073.1	RREB1	54.1	1.32E-17	3.37E-15	1.2
MA0685.1	SP4	43.9	3.71E-23	9.46E-21	1.2
MA0732.1	EGR3	35.5	2.05E-23	5.23E-21	1.3
MA0597.1	THAP1	40.8	9.47E-35	2.41E-32	1.2
MA0159.1	RARA::RXRA	32.7	7.30E-27	1.86E-24	1.3
MA0100.2	Myb	29.7	2.01E-15	5.13E-13	1.3
MA0471.1	E2F6	49.2	1.49E-14	3.80E-12	1.2
MA0753.1	ZNF740	40.0	6.60E-26	1.68E-23	1.2
MA0113.3	NR3C1	14.3	1.38E-16	3.52E-14	1.5
MA0088.2	ZNF143	28.0	2.06E-19	5.25E-17	1.3
MA0007.3	Ar	16.3	2.27E-17	5.79E-15	1.4
MA0730.1	RARA	29.8	2.63E-15	6.71E-13	1.2
MA0526.1	USF2	31.8	2.66E-25	6.78E-23	1.2
MA0797.1	TGIF2	20.9	5.23E-16	1.33E-13	1.3
MA0782.1	PKNOX1	23.9	2.08E-16	5.30E-14	1.3
MA0147.2	Myc	25.3	7.57E-14	1.93E-11	1.3
MA0801.1	MGA	23.8	1.40E-17	3.57E-15	1.3
MA0057.1	MZF1	39.1	1.97E-14	5.02E-12	1.2
MA0495.1	MAFF	26.9	1.53E-11	3.90E-09	1.2
MA0138.2	REST	40.1	2.92E-15	7.45E-13	1.1
MA0160.1	NR4A2	26.4	1.56E-36	3.98E-34	1.2
MA0494.1	Nr1h3::Rxra	27.8	1.29E-07	3.29E-05	1.2
MA0861.1	TP73	14.3	8.98E-15	2.29E-12	1.3
MA0496.1	MAFK	26.7	1.88E-13	4.79E-11	1.2
MA0105.4	NFKB1	22.5	4.20E-21	1.07E-18	1.2
MA0074.1	RXRA::VDR	16.5	5.14E-10	1.31E-07	1.3
MA0018.2	CREB1	13.2	1.96E-26	5.00E-24	1.3
MA0659.1	MAFG	14.7	1.08E-14	2.75E-12	1.3
MA0101.1	REL	25.8	3.58E-06	9.13E-04	1.2
MA0509.1	Rfx1	26.7	6.13E-15	1.56E-12	1.2
MA0796.1	TGIF1	20.7	2.15E-12	5.48E-10	1.2
MA0778.1	NFKB2	21.8	1.75E-18	4.46E-16	1.2
MA0677.1	Nr2f6	27.5	1.61E-19	4.11E-17	1.2
MA0819.1	CLOCK	10.8	9.30E-08	2.37E-05	1.3
MA0860.1	Rarg	25.3	4.26E-09	1.09E-06	1.2
MA0093.2	USF1	25.5	5.04E-17	1.29E-14	1.2
MA0106.3	TP53	13.6	5.76E-16	1.47E-13	1.3
MA0107.1	RELA	27.8	3.85E-06	9.82E-04	1.2
MA0592.2	Esrra	22.8	8.39E-11	2.14E-08	1.2
MA0498.2	MEIS1	13.0	3.34E-34	8.52E-32	1.3

Table 2. Complete list of known TF motifs enriched in E2 constantly induced DHSs (continued)

MA0736.1	GLIS2	29.2	4.65E-22	1.19E-19	1.1
MA0657.1	KLF13	23.6	3.44E-12	8.77E-10	1.2
MA0058.3	MAX	13.0	1.27E-11	3.24E-09	1.2
MA0798.1	RFX3	11.6	1.12E-05	2.86E-03	1.3
MA0117.2	Mafb	21.0	1.24E-10	3.16E-08	1.2
MA0831.1	TFE3	13.3	3.76E-19	9.59E-17	1.2
MA0833.1	ATF4	14.7	7.81E-07	1.99E-04	1.2
MA0839.1	CREB3L1	14.4	2.51E-14	6.40E-12	1.2
MA0059.1	MAX::MYC	16.7	2.59E-08	6.60E-06	1.2
MA0605.1	Atf3	8.6	6.94E-39	1.77E-36	1.3
MA0076.2	ELK4	26.5	2.28E-11	5.81E-09	1.1
MA0663.1	MLX	8.9	1.45E-05	3.70E-03	1.3
MA0626.1	Npas2	12.6	4.48E-10	1.14E-07	1.2
MA0115.1	NR1H2::RXRA	21.4	2.01E-11	5.13E-09	1.2
MA0808.1	TEAD3	21.6	1.05E-06	2.68E-04	1.2
MA0664.1	MLXIPL	9.1	1.59E-13	4.05E-11	1.3
MA0825.1	MNT	12.6	3.24E-16	8.26E-14	1.2
MA0060.2	NFYA	17.8	6.21E-10	1.58E-07	1.2
MA0512.2	Rxra	24.2	2.80E-14	7.14E-12	1.1
MA0511.2	RUNX2	12.8	1.19E-09	3.03E-07	1.2
MA0617.1	Id2	13.9	5.87E-18	1.50E-15	1.2
MA0844.1	XBP1	8.3	3.34E-10	8.52E-08	1.2
MA0855.1	RXRB	24.0	1.43E-11	3.65E-09	1.1
MA0608.1	Creb3l2	10.0	8.35E-11	2.13E-08	1.2
MA0750.1	ZBTB7A	21.3	1.29E-16	3.29E-14	1.1
MA0603.1	Arntl	10.7	1.13E-13	2.88E-11	1.2
MA0616.1	Hes2	13.4	8.14E-09	2.08E-06	1.1
MA0781.1	PAX9	11.3	2.61E-14	6.66E-12	1.2
MA0834.1	ATF7	7.5	6.46E-10	1.65E-07	1.2
MA0622.1	Mlxip	12.8	3.94E-18	1.00E-15	1.1
MA0464.2	BHLHE40	9.6	3.14E-12	8.01E-10	1.2
MA0259.1	ARNT::HIF1A	11.8	3.56E-15	9.08E-13	1.1
MA0062.2	Gabpa	21.2	2.72E-08	6.94E-06	1.1
MA0645.1	ETV6	20.6	4.19E-08	1.07E-05	1.1

2.2.4.2 DNase I footprint analysis of E2-induced DHSs

As just mentioned, motif analysis based on DNA sequence composition has revealed various TF motifs especially ERE, AP-1, SP/KLF and AP-2 related motifs enriched in E2 constantly induced DHSs. Considering that the presence of motif is only indication of potential TF recruitment, we further predicted DNase I footprints on E2-induced DHSs from DNase-seq data to see if we could detect protection from DNase I cleavage on predicted TF motifs. The underlying rationale for DNase I footprint analysis is that when a TF protein physically interacts with cis-chromatin, it would protect the occupied DNA motif from DNase I cleavage and therefore leave valley-like footprints within DNase I hypersensitivity sites. We were able to detect DNase I footprints on JUN, KLF4 and TFAP2C motifs in E2 constantly induced DHSs, and footprints on all three TF motifs were deeper upon E2 treatment than vehicle control (Wilcox $p < 2.23\text{-}16$), indicating increasing protection on these sites due to stronger protein-DNA interaction triggered by E2 (**Figure 11**). We detected very poor and shallow footprints on ERE sites (data not shown), which is consistent with previous studies that showed footprints of NRs such as ER and GR are barely detectable in DNase-seq data (Sung et al. 2014). This latter result may arise because their residence time on DNA is usually very short and transient (McNally et al. 2000), giving insufficient protection of DNA from DNase I cleavage. Nonetheless, our DNase I footprint analysis has further suggested AP-1, SP/KLF and AP-2 family TFs frequently interact with E2 induced DHSs and such interaction is triggered by E2 stimulation.

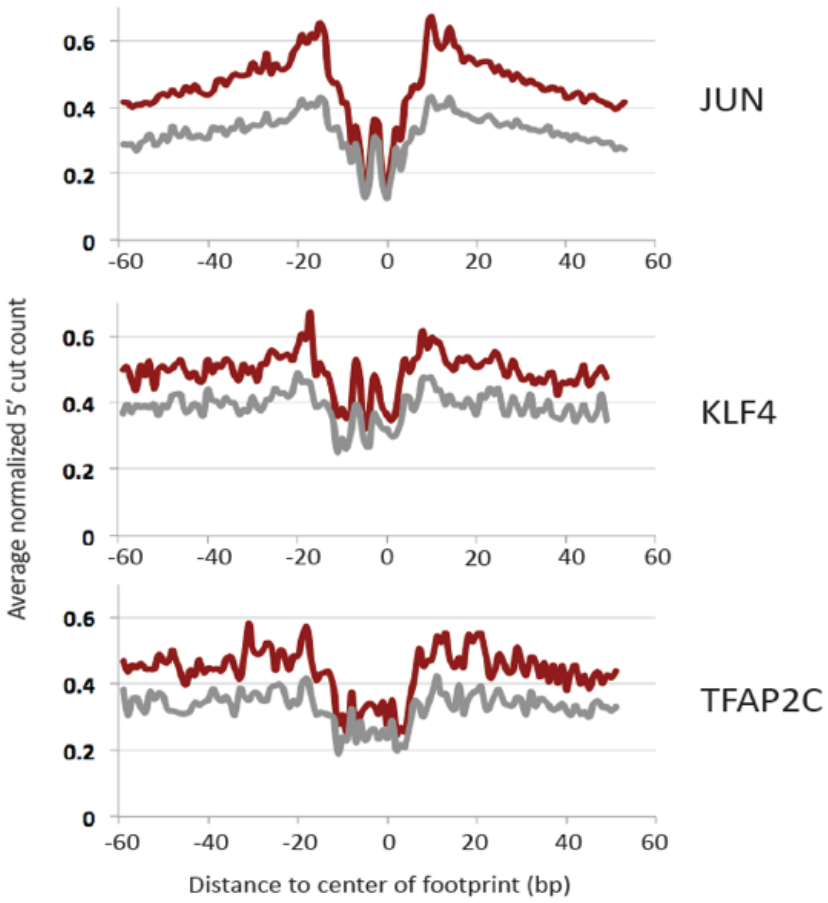

Figure 11. Signal of DNase I cleavage on JUN, KLF4 and TFAP2C motifs in E2 constantly induced DHSs

Red: E2 treatment for 12 hours. Grey: vehicle treatment for 12 hours. Average DNase-seq 5' end cut count was calculated on every base pair within +/- 50 bp flanking regions centered by predicted TF motifs, and normalized by effective library size equivalent to 10 million reads.

2.2.4.3 TF motifs enriched in E2 dynamically induced DHSs

We next estimated the enrichment of ERE, AP-1, SP/KLF and AP-2 motifs in E2 dynamically induced DHSs, and we found the four motif families are enriched sequentially after E2 treatment (**Figure 12**). As expected, ERE motifs are highly enriched in 45min max DHSs but their enrichment is no longer maintained in DHSs maximally induced at 4h, 12h and 24h. AP-1 motifs are not strongly enriched in 45min max DHSs, but are strongly enriched in 4h max DHSs and remain highly enriched in DHSs maximally induced at later time points, which suggests that AP-1 TFs respond to E2 stimulation quickly after ER. SP/KLF and AP-2 motifs are not enriched in DHSs maximally induced at early time points (45min and 4h) but significantly enriched in DHSs maximum induced at late time points (12h and 24h), indicating that those TFs respond to E2 stimulation more slowly than ER and AP-1 TFs but may exert their gene regulatory function in secondary transcriptional regulation. From RNA-seq data, we have found many AP-1, SP/KLF and AP-2 genes, such as FOSL2, KLF4 and TFAP2C, are rapidly transcriptionally up-regulated by E2 (differentially expressed genes with E2 treatment for 45 min, exact test on negative binomial model, FDR < 0.001), indicating that these TFs are E2 primary target genes that act as important regulators in secondary transcriptional regulation.

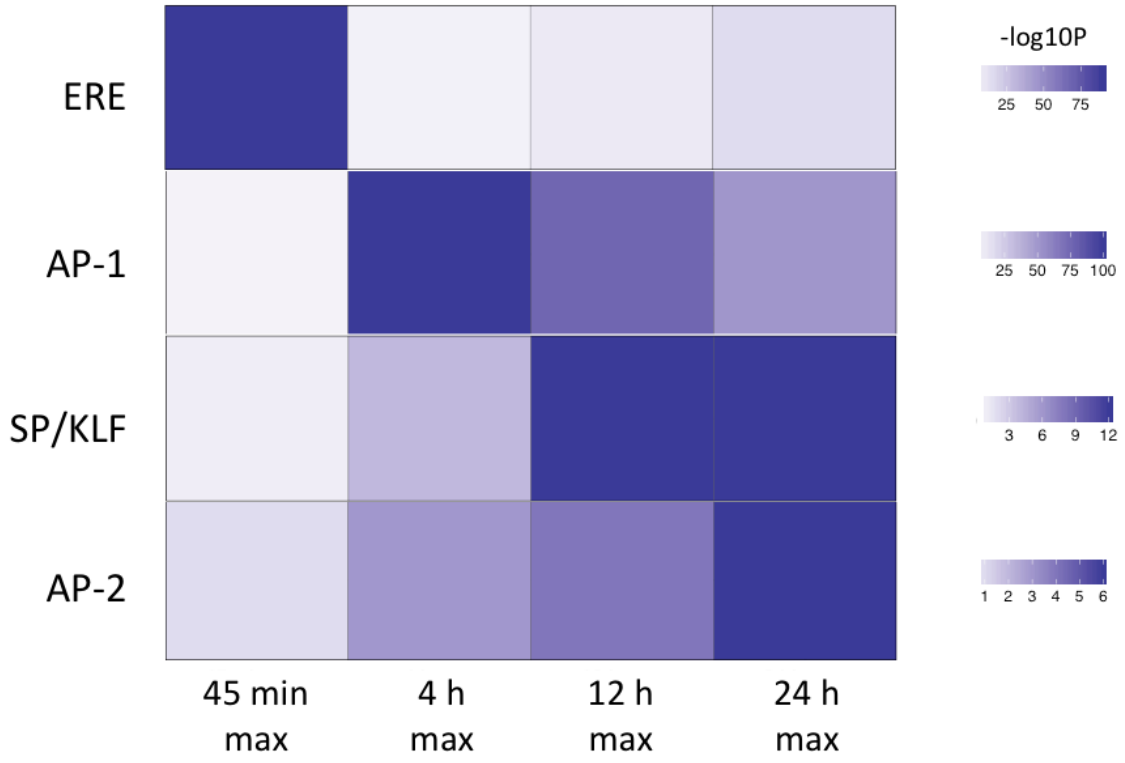

Figure 12. Enrichment of ERE, AP-1, SP/KLF and AP-2 motifs in E2 dynamically induced DHSs

Motif appearance of these TFs was detected by FIMO using JASPAR database. One-tailed Fisher's exact tests were performed using occurrence of TF motifs in random DHSs as background control. P values were corrected for multiple testing and plotted in heat map after –log10 transformed. The darker the square is, the more significant p value it represents.

2.2.5 Assessment of regulatory activity of E2 signaling related DHSs by CapSTARR-seq

2.2.5.1 Overview of CapSTARR-seq experiment

To directly interrogate regulatory activity of E2 related DHSs in a high throughput manner, we employed CapSTARR-seq assays in MCF7 cells with E2 treatment over time. We captured 10,825 regions spanning a total of 9.74 Mb of human genome to enrich for E2-responsive DHSs (E2 constantly induced DHSs, E2 constantly repressed DHSs and E2 dynamically induced DHSs) and DHSs associated with E2-regulated genes. We included 500 "control DHSs" that are randomly selected DHSs and are neither E2 responsive nor near E2-regulated genes, and 500 "complete negative controls" that are randomly selected closed chromatin regions lacking active H3K27ac signal and breast cancer associated TF/NR binding sites in MCF7 mapped previously (Kittler et al. 2013) (**Figure 13A**). We generated input screening libraries by cloning these candidate regions into reporter constructs driven by two different minimal promoters, SCP1 and miniCMV, respectively, and placing candidate regions in the 3'UTR of reporter gene eGFP. SCP1 is a synthetic core promoter used in original STARR-seq and CapSTARR-seq papers, which is expected to be a stronger core promoter than miniCMV. The reasons to use two different minimal promoters are to better estimate data reproducibility and to robustly assess regulatory activity. After transiently electroporated screening libraries into MCF7 cells and at least 48 hours of hormone starvation, we applied E2 or ethanol in the medium and collected cells in time-course manner. Target-specific mRNA-seq libraries were prepared, in which only transcripts produced from screening library plasmids were enriched and amplified for Illumina sequencing (**Figure 13B**).

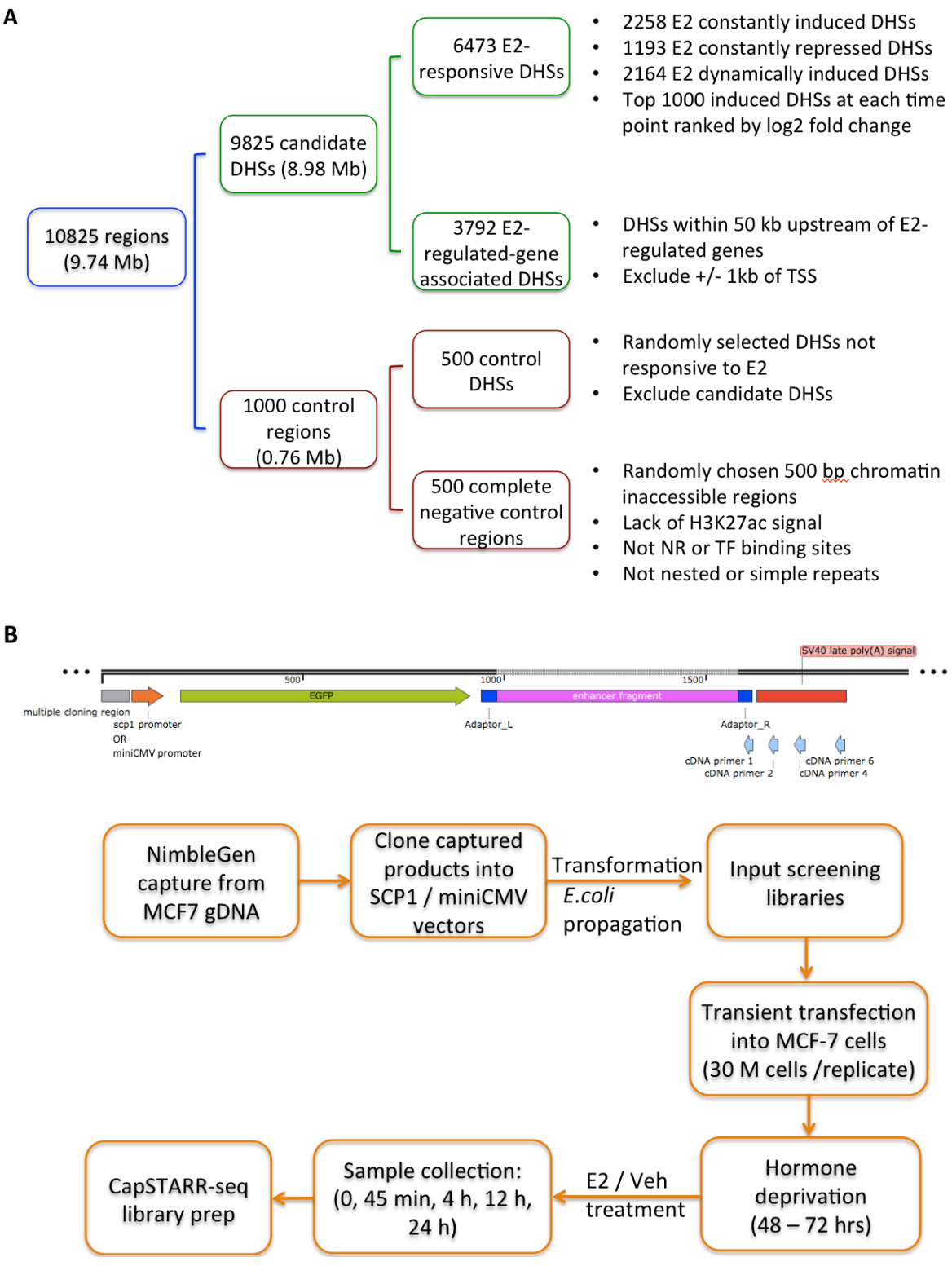

Figure 13. Overview of CapSTARR-seq experiment

A) Composition of captured regions in CapSTARR-seq. B) Design of CapSTARR-seq input screening vectors and experimental workflow.

2.2.5.2 Quality metrics of CapSTARR-seq libraries

We obtained 100 bp paired-end reads and recovered genomic coordinates of sequenced fragments from uniquely mapped reads. For input screening libraries, we successfully captured nearly all designed regions with close to 90% on-target rate and super deep per bp coverage (**Table 3, Figure 14A and 14B**). Since we have not detected signal indicating saturation of sequencing and we estimated total independent clones by counting single colonies on LB agar plates, the complexity of input libraries calculated from sequencing data is very likely to under-represent the actual complexity of screening libraries. Compared to input libraries, the output libraries have much lower but still satisfying complexity on captured regions (**Table 4**). While 88% of regions in SCP1 libraries were covered at least 10 times, 48% of regions in miniCMV libraries were covered at least 10 times. The difference in complexity between libraries driven by the two promoters may be due to basal promoter strength of SCP1 and miniCMV promoters. After normalizing output library signal by input library representation, the data between biological replicates is very reproducible for both SCP1 and miniCMV libraries (**Figure 15A and 15B**). Therefore, we combined data from biological replicates for later analyses in order to increase data complexity and statistical power to detect enhancers.

Sample	Total frag #	Mapped frag #	Mapping rate %	Distinct frag #	On-target rate %	Capture region covered %	Mean per bp coverage	Median per bp coverage
SCP1 input	53,908,601	53,165,477	98.62	12,076,986	89.67	99.83	484	394
miniCMV input	12,324,101	12,020,367	97.53	6,997,737	88.83	99.79	279	222

Table 3. Quality metrics of CapSTARR-seq input screening libraries

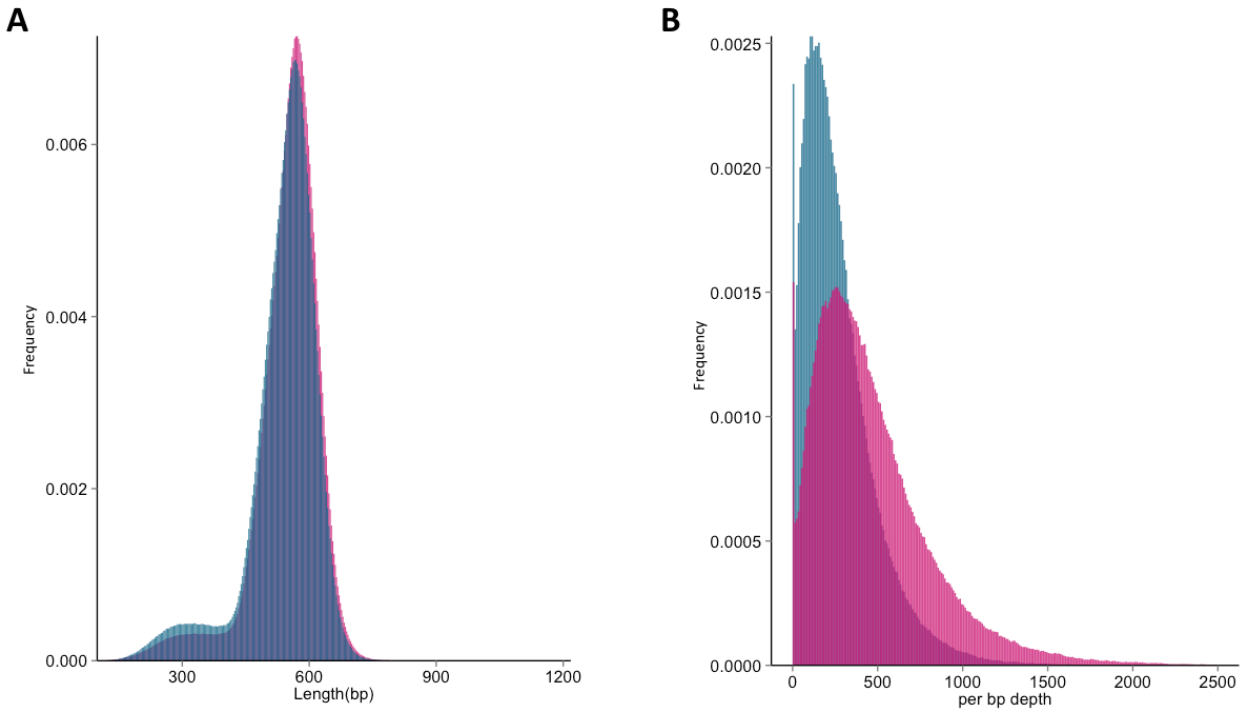

Figure 14. Quality metrics plots of CapSTARR-seq input screening libraries
A) Distribution of insert fragment lengths of input libraries. Pink: SCP1 input library. Blue: miniCMV input library. B) Per base pair coverage of input libraries. Pink: SCP1 input library. Blue: miniCMV input library. Only distinct fragments overlapping with captures regions were counted for estimating per base pair coverage.

Sample	Mean distinct frag #	Median distinct frag #	Mean per bp coverage	Median per bp coverage	Max per bp coverage	% of regions with >= 10× coverage
SCP1	1.24 M	1.3 M	54	39	1931	88.4 %
miniCMV	0.33 M	0.32 M	13	9	524	48.3 %

Table 4. Quality metrics of CapSTARR-seq output libraries

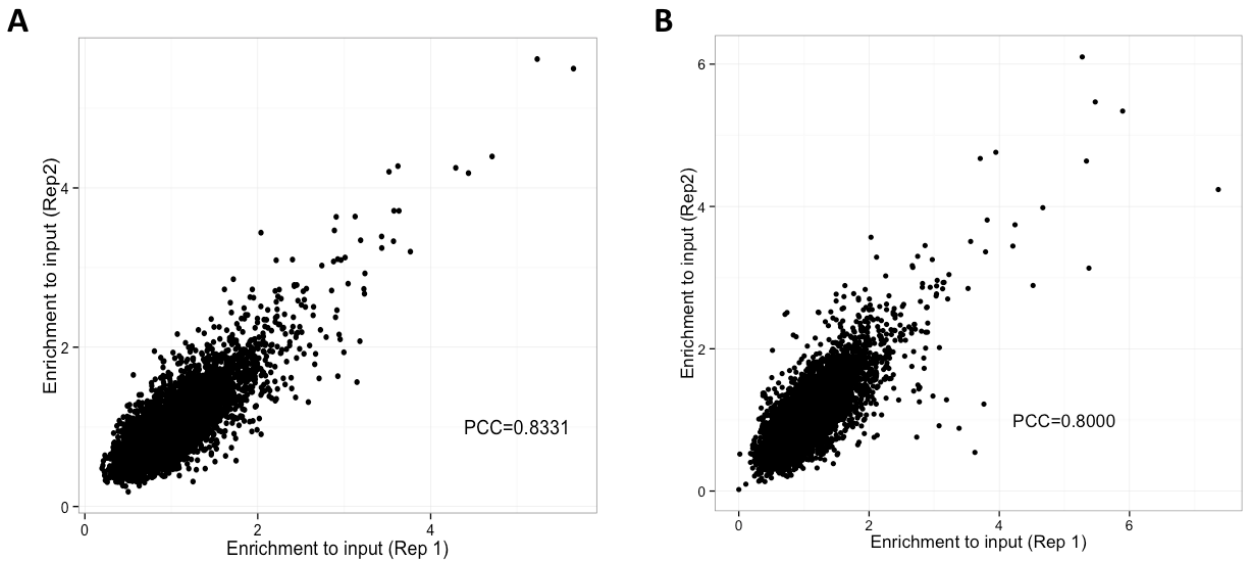

Figure 15. Data reproducibility of CapSTARR-seq output libraries
A) Dot plot of correlation between biological replicates of SCP1 libraries. B) Dot plot of correlation between biological replicates of miniCMV libraries. Note: Data of cells prior to ligand treatment was used. Number of fragments on each captured region was counted and normalized by on-target library size. Enrichment to input value for each region was calculated as dividing normalized count of output library by normalized count of input library.

2.2.5.3 Identification of active enhancers

After normalizing output library signal by input library signal, complete negative control regions on average showed the lowest fold of enrichment. Therefore, we applied the following formula to assign enhancer activity E to each captured region i:

$$E_i = \frac{O_i/I_i}{\overline{O^{neg}/I^{neg}}}$$

In this formula, O_i and I_i are the numbers of uniquely mapped fragments in output and input library on the i-th captured region after normalized by library sizes, respectively. The numerator denotes the mean O/I for all complete negative control regions, which is 0.65 for both SCP1 and miniCMV data. Candidate enhancers were preliminarily called by MACS2 (Zhang et al. 2008) using input library as control (FDR < 0.05). An active enhancer is called if E_i is greater than 1.5 in either all E2 or all vehicle treatment conditions for a candidate enhancer. We reproducibly detected 3,039 active enhancer regions from SCP1 data and 3,255 active enhancer regions from miniCMV data (**Figure 16B**). About 81% of the called enhancers are shared by the two datasets, and quantitatively, the enhancer activity estimated from SCP1 data and miniCMV data correlates well with each other (Pearson correlation coefficient = 0.7845, **Figure 15A**). We measured enhancer activity of randomly chosen 23 called enhancers by conventional dual luciferase reporter assays. The enhancer activity measured by CapSTARR-seq is in good correlation with that measured by luciferase reporter assay (Pearson correlation coefficient = 0.6518, **Figure 15C**). Among all categories of captured regions, E2-regulated gene associated DHSs have the highest enhancer calling rate, while E2 constantly repressed DHSs have the lowest enhancer calling rate which is even lower than control DHSs (**Figure 15D**).

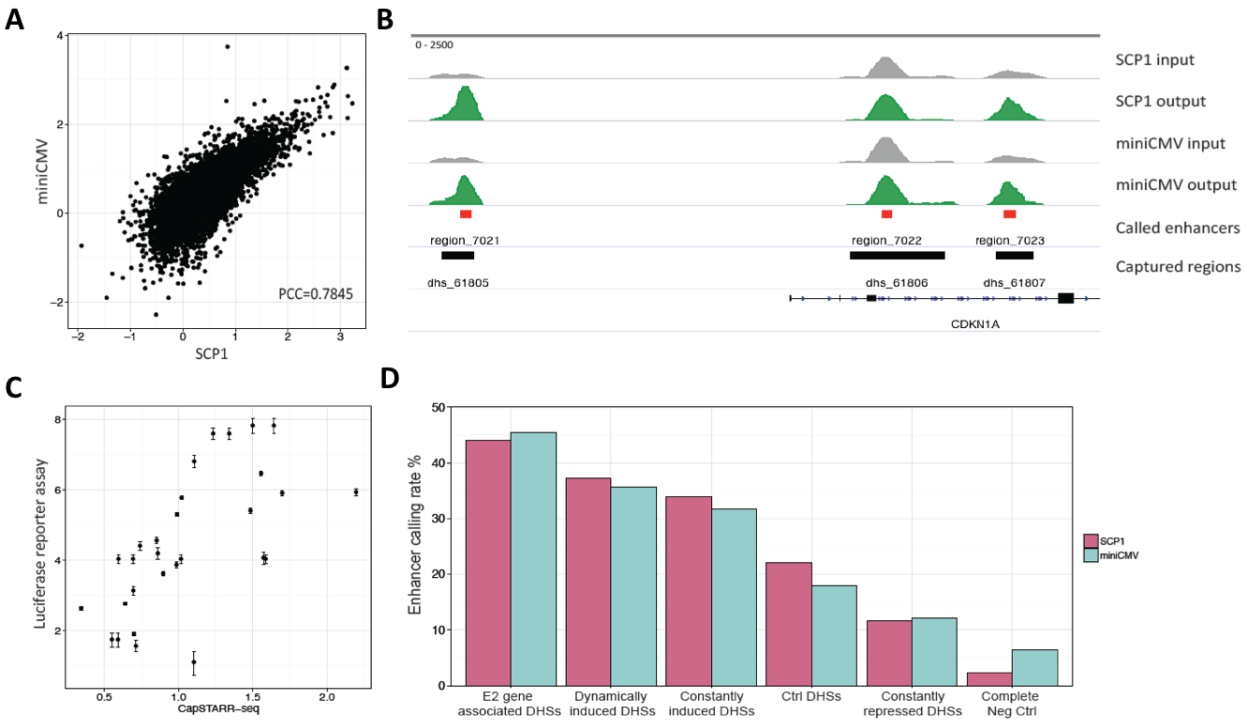

Figure 16. Active enhancers identified from CapSTARR-seq

A) Correlation of enhancer activity between SCP1 data and miniCMV data. Enhancers called from both SCP1 and miniCMV data were included. Enhancer activity under untreated condition (prior to ligand treatment but hormone deprived for 48 hours) was plotted in log2 scale. B) Snapshot of raw CapSTARR-seq signal on called active enhancers. Output data from untreated condition was graphed after being normalized by on-target library size equivalent to 15 million reads. C) Correlation between CapSTARR-seq and luciferase reporter assay. Enhancer activity under untreated condition was plotted on x-axis in log2 scale. Relative luciferase activity was estimated by comparing internal control normalized luciferase activity of candidate regions to that of pGL4.23[minP/luc2] negative control plasmid, and plotted in log2 scale. Error bars represent standard errors from three biological replicates. D) Bar graph of enhancer calling rate for different classes of captured regions. Enhancer calling rate was calculated by dividing the number of called enhancers by the number of regions that passed filters.

2.2.5.4 Identification of E2-responsive enhancers

In addition to active enhancers without hormone treatment context, we further identified E2-responsive enhancers where differential enhancer activity was observed in response to E2 treatment. By comparing enhancer activity between E2 and vehicle treatment conditions, we identified 440 regions with E2-induced enhancer activity for at least one time point (fold change of enhancer activity in E2 compared to vehicle > 1.2, **Figure 17A**), and 311 regions with E2-repressed enhancer activity for at least one time point (fold change of enhancer activity in E2 compared to Veh < 0.8, **Figure 17B**). E2-induced DHSs are significantly enriched for E2-induced enhancers and depleted for E2-repressed enhancers compared to control DHSs (**Figure 17C**), indicating most E2-induced DHSs have E2-induced transcriptional activation function. We randomly chose 17 E2-induced enhancers and performed dual luciferase reporter assay with E2 and vehicle treatment for 12 hours. Among them, 15 regions showed more than 0.5 log2 fold induction of enhancer activity with E2 treatment (**Figure 17D**), which successfully recapitulated CapSTARR-seq results.

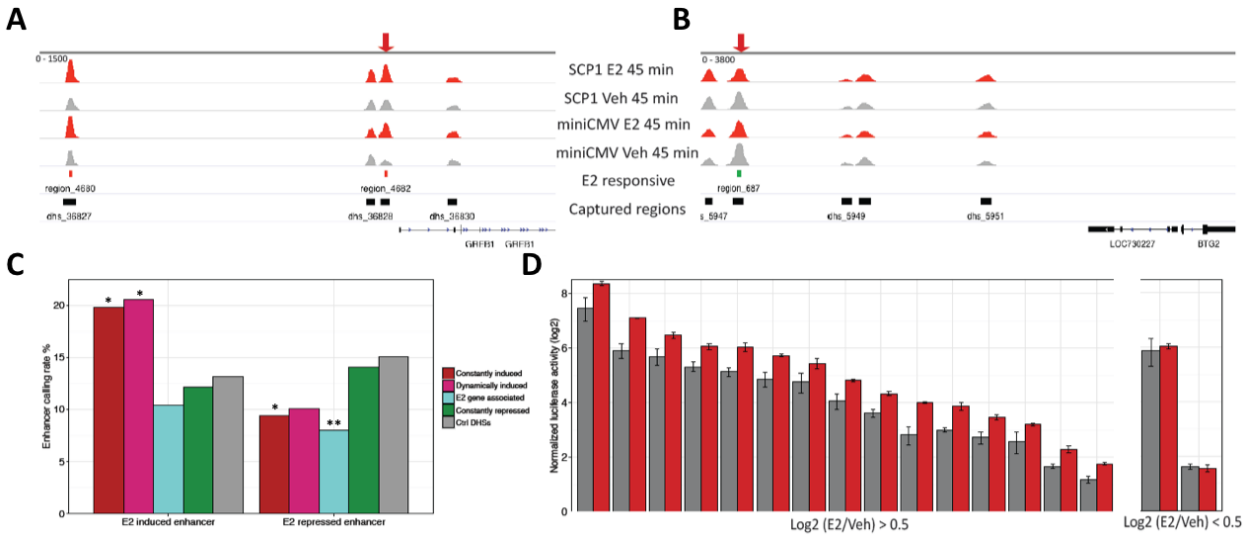

Figure 17. Identification of E2-responsive enhancers

A) Snapshot of raw CapSTARR-seq signal on E2-induced enhancers and B) Snapshot of raw CapSTARR-seq signal on E2-repressed enhancers. CapSTARR-seq data was normalized by on-target library size equivalent to 15 million reads. C) E2-responsive enhancer calling rate for different classes of captured regions. Calling rate was calculated by dividing the number of called E2-responsive enhancers by the number of regions that passed filters. D) E2-responsive enhancers investigated by luciferase reporter assay. Red: E2 treatment for 12 hours. Grey: vehicle treatment for 12 hours. Relative luciferase activity was estimated by comparing internal control normalized luciferase activity of candidate regions to that of pGL4.23[minP/luc2] negative control plasmid, and plotted in log2 scale. Error bars represent standard errors from three biological replicates.

2.2.5.5 Correlation of enhancer activity and gene expression

Although we limited our assay to a subset of chromatin accessible regions, and CapSTARR-seq measures regulatory activity regardless of chromatin context, we were able to observe correlation between endogenous gene expression level and the number of nearby enhancers (**Figure 18**). We found that genes associated with at least one active enhancer are expressed significantly higher than genes associated with captured regions but no called active enhancer (one-tailed Wilcox p = 1.793e-6). In addition, we discovered a positive correlation between the E2 response of gene expression and E2 response of nearby enhancer. We found that 111/440 (25.2%) of E2-induced enhancers are associated with E2 up-regulated genes, which is significantly higher than E2-repressed enhancers (50/311 = 16%, Fisher's exact p = 3.422e-4). Quantitative analysis of gene expression in response to E2 associated with nearby E2-responsive enhancers revealed that genes associated with E2-induced enhancers are induced by E2 significantly higher than those associated with E2-repressed enhancers (one-tailed Wilcox p = 5.524e-3, **Figure 19A**). We also found the number of E2-induced enhancers to be positively correlated with the induction level of gene expression by E2. Genes associated with at least two E2-induced enhancers are up-regulated significantly higher than those associated with only one E2-induced enhancer (Wilcox p = 8.715e-4, **Figure 19B**). One of the E2 most strongly up-regulated genes, GREB1, has nine E2-induced enhancers located within 50 kb of its TSS, and its expression is induced by E2 for 4.82 fold. Our results indicate that strong E2 induction of gene transcription is mediated through collaboration of multiple enhancers.

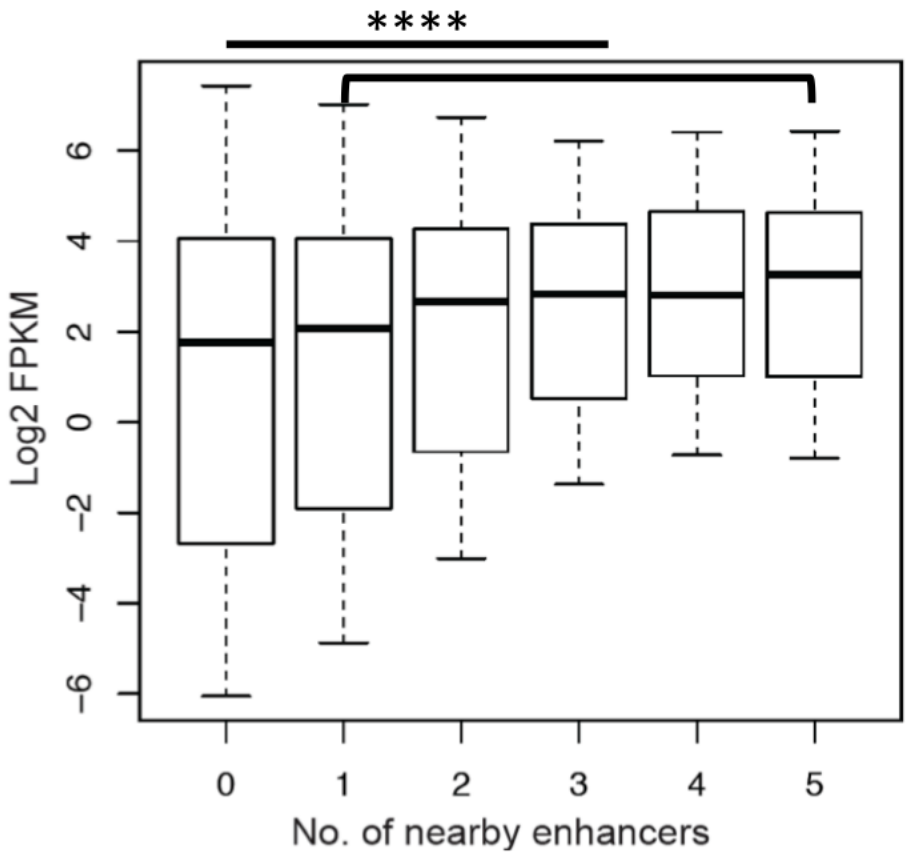

Figure 18. Correlation between active enhancers and gene expression
Boxplot of gene expression level binned by number of nearby enhancers. Enhancers within 50 kb of TSS were counted.

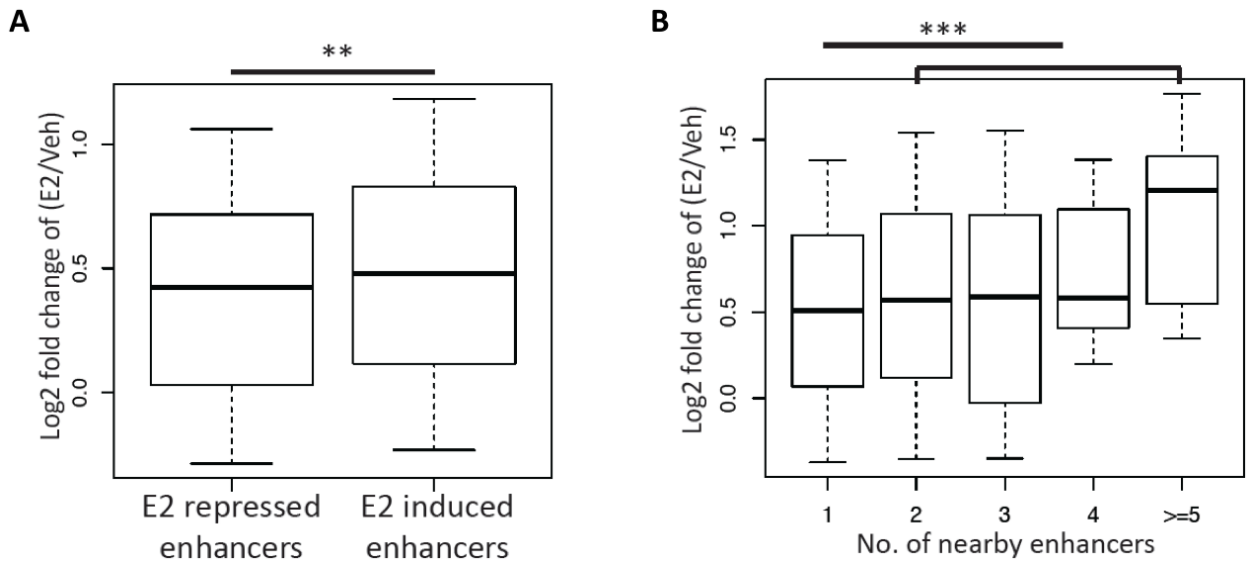

Figure 19. Correlation between E2-regulated gene expression and enhancer activity in response to E2

A) Correlation between E2 response of gene expression and nearby E2-responsive enhancers. Gene expression data with E2 treatment for 24 hours was plotted. **: $p<0.01$. B) Correlation between E2 response of gene expression and number of nearby E2-induced enhancers. Gene expression data with E2 treatment for 24 hours was plotted. ***: $p<0.0001$.

2.2.5.6 E2-induced enhancers reveal the dynamics of TF interaction in E2 signaling

One advantage of CapSTARR-seq compared to conventional luciferase reporter assays is that we are able to detect E2 induced enhancer activity of many enhancers at as early as 45min by CapSTARR-seq but not by luciferase reporter assay, because the readout of CapSTARR-seq is mRNA molecules that are accumulated much faster than the readout of luciferase assays (e.g. mature protein). Therefore, CapSTARR-seq data reflects enhancer activity dynamics in a more real-time manner. While E2-induced enhancers are more enriched for ER binding sites than E2-repressed enhancers (Fisher's exact p=8.953e-14), E2-induced enhancers with ER binding sites are more induced than those without ER binding sites specifically at 45 min (Wilcox p<2.2e-16) but not later time points (**Figure 20**), which is consistent with DNase-seq data, reinforcing that ER is the master regulator in primary transcriptional regulation of E2. We were not able to stratify E2-responsive enhancer activity by using any other TF motif alone besides ER, indicating that multiple TFs interact with E2-responsive enhancers to mediate gene transcription.

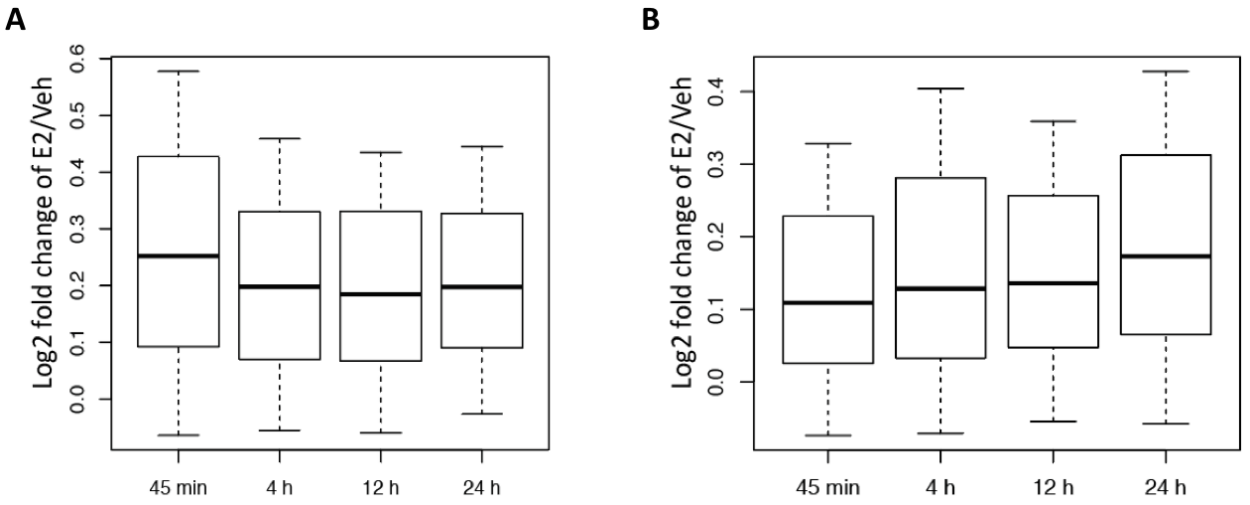

Figure 20. E2-responsive enhancer activities over time in E2-induced enhancers
A) with ER binding sites and B) without ER binding sites. ER binding sites were determined by ER ChIP-seq data of E2 treatment for 45 min. Fold of change of enhancer activity between E2 and vehicle treatment was calculated and plotted on y-axis in log2 scale.

We further classified E2-induced enhancers into four categories by k-means clustering based on their regulatory activity dynamics in response to E2 over time. We compared enhancers maximally induced at 45min and 4h (E2-early induced enhancers) to those maximally induced at 12h and 24h (E2-late induced enhancers) for the occurrence of known TF motifs. Consistent with motif analysis on DHSs, E2 early-induced enhancers are more enriched for ERE and AP-1 motifs, while E2 late-induced enhancers are more enriched for SP/KLF and AP-2 motifs (**Figure 21**), supporting the hypothesis that these TFs are activated by E2 in order and play regulatory function in primary and secondary response, respectively. Additionally, ER pioneer factors related motifs (forkhead, GATA and PBX1) are all enriched in E2 early-induced enhancers. Although CapSTARR-seq measures regulatory activity of exogenous DNA fragments without chromatin context, the plasmid DNA might be chromatinized in the cells so that pioneer factors may still play important roles in facilitating ER loading to enhancer fragments.

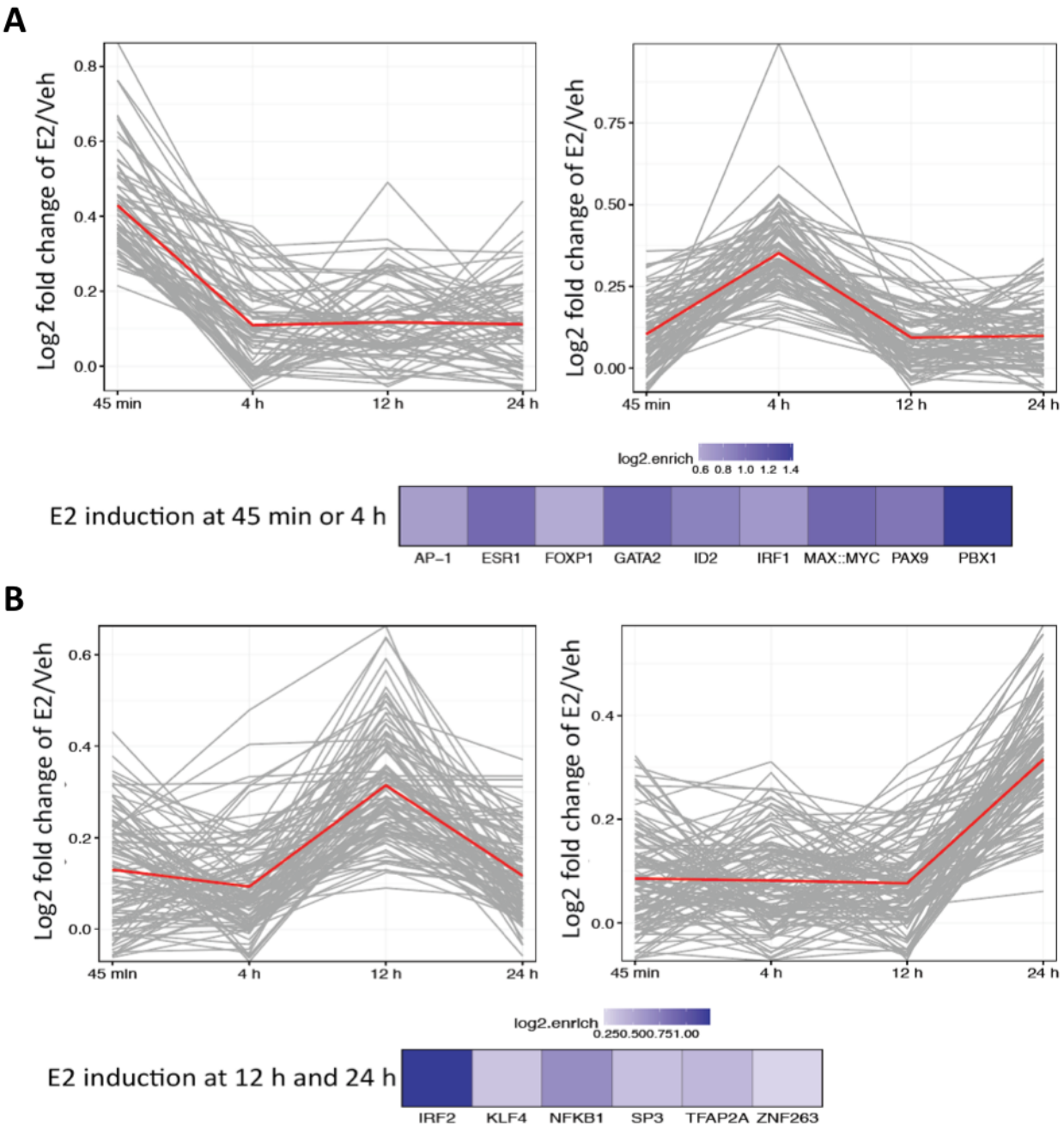

Figure 21. E2 dynamically induced enhancers and enriched TF motifs

A) Trajectory of the dynamics of E2 early-induced enhancers and enriched TF motifs. B) Trajectory of the dynamics of E2 late-induced enhancers and enriched TF motifs.

Notes:

a. E2-induced enhancers were clustered by k-means clustering. The number of clusters was determined by comparing sum of squares within clusters and between clusters. Red lines indicate the trajectory of centroid in each cluster.

b. TFs expressed in MCF7 cells (FPKM cut-off is 1) were included for motif analysis. Motif occurrence was predicted by FIMO using JASPAR database. Fisher's exact tests were performed

Figure 21. E2 dynamically induced enhancers and enriched TF motifs (continued)
comparing the frequency of motifs in E2 early-induced enhancers and E2 late-induced enhancers, and p values were adjusted for multiple testing. Enrichment of motif frequency in each class of enhancers was calculated and plotted in heat map in log2 scale.

2.3 Discussion

In this chapter, we have performed genome-wide investigation and functional assessment of cis-regulatory elements in the E2-dependent transcriptional regulation network. We started by profiling the dynamic changes of chromatin accessibility landscape in response to E2 stimulation together with the dynamics of gene expression regulated by E2. We identified thousands of E2-responsive DHSs associated with distinct chromatin signatures and gene correlations. We revealed various TFs including many novel ones that potentially interact with E2-responsive DHSs to modulate E2-dependent gene expression. We further experimentally interrogated the regulatory function of E2-responsive DHSs quantitatively in a high throughput manner, not only demonstrating that E2-responsive enhancers are frequently located in E2-responsive DHSs, but also uncovering the temporally coordinated action of four key TF families (ER, AP-1, SP/KLF, AP-2) at E2-responsive cis-regulatory elements.

Genomic analyses of chromatin accessibility data have revealed that ER binding sites and MCF7 cell HOT regions are strongly enriched in E2-induced DHSs. HOT regions are cis-regions that are coordinately bound by multiple TFs including NRs. They are closely associated with active chromatin marks, nearby gene transcription and sites of genomic amplification in breast cancer (Kittler et al. 2013). The frequent overlaps of ER binding sites, HOT regions and E2-induced DHSs suggest E2-triggered transcriptional regulation is predominantly mediated by ER, but more importantly, it is also highly dependent on other TFs and NRs. ER binding sites located in E2-induced DHSs are likely to be most relevant ones for transcriptional response, because despite the discrepancy among different datasets profiling ER genomic binding sites, the majority of ER binding sites in E2-induced DHSs are reproduced across multiple studies. The average ER ChIP-seq signal

in these regions is significantly higher than that in non-E2-responsive DHSs (**Figure 28**). ER occupancy is highly associated with early responsive E2-induced DHSs and enhancers (maximally induced at 45min after E2 treatment) but not with late induced DHSs and enhancers, indicating that much of ER activity as an upstream master regulator in the E2-governed gene regulation network is transient. It is reasonable to assume that secondary transcriptional regulation more heavily relies upon other TFs, and that it is initiated soon after E2 stimulation (sometime between 45min and 4h). Historically there has been significant interest to identify E2 signaling early-activated TFs/NRs, the mechanisms of their activation and how the activation is regulated. Combined results from differentially expressed gene detection and motif analysis have provided a short list of TFs that are potentially E2 primary transcription targets and regulators of E2-triggered secondary transcription, such as KLF4, FOSL2, EGR3, FOXC1, MYC, TFAP2C and so on.

Among the numerous motifs identified as enriched in E2-induced DHSs besides the ERE motif, the AP-1, SP/KLF and AP-2 family motifs occur in a large fraction of E2-induced DHSs whether or not ER binding sites are present, suggesting that these TFs widely exert regulatory function in E2-induced DHSs through both tethering and coordinated binding of ER. A number of studies on particular E2-responsive enhancers have demonstrated the importance of these TFs in mediating E2 effects through tethering ER to their binding sites. AP-1 was identified as an important TF mediating ER transcriptional regulation on non-classical ERE sites long before genome-wide studies were conducted (Kushner et al. 2000), and genome studies focusing on ER cistrome revealed AP-1 motifs frequently occur in ER binding sites (Carroll et al. 2005). In E2-induced DHSs, AP-1 motifs are present in the same DHS of ~67% ER binding sites, but interestingly AP-1 motifs occur in up to 80% of DHSs where we did not detect signal or detected very faint signal of ER binding, indicating that E2 signaling on gene transcription is extensively mediated by AP-1 TFs, and AP-1 TFs can function both with involvement of ER and without ER

on cis-regulatory elements. It has been demonstrated that the transactivation activity of AP-1 TFs can be directed by E2 non-genomic signaling through MAPK-mediated phosphorylation (Bjornstrom and Sjoberg 2004; Karin 1995). Sp1 from the SP/KLF family is another well-studied TF important for ERE-independent ER transcriptional regulation. The SP/KLF family consists of more than a dozen of TFs with similar structures of DNA binding domains (DBDs) that recognize GC-rich genomic sequences. Many genes have been found to be regulated by E2 through ER-Sp1 complexes at Sp1 binding sites, or Sp1 sites in proximity with ERE half-sites (Bjornstrom and Sjoberg 2005). While interaction between ER and Sp1 has been widely studied, little is known about the roles of other SP/KLF family members in E2-goverened transcriptional regulation. SP/KLF family TFs share highly conserved DBDs but they exert distinct functions in transcriptional regulation and interact with each other to coordinate transcription (Kaczynski, Cook, and Urrutia 2003). Our gene expression data indicates that KLF4 and KLF10 are among the handful of E2 up-regulated genes at 45min, and the KLF4 motif is over-represented in E2-induced DHSs and enhancers. Therefore, KLF4 or other KLF family TFs may play important roles in mediating E2 signaling on gene transcription. TFAP2C belonging to AP-2 family has been proposed as a collaborative factor important for hormone response in breast cancer. It transcriptionally regulates the expression of ER by interacting with ER promoter, and it also regulates expression of the GPR30 receptor for E2 (Woodfield et al. 2007). In addition, TFAP2C may act as "pioneer factor" of ER since it binds to ER binding sites prior to ER and loss-of-function of TFAP2C diminished ER genomic binding and long-range chromatin interactions (Tan et al. 2011). TFAP2C, like KLF4, is also an E2 early-induced gene at 45min revealed by RNA-seq data. According to motif analysis, dynamic enrichment of ERE, AP-1, SP/KLF and AP-2 motifs in E2-induced DHSs and enhancers suggests the sequential recruitment of these TFs on E2-responsive cis-regulatory elements and temporally coordinated actions of these TFs in E2-dependent gene regulation. Here, we propose

these four families of TFs participate in E2-triggered transcriptional regulation in the following way: ER is directly activated by E2 and initiates various intracellular signal transduction cascades in cytoplasm. Ligand-activated ER is translocated into nucleus and acts as master regulator in primary gene transcription. Meanwhile, AP-1 TFs are rapidly activated through signaling pathways like MAPK via post-translational modification, and regulate primary and secondary gene transcription through both ER-dependent and ER-independent mechanisms. KLF4 and TFAP2C are E2 primary regulated genes so that their expression is quickly induced with short-term E2 stimulation. They further act as key regulators in modulating secondary transcriptional response in E2 signaling (**Figure 22**).

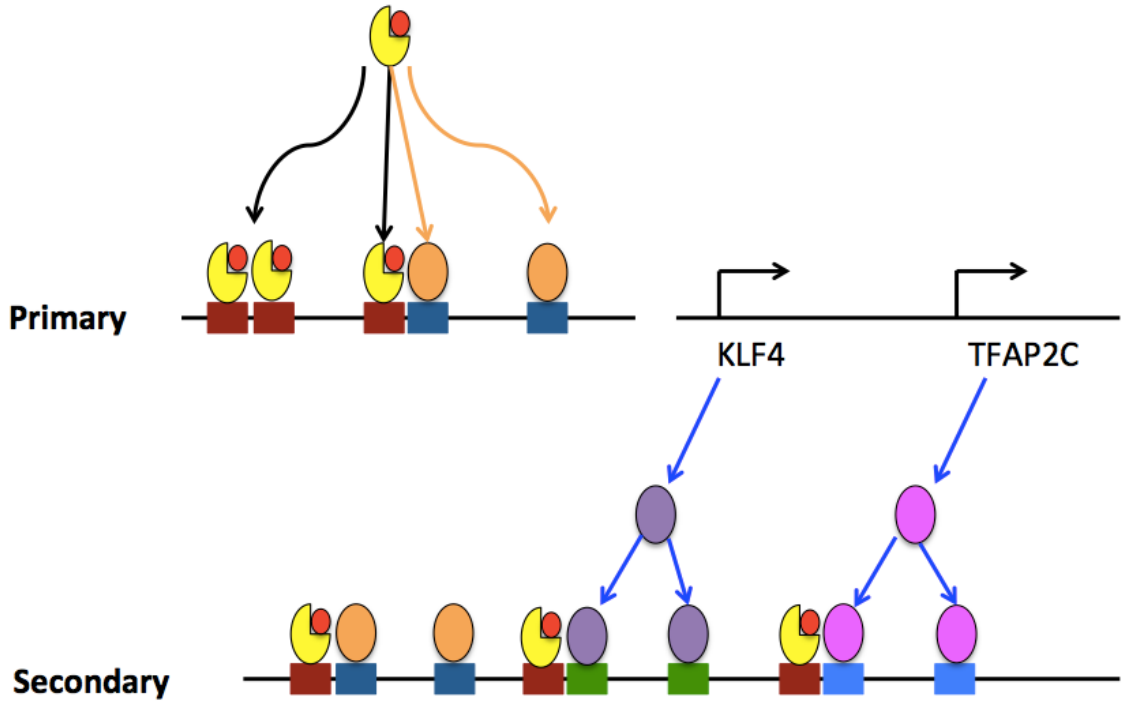

Figure 22. Model for action of ER, AP-1, SP/KLF4 and AP-2 family TFs in E2-governed primary and secondary transcriptional regulation

Notes:

a. TFs

Yellow nicked circle: ER. Orange circle: AP-1 TFs. Purple circle: SP/KLF TFs. Pink circle: AP-2 TFs.

b. Cis-regulatory elements

Dark red box: EREs. Dark blue box: AP-1 motif. Green box: SP/KLF motif. Light blue box: AP-2 motif.

c. Regulation

Black arrow: ligand-induced activation. Orange arrow: post-translational modification. Blue arrow: gene transcription and translation.

Previously, various studies using different genomic and computational approaches have predicted numbers of cis-regions involved in gene transcriptional regulation of E2 signaling. However, very few predicted regions have been functionally interrogated, which has been a major hurdle in building E2-governed gene regulation network. In our study, we directly assessed the E2-dependent regulatory activity of close to 10,000 E2-responsive cis-regions for the first time, and we successfully characterized more than 3,000 regulatory elements and more than 800 E2-responsive regulatory elements. Again, CapSTARR-seq results prove that DNase I hypersensitivity sites harbor numbers of cis-regulatory elements important for controlling gene expression. The regulatory activity of identified enhancers is closely associated with E2-regulated gene expression, and we also revealed that the number of E2-induced enhancers is positively correlated with the induction level of gene expression regulated by E2. For the classical E2-induced gene GREB1, we found 8 E2-induced enhancers located within 50 kb of its TSS, mediating the robust up-regulation of GREB1 transcription in response to E2. Since the CapSTARR-seq conducted in this study is limited to a subset of chromatin accessible regions, it would be plausible to expand the experiments to whole MCF7 genome so as to identify more E2-responsive cis-regulatory elements in a less biased way. Both Drosophila whole genome (Arnold et al. 2013) and human prostate cancer LNCap cell whole genome (submitted) STARR-seq data have revealed numerous cis-regions in closed chromatin with regulatory activity. It would be interesting to know whether there are E2-responsive cis-regulatory elements located in chromatin inaccessible regions. Such knowledge may shed light on the mechanisms of cell type specific hormone responses. Given STARR-seq and CapSTARR-seq still need to be further optimized technically to improve signal-to-noise ratio, the current data are not yet available to build quantitative models based on enhancer activity to infer associated gene expression outcome. However, STARR-seq and CapSTARR-seq can serve as powerful tools to annotate E2-dependent regulatory potential of cis-regulatory elements, and they can be applied in combination

with endogenous functional studies using CRISPR/Cas9-mediated large-scale genetic screen in coding and noncoding regions (Chen et al. 2015; Wright and Sanjana 2016).

2.4 Materials and methods

2.4.1 Cell culture and hormone treatment conditions

MCF7 cells (ATCC: HTB-22) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1% antibiotic-antimycotic. For hormone starvation, cells were cultured in phenol red free DMEM supplemented with 10% charcoal:dextran stripped FBS and 1% antibiotic-antimycotic. 10 nM of 17β-estradiol (Sigma) was used to stimulate cells after starvation, and ethanol was used as vehicle control.

2.4.2 DNase I hypersensitivity sites mapping

After 72 hours of hormone starvation, 2.0×10^7 cells (at ~80% confluency) were treated with E2 or vehicle. Cells were harvested at 45 min, 4 h, 12 h and 24 h post-treatment. Four biological replicates were generated under each condition. Nuclei isolation and DNase I digestion were performed as previously described (Song and Crawford 2010). Cells were lysed in buffer RSB containing 0.1% NP-40. Nuclei were pelleted, subjected to DNase I (Roche) digestion at 37°C for 15 min, and embedded into 0.5% low melt agarose gel (Lonza). High molecular weight DNA (50 kb – 1 Mb) was extracted from agarose gel plugs by phenol:chloroform purification, and equally pooled from nuclei digested by different amounts of DNase I (0.4 U, 1.2 U and 4.0 U). 1 ug DNase I cut DNA was used to make DNase-seq library. Library preparation was adapted from previous protocol (Song and Crawford 2010), but adaptors and PCR primers were redesigned in order to achieve library compatibility with Illumina platforms and simplify library preparation procedure (**Table 5**). Given the new design, adaptor I ligated DNA was sheared by sonication in Covaris S2 to generate DNA fragments with an average size of 150 bp following manufacturer's protocol, then directly

bound to streptavidin beads for enrichment of 5' end of DNase I cut fragments. DNA was further ligated to adaptor II on the beads and slightly PCR amplified using Fusion Hot Start Flex DNA Polymerase (NEB, PCR program: 98°C for 30s; 12 cycles of 98°C for 10s, 65°C for 30s and 72°C for 15s). After final PCR, streptavidin beads were removed on magnet, and reaction was cleaned up by SPRISelect beads (Beckman, beads/reaction ratio=0.85). DNase-seq libraries were on Illumina HiSeq2000/2500 platform in 50 bp single-end format.

Oligo name	Sequence (5'->3')
sonic_oligo_1a	Bio-ACACTCTTTCCCTACACGACGCTCTTCCGATC*T
sonic_oligo_1b	Phos-GATCGGAAGAGCGTCGTGTAGGGAAAGAGTG-Amm
sonic_oligo_2a	Phos-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
sonic_oligo_2b	CTGGAGTTCAGACGTGTGCTCTTCCGATC*T
sonic_primer_f	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTT
sonic_primer_r	CAAGCAGAAGACGGCATACGANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTT

Table 5 Sequence of adaptor oligos and PCR primers in DNase-seq library preparation

Notes:

Bio: biotin
*: phosphorothioate bond
Phos: phosphorylation
Amm: amino modifier.
NNNNNN: index sequence for multiplex sequencing

2.4.3 CapSTARR-seq experiment and library preparation

The overall workflow of generating input screening library was similar to previously published CapSTARR-seq (Vanhille et al. 2015). MCF7 genomic DNA extracted using PerfectPure DNA Cultured Cell Kit (5PRIME) was sheared by sonication (Covaris S2) following manufacturer's protocol and sized selected on 1% agarose gel to enrich for fragments from 350 to 650 bp. 1 ug size-selected gDNA was ligated to customized adaptors (**Table 6**) and subjected to NimbleGen SeqCap EZ Capture (Roche). Post-capture DNA was PCR amplified using Fusion Hot Start Flex DNA Polymerase to add homology arms for subsequent cloning (**Table 6**, PCR program: 98°C for 30s; 18 cycles of 98°C for 10s, 65°C for 30s and 72°C for 30s). The vector backbone was modified from pGL4.23[luc2/minP] (Promega). In brief, minP promoter was replaced with SCP1 promoter (kindly provided by Dr. Alexander Stark), luc2 was replaced with eGFP ORF, and a CmR-ccdB cassette was cloned between eGFP ORF and SV40 poly (A). The Post-capture PCR product was cleaned up by AmPureXP beads (Michailidou et al.), then cloned into the vector backbone (linearized by Sph I and Nde I) through Gibson Assembly (NEB) and transformed into MegaX DH10BTM T1R electrocompetent cells (ThermoFisher Scientific). A total of 16 Gibson Assembly reactions and transformations were performed. All transformed E.coli cells were pooled and cultured in 4 L LBAMP medium, and harvested when OD600 reached 1.0. The input screening plasmids were extracted using Plasmid Giga Kit (QIAGEN) and drop dialyzed. For library transfection, input screening plasmids were electroporated into MCF7 cells using BTX Electrofusion Systems (Harvard Apparatus) according to manufacturer's protocol (1 ug DNA/1 million cells). After transfection, cells were plated into starvation medium for 48 hours, then treated with E2 or vehicle. Cells were harvested at 45 min, 4 h, 12 h and 24 h post-treatment. 3.0×10^7 cells were electroporated in each replicate and two replicates were generated under each condition. For CapSTARR-seq library preparation, total RNA was extracted using PerfectPure RNA Cultured Cell Kit (5PRIME) followed

by poly (A) tail mRNA isolation using Dynabeads Oligo (dT) beads (ThermoFisher Scientific). mRNA was treated with TURBO DNase (ThermoFisher Scientific) and purified by RNAClean XP beads (Michailidou et al.). Target-specific first strand cDNA synthesis was performed with SuperScript III (ThermoFisher Scientific, 500 ng mRNA/reaction) to specifically reverse transcribe mRNA produced from screening library (**Table 6**). After reverse transcription, reactions were treated with RNase A+H. In final PCR amplification, every 5 ul cDNA reaction was used as template in every 50 ul PCR reaction using Fusion Hot Start Flex DNA polymerase (**Table 6**, PCR program: 98°C for 30s; 21 cycles of 98°C for 10s, 65°C for 30s and 72°C for 30s). PCR products were cleaned up by AmPure XP beads (beads/reaction ratio=0.8). Generating sequencing library for input screening plasmids followed the same PCR setup as generating CapSTARR-seq library. Products from two PCR reactions were pooled (10 ng plasmid/reaction, 12 cycles of PCR) and cleaned up. Input screening libraries and CapSTARR-seq libraries were sequenced on Illumina HiSeq2000/2500 platform in 100 bp paired-end format.

Oligo name	Sequence (5'->3')
Adaptor oligo_1	GATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Adaptor oligo_2	Phos-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
pre-capture PCR primer 1	TAATTCTAGAGTCGGGGCGGGCATGGATCTACACTCTTTCCCTACACGAC
pre-capture PCR primer 2	CTTATCATGTCTGCTCGAAGCGGCAGTGACTGGAGTTCAGACGTGT
Capture_blocker1	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATC
Capture_blocker2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Amm
cDNA synthesis primers	CAGTGACTGGAGTTCAGACG
	GCAGCTTATAATGGTTACAAATAAAGC
	ACATTTGTAGAGGTTTTACTTGCT
	TCAATGTATCTTATCATGTCTGCTCGAAG
Final PCR primer 1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA*C
Final PCR primer 2	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTG*T

Table 6 Sequence of oligos used in CapSTARR-seq

Notes:

Phos: phosphorylation
Amm: amino modifier
*: phosphorothioate bond
NNNNNN: index for multiplex sequencing

2.4.4 Gene expression profiling and differentially expressed gene detection

After 72 hours of hormone starvation, 1.0×10^6 cells (at ~80% confluency) were treated with E2 or vehicle. Cells were harvested at 45 min, 4 h, 12 h and 24 h post-treatment. Two biological replicates were generated under each condition. Total RNA was extracted from E2/vehicle treated cells using MagNA Pure Compact RNA Isolation Kit (Roche). RNA-seq library preparation was conducted using TruSeq Stranded mRNA Library Prep Kit (Illumina). RNA-seq libraries were sequenced on Illumina HiSeq2000/2500 platform in 50 bp single-end format. RNA-seq raw reads were mapped to human genome hg19 using TopHat. Transcript assembly (RefSeq annotation) and gene FPKM estimation were done by Cufflinks. Read counts of genes were summarized by HTSeq, and differentially expressed genes between E2 and vehicle at each time point were detected by edgeR (Robinson, McCarthy, and Smyth 2010) using exact negative binomial test (FDR<0.01 and average FPKM > 1).

2.4.5 ChIP-seq and peak calling

2.0×10^7 cells (at ~80% confluency) were treated with E2/vehicle for 45 min after 72 hours of hormone starvation. ChIP experiment was done as previously described (Hua et al. 2008). Chromatin was lysed in 0.1% SDS lysis buffer and sheared by sonication (Covaris S2, intensity=4, duty cycle=5%, cycles per burst=200, time=7min). Chromatin immunoprecipitatino was performed using anti-ERα (Santa Cruz: HC-20) antibody. ChIP-seq library preparation was performed as previously described (Hua et al. 2008) using adaptors and primers listed in **Table 7**. ChIP-seq libraries were sequenced on Illumina HiSeq2000/2500 platform in 50 bp single-end format. Raw sequencing reads were aligned to human genome hg19 by BWA using default parameters. Only uniquely mapped reads were used in downstream analysis. ChIP-seq peaks were called using MACS2 (Zhang et al. 2008) with default parameters (FDR<0.05).

Oligo name	Sequence (5'->3')
Adaptor oligo 1	GATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
Adaptor oligo 2	Phos-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Final PCR primer 1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA*C
Final PCR primer 2	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTG*T

Table 7 Sequence of oligos used in ChIP-seq library preparation

Notes:

*: phosphorothioate bond

Phos: phosphorylation

NNNNNN: index for multiplex sequencing

2.4.6 Dual luciferase reporter assay

Luciferase reporter constructs were manufactured by VectorBuilder. In brief, candidate regions were PCR amplified using MCF7 genomic DNA as template, and cloned into pGL4.23[luc2/minP] multiple cloning site which is upstream of the minP promoter. The day prior to transfection, 1.0×10^4 cells were plated into 96-well plates in starvation medium. During transfection, Firefly luciferase reporter plasmids were co-transfected with NanoLuc luciferase control plasmid pNL1.1.TK[Nluc/TK] (Promega) using ViaFect Transfection Reagent (Promega). For each well, 100 ng total DNA (Firefly plasmid : NanoLuc plasmid = 99:1) was mixed with 0.3 ul transfection reagent in 10 ul serum-free medium and incubated for 10 min at room temperature before added into cells. 24 hours after transfection, cells were treated with E2 or vehicle. Luciferase assays were performed using Nano-Glo Dual-Luciferase Reporter Assay Kit (Promega) at desired time points post E2/vehicle treatment on Synergy HT plate reader (BioTek) following manufacturer's protocol.

2.4.7 DNase-seq data processing, peak calling and differential peak detection

Raw sequencing reads were aligned to human genome hg19 by BWA using default parameters. Only uniquely mapped reads were used for downstream data analysis. Peaks for individual replicates were called by F-Seq (Boyle, Guinney, et al. 2008) (default parameters with ploidy and copy number correction turned on), and high confident peaks under each condition were identified by IDR analysis (IDR<0.01). 5' end cut counts on DHS union peaks from different conditions were summarized and analyzed by edgeR to detect E2 responsive DHSs. E2 constantly responsive DHSs were detected using GLM likelihood ratio test after fitting data into negative binomial GLM model ($p<0.001$). Differential DHSs between E2 and vehicle treatments at each time point were detected using exact negative binomial test ($p<0.001$).

2.4.8 CapSTARR-seq data processing and peak calling

Raw sequencing reads were aligned to human genome hg19 by BWA using default parameters. Reads mapped in proper pair were used and biological replicates were combined for downstream analysis. For each condition, peak summits were called by MACS2 using default parameters (FDR<0.05), and 300 bp windows centered on peak summits were identified as peaks. Union peak set was defined by combining peaks from different conditions (peaks that summits were within 100 bp were merged). Fold of enrichment over input for union peaks were calculated using total fragment counts, and peaks with lowest 20% complexity (calculated as # of distinct fragments / region length (bp)) were removed from analysis. Enhancer activity was estimated by normalizing fold of enrichment by the average fold of enrichment of complete negative control regions. Requirements of being E2 induced enhancers are a) enhancer activity > 1.5 for at least one E2 condition; b) fold change of E2 versus vehicle > 1.2 for at least one time point; c) minimum fold change of E2 versus vehicle across all time points > 0.95. Similarly, requirements of being E2 repressed enhancers are a) enhancer activity > 1.5 for at least one vehicle condition; b) fold change of E2 versus vehicle < 0.8 for at least one time point; c) maximum fold change of E2 versus vehicle across all time points < 1.05.

2.4.9 Analysis of chromatin signal intensity data and HOT region data

ChIP-seq data of chromatin marks H3K14ac, H3K27me3, H3K9me3, H3K4me3, H3K9ac and RNA polymerase II was obtained from GEO dataset GSE23701 (Joseph et al. 2010; Kong et al. 2011). ChIP-seq data of chromatin mark H3K27ac was obtained from GEO dataset GSE45822 (Li, Notani, et al. 2013). ER ChIP-seq data of E2 treatment for 4 hours was obtained from GEO dataset GSE54855 (Guertin et al. 2014). List of HOT regions was obtained from previously published paper (Kittler et al. 2013). Data that was aligned to hg18 was lifted over to hg19 annotation. Average tag

count in the +/- 3 kb regions flanking the center of interested DHSs was computed using coverage function of BEDTools and scaled by library size equivalent to 10 million reads. One-tailed Wilcox rank tests were performed on normalized average ChIP-seq signal. To make meta plot of the signal intensity of repressive marks on different categories of DHSs, average ChIP-seq tag count in every 25 bp windows within +/- 3 kb regions flanking the center of interested DHSs was computed and scaled by library size. To find overlapping sites between HOT regions, ER binding sites and DHSs, intersect function of BEDTools was used following default parameters. To generate heatmap of ChIP-seq signal in interested regions, data matrix of ChIP-seq tag count in every 25 bp window of +/- 5 kb regions flanking the center of DHSs was computed using annotatePeaks function in HOMER software (Heinz et al. 2010). Data matrix was imported into Cluster 3.0 software and re-arranged by gene centered correlation method, then visualized in Java Treeview software (Saldanha 2004).

2.4.10 Clustering analysis of E2-regulated genes and E2-induced enhancers

K-means clustering was performed using pheatmap function of pheatmap package in R based on log2 transformed fold of changes of FPKM between E2 treatment and vehicle treatment for gene expression data, and log2 transformed fold of changes of enhancer activity between E2 treatment and vehicle treatment for CapSTARR-seq data. The number of clusters was determined by plotting the within cluster sum of squares as a function of k. For gene expression data, genes with maximum expression change over 1.5 times of log2 fold of change were included. Centroid data in each cluster was extracted after clustering, and plotted using ggplot2.

2.4.11 Association between cis-regions (DHSs, enhancers) and genes

DHSs or enhancers were assigned to nearby genes if they are located within 50 kb of any TSS using window function of BEDTools. Odds ratio representing the association between DHSs and E2-regulated genes was computed using the following formula: (# of regulated genes with >= 1 nearby DHSs / # of non-regulated genes with >= 1 nearby DHSs) / (# of regulated genes with 0 nearby DHS / # of non-regulated genes with 0 nearby DHS). Only genes with FPKM greater than 1 were included for DHS association analysis.

2.4.12 Motif analysis

PWMs of known TF motifs were obtained from JASPAR 2016 vertebrate core motif database (Mathelier et al. 2016). TFs not expressed in MCF7 cells (FPKM < 1) were filtered from analysis. Occurrence of TF motifs in DHSs was identified by FIMO (Grant, Bailey, and Noble 2011) using default cut-off. One-tailed Fisher's exact tests were performed and p values were adjusted by Bonferroni correction.

2.4.13 DNase I footprint analysis

DNase-seq reads were trimmed to only remain the first base pair on 5' end. DNase I footprints in E2-responsive DHSs were identified using dnase2TF software (Sung et al. 2014) with default parameters. Footprints of ERE, JUN, KLF4 and TFAP2C were predicted by overlapping motif coordinates extracted from FIMO results with identified DNase I footprints using intersect function in BEDTools. Average DNase-seq 5' end cut count was computed for each base pair within +/- 50 bp regions flanking predicted TF footprints. One-tailed Wilcox rank tests were formed to compare the signal of DNase I cleavage on footprints.

2.5 Contributions

Grace Yu and Kevin White designed the project. Grace Yu adapted and optimized original DNase-seq protocol with the help of Yuwen Liu. Jie Zhou and Yuwen Liu adapted and optimized CapSTARR-seq. Grace Yu generated all experimental data with the help of Amber Thomas, Lijia Ma and Mike Bolt. Grace Yu performed all data analyses and composed all figures and tables. Grace Yu and Kevin White wrote the manuscript.

CHAPTER 3

Functional characterization of KLF4 in estrogen governed transcriptional regulation

3.1 Introduction

KLF4 belongs to the Krüppel-like transcription factor family, which binds to a 5'-CACCC-3' core sequence to regulate target gene expression both as activator and as repressor in a context-dependent manner (Evans and Liu 2008). Its target genes are involved in many cellular functions, particularly in cell proliferation, apoptosis and differentiation. It is extensively recognized that KLF4 plays an important role in maintaining "stemness" of embryonic stem cells (ESCs). After introducing expression of Klf4 in combination with three other TFs (Oct3/4, Sox2 and c-Myc), adult fibroblasts can be reprogrammed into induced pluripotent stem cells (iPSCs) that resemble ES cell properties, including ability to differentiate into a variety of tissues (Takahashi and Yamanaka 2006). In cancers, more and more studies have revealed the emerging roles of KLF4 as either a tumor suppressor or an oncogene depending on cancer type. However, its cancer specific targets and gene regulation mechanism remain largely unknown. For example, the loss of KLF4 expression has been shown to be correlated with advanced tumor malignancy and accelerated tumor progression in colon cancers (Dang et al. 2001) and gastric cancers (Wei et al. 2005) due to promoter methylation or loss of heterozygosity. KLF4 exerts tumor suppressive function at least in part through its transcriptional regulation target, p21, which is a CDK inhibitor important for cell cycle

arrest (Chen et al. 2001), and through its transcriptional regulation target p53 (el-Deiry et al. 1993). KLF4 has also been found to be often over-expressed in squamous cell carcinoma (Foster et al. 1999) and breast cancer (Foster et al. 2000), suggesting its potential oncogenic role in these cancers. One way KLF4 is thought to exhibit oncogenic function is to through transcriptional repression of p53 expression, allowing for resistance to DNA damage induced apoptosis (Rowland, Bernards, and Peeper 2005).

Compared to other ubiquitously expressed KLF family members, KLF4 is only expressed in a subset of tissues, for example, embryonic stem cells, gastrointestinal tract, skin, lung, bone marrow, breast and ovary. Elevated expression of KLF4 mRNA and protein have been detected in most breast ductal and invasive carcinomas compared to adjacent uninvolved epithelium cells (Foster et al. 2000), and the nuclear localization of KLF4 is associated with more aggressive phenotype in early-stage breast cancer (Pandya et al. 2004). A few molecular studies have uncovered KLF4 tumor related function in breast cancer cell lines and mouse xenograft models. Loss and gain of function experiments in both ER+ and ER- breast cancer cell lines demonstrated KLF4 expression is positively associated with breast tumorigenesis by investigating phenotypes such as proliferation, apoptosis, cell migration, cell invasion, epithelial-mesenchymal transition (EMT), and xenografted tumor growth (Rowland, Bernards, and Peeper 2005; Yu et al. 2011; Hu et al. 2012; Tiwari et al. 2013). It has been found that as a rapid turnover protein, KLF4 protein stability is highly regulated by estrogen signaling through estrogen induced down regulation of pVHL, a protein that governs KLF4 turnover. KLF4 protein accumulation in turn facilitates estrogen triggered mitogenic growth (Hu et al. 2012). Furthermore, arginine methylation of KLF4 regulated by PRMT5 is also crucial for stabilizing KLF4 protein, and abruption of KLF4 methylation results in up-regulation of Bax and down-regulation of a series of oncogenes like cyclin D2, cyclin B1 and Myc in breast cancer (Hu et al. 2015). In addition, genome-wide association studies (GWASs) have also

revealed that KLF4 is potentially an important master regulator in breast cancer, as one breast cancer associated risk allele rs471467 located near KLF4 was found to physically interact with KLF4 promoter, and the KLF4 motif has been shown to be over-represented in promoter regions of this eQTL associated genes, which suggests that this breast cancer risk allele may affect downstream gene expression by acting through KLF4 (Li, Seo, et al. 2013). However, there are many major questions remained to be answered about KLF4 function in breast cancer. First of all, which are KLF4 transcriptional regulation targets that are associated with breast cancer, and how does KLF4 modulates its target gene expression? Secondly, whether and how does KLF4 transcriptional regulation cross-talk with major signaling pathways in breast cancer, including the estrogen receptor, progesterone receptor and epidermal growth factor receptor signaling? Though several studies found KLF4 physically interacts with ER in MCF7 cells (Akaogi et al. 2009; Mohammed et al. 2013), and KLF4 knockdown attenuated E2 response of ERE driven luciferase reporter gene expression (Hu et al. 2012), we do not have an overall picture of the role of KLF4 in E2-governed transcriptional regulation. Therefore, in this chapter, we have applied genomic approaches to functionally characterize KLF4 in E2-dependent transcriptional regulation, and we have identified KLF4 as an early effector of E2 signaling and a cooperating TF in the E2 gene regulation network.

3.2 Results

3.2.1 KLF4 is an E2 early up-regulated effector TF

In Chapter 2, we discovered that the KLF4 motif is over-represented in E2 induced DHSs and E2 induced enhancers. In order to ask whether KLF4 is a functional TF in MCF7 cells, we first examined KLF4 expression in response to E2. In MCF7 cells, KLF4 is a moderately expressed gene, and it is detected as one of just a handful of differentially expressed genes within 45 min of E2 stimulation from RNA-seq data (exact test based on negative binomial model, FDR < 0.001). As shown in **Figure 23A**, KLF4 mRNA was rapidly up-regulated upon E2 treatment for 45 min and retained this increasing expression over time. Noticeable accumulation of KLF4 protein was also observed upon E2 treatment for 4 h (**Figure 23B**). From ChIP-seq data, we found that there is a strong ER binding site located ~50 kb upstream of KLF4 TSS (**Figure 23C**). Therefore, KLF4 is likely to be an E2 primary target gene, and its transcription is directly and rapidly regulated by ER.

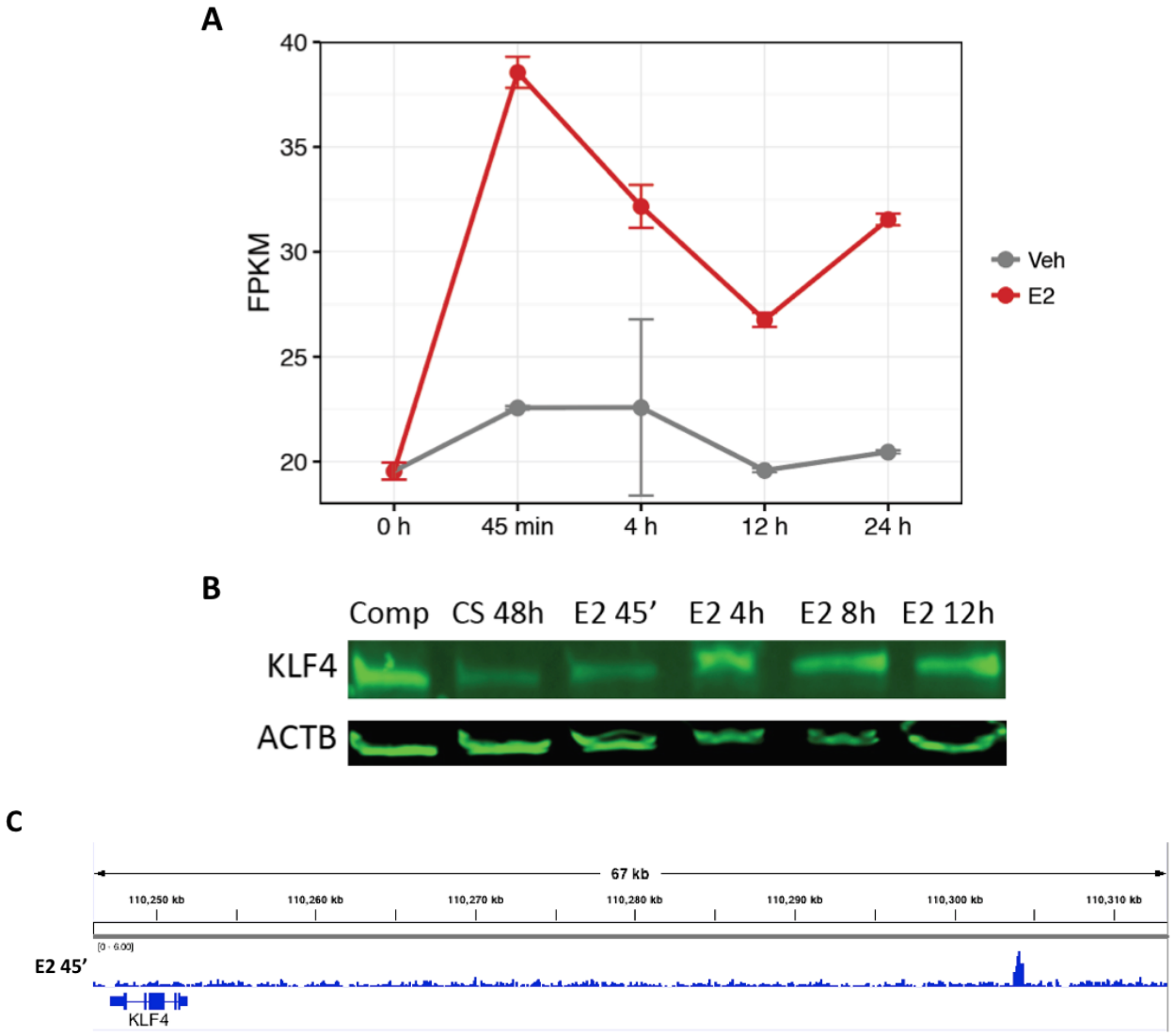

Figure 23. KLF4 is an E2 early up-regulated gene

A) KLF4 mRNA level upon E2 treatment over time measured by RNA-seq. Error bars denote standard error from two biological replicates. B) KLF4 protein level upon E2 treatment over time. Comp: complete media. CS 48h: hormone deprivation for 48 hours. C) Snapshot showing ER binding site near KLF4 TSS. ER ChIP-seq library size was normalized to 10 million reads.

3.2.2 KLF4 knockdown attenuated E2-dependent cell proliferation

In order to interrogate KLF4 function in E2-induced cell growth, we built a KLF4 loss-of-function model using shRNA mediated stable knockdown. We compared three different shRNAs targeting regions of human KLF4 mRNA (**Table 8**). Though the three shRNAs yielded different efficiency of reducing KLF4 level, we observed similar phenotypic changes in all three KLF4 stable knockdown cells. Among the three shRNA we tested, #3 shRNA targeting the 3rd and 4th exons and spanning exon-exon junction most effectively reduced KLF4 mRNA and protein level, to about one third of scrambled control knockdown (**Figure 24A and 24B**). Meanwhile, we did not observe noticeable change of ER protein level in KLF4 knockdown cells (**Figure 24B**). Therefore, we used #3 shRNA for downstream experiments.

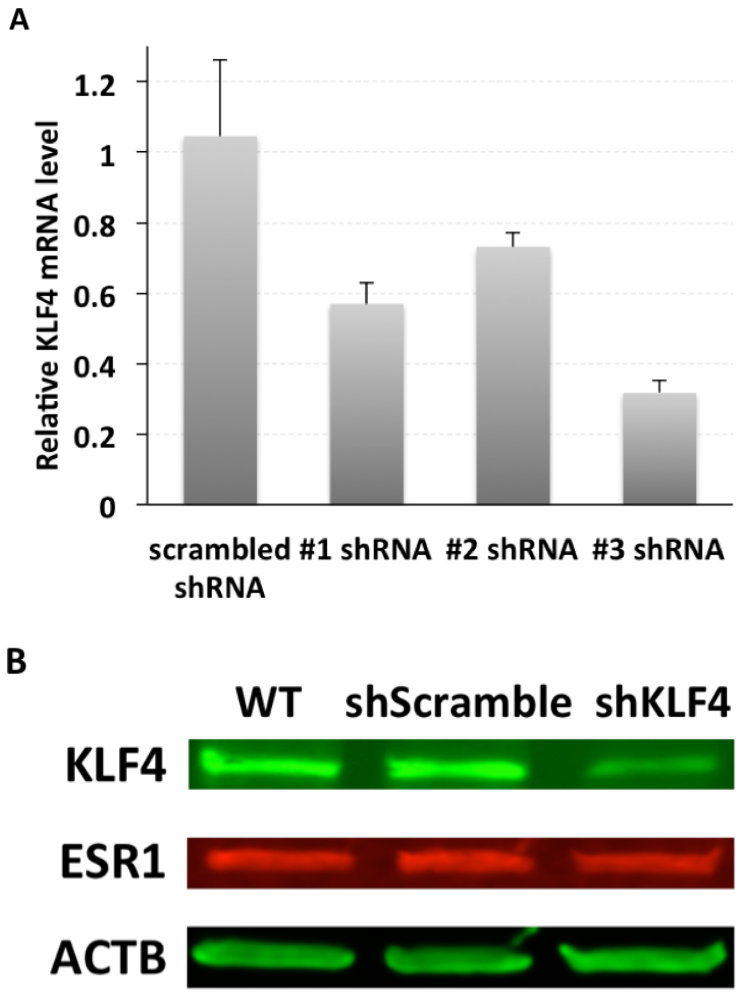

Figure 24. shRNA mediated KLF4 knockdown in MCF7 cells
A) KLF4 mRNA level in different lentiviral shRNA transduced cells measured by RT-qPCR. Error bars represent standard error among three biological replicates. B) KLF4 and ER protein level in lentiviral shRNA transduced cells.

Compared to scrambled control cells, KLF4 knockdown cells were obviously slowed down in cell growth, and many dead cells were seen in KLF4 knockdown cell culture dishes. We therefore measured KLF4 knockdown and scrambled control cell proliferation in response to E2 using an EdU proliferation assay, which precisely labels cells in S-phase of mitotic cell cycle. We tested the contribution of KLF4 expression level and E2 treatment to cell proliferation using the following linear model with interaction term:

$$\text{S-phase cell\%} = \beta_0 + \beta_1 \cdot \text{KLF4} + \beta_2 \cdot \text{E2} + \beta_3 \cdot \text{KLF4} \times \text{E2} + \varepsilon$$

While KLF4 knockdown in hormone deprived vehicle treatment condition did not significantly affect cell proliferation, KLF4 knockdown significantly attenuated E2-induced cell proliferation with E2 treatment for both 24 hours and 48 hours (**Figure 25**), demonstrating that KLF4 is required for maintaining E2 effect on cell proliferation.

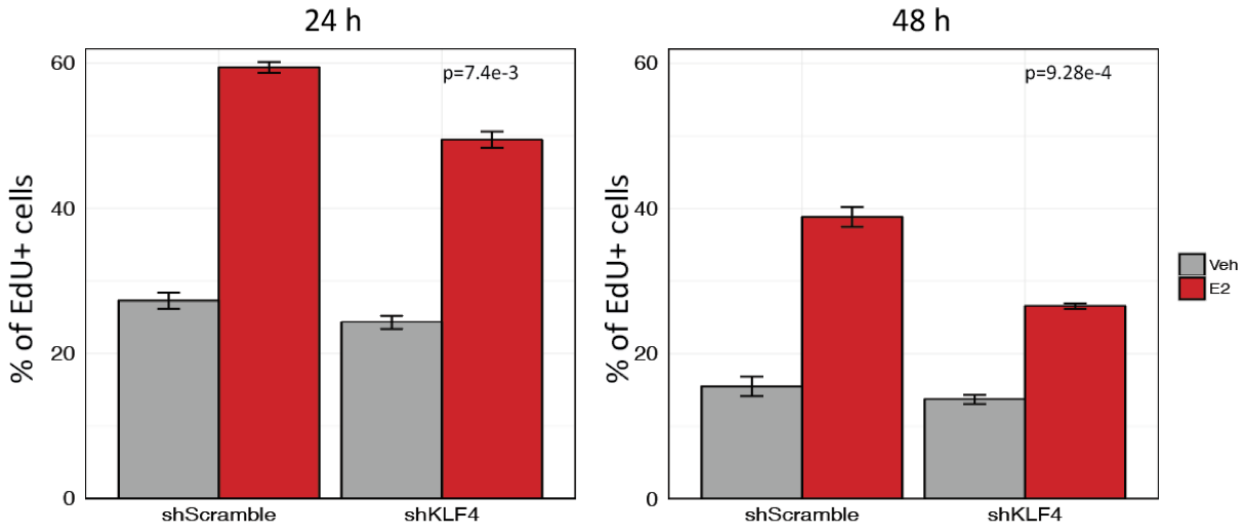

Figure 25. Effect of KLF4 knockdown in E2-dependent cell proliferation
Percentage of EdU+ cells with E2 treatment were plotted on y-axis with E2 treatment for A) 24 hours and B) 48 hours. Error bars represent standard error among three biological replicates. P-values denote the significance level of the interaction between KLF4 expression level and E2 treatment in the linear model.

3.2.3 KLF4 is a cooperative TF in E2-goverened gene regulation

Since KLF4 knockdown had significant impact on E2-regulated cell proliferation, we were interested in the genes regulated by both E2 and KLF4. We identified 2,490 genes down-regulated and 1,296 genes up-regulated in KLF4 knockdown cells compared to scrambled control cells from RNA-seq data (exact test on negative binomial model, FDR < 0.001), which are potentially KLF4 activated (up-regulated) and repressed (down-regulated) genes, respectively. We found that there is a substantial overlap between E2 regulated genes and KLF4 regulated genes. About half of E2 regulated genes were also affected after knocking down KLF4, and both E2 and KLF4 regulated genes were significantly enriched for coordinated up-regulation (Fisher's exact p < 2.2e-16) and coordinated down-regulation (Fisher's exact p = 1.048e-3, **Figure 26A**). Gene ontology enrichment analysis revealed that both E2 and KLF4 up-regulated genes are extremely enriched for genes related to cell cycle and DNA metabolic process. Such enrichment is far higher than E2 only up-regulated genes (**Figure 26B**), indicating that KLF4 is heavily involved in regulating E2-induced cell proliferation. In order to investigate whether KLF4 is important for E2-dependent gene regulation, we performed RNA-seq in KLF4 knockdown versus scrambled control knockdown cells with and without E2 treatment for 24 hours. While knocking down KLF4 did not seem to widely confer E2-dependent gene regulation, we were able to detect interactive effects of E2 and KLF4 knockdown on expression of 340 genes (GLM likelihood ratio test on negative binomial model, FDR < 0.05). Among these genes, KLF4 knockdown resulted in decreased E2 response for 83% of the genes and increased E2 response for 17% of the genes. Though both putative synergistic and antagonistic effects of KLF4 in E2-dependent gene regulation have been observed, and context-dependent mechanisms may exist determining whether KLF4 can activate or repress transcription, the overall

outcome of gene regulation and cell proliferation suggest that KLF4 plays a cooperative role in E2-governed gene regulation.

Next, we asked whether there is potential cooperative crosstalk at transcriptional regulation between ER and KLF4. We profiled 5,216 ER binding sites and 6,896 KLF4 binding sites in MCF7 cells by ChIP-seq (FDR < 0.05). We found both ER and KLF4 bindings sites near 50 kb of TSS of 212 out of 711 genes that are up-regulated by both E2 and KLF4. The enrichment of both TF binding sites near genes up-regulated by both E2 and KLF4 are significantly higher than genes only up-regulated by E2 or KLF4 (Fisher's exact p = 2.429e-3). Furthermore, among the 212 genes associated with both ER and KLF4 binding sites, we found ER and KLF4 binding sites co-occur in 106 regions. The co-occurrence is also more frequent than ER and KLF4 binding sites near genes up-regulated only by E2 or KLF4 (Fisher's exact p = 5.004e-3, **Figure 26C**), indicating significant cooperation at the transcriptional regulation level between KLF4 and ER.

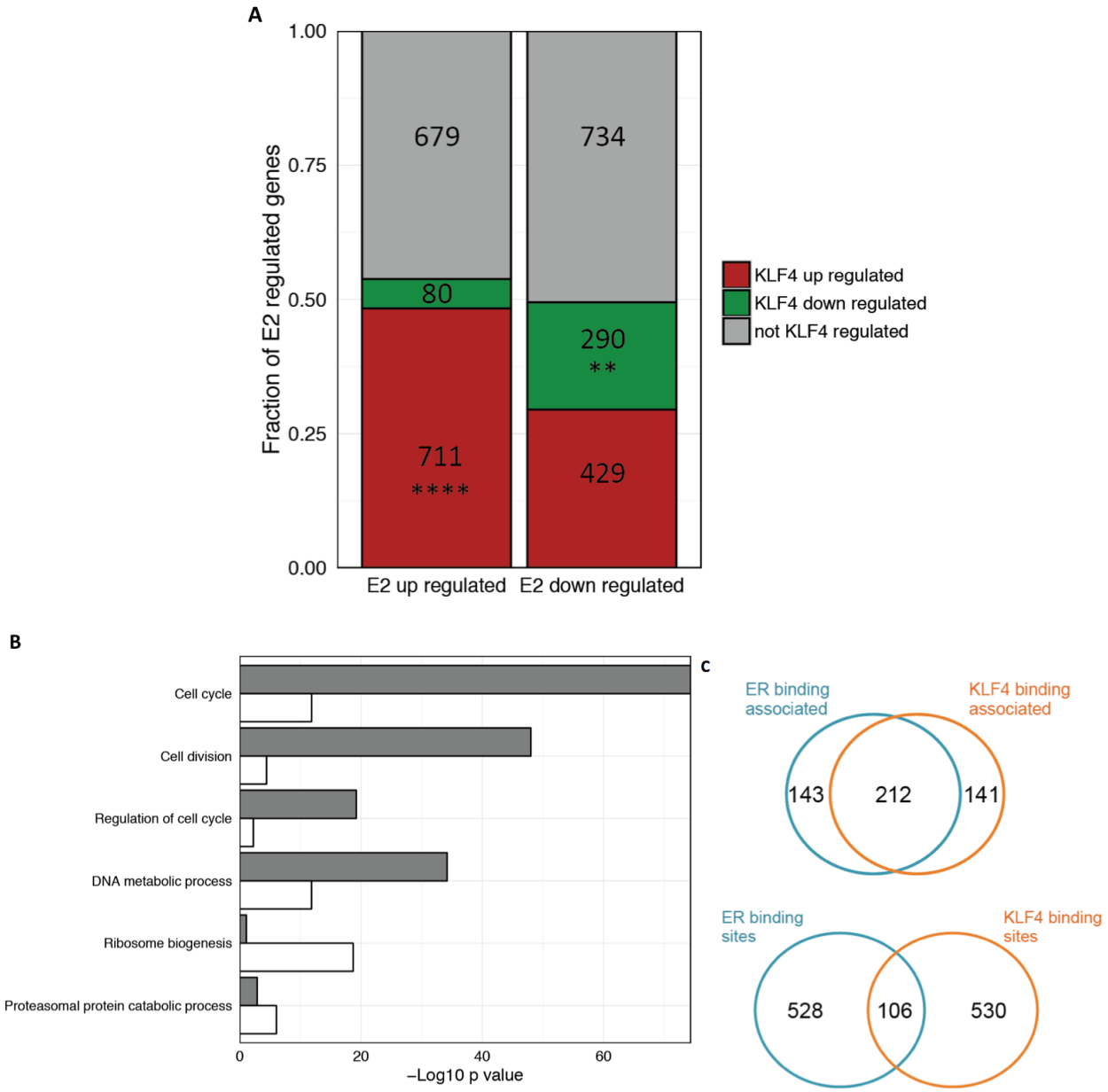

Figure 26. E2 and KLF4 regulated genes

A) Fraction of KLF4 regulation on E2-regulated genes. Fisher's exact tests were performed comparing the ratio between KLF4 up- and down-regulated genes that were also E2-regulated genes to the ratio between KLF4 up- and down-regulated genes in the whole transcriptome. ****: $p < 0.0001$. **: $p < 0.01$. B) Significantly enriched GO terms in E2 and KLF4 up-regulated genes. Fisher's exact tests were performed comparing gene list of interest to all expressed genes in MCF7, and p values were adjusted for multiple testing. –log10 transformed adjusted p values were plotted on x-axis. C) E2 and KLF4 co-up regulated genes associated with ER and KLF4 binding sites. Up: overlap between genes associated with ER and KLF4 binding sites. Down: overlap between ER and KLF4 binding sites associated with co-up regulated genes. Only bindings sites within 50 kb of TSS were included.

3.2.4 KLF4 and ER highly co-bind to E2 induced DHSs

409 out of 2,258 E2 induced DHSs contain high confident KLF4 ChIP-seq peaks, which confirms that many KLF4 motifs predicted from motif analysis of DNase-seq data are true positives. Although in all DHSs, KLF4 and ER do not co-occur frequently, their co-occurrence is much more frequent in E2 induced DHSs (Fisher's exact $p < 2.2e-16$, **Figure 27A and 27B**). From the heat map of ER and KLF4 ChIP-seq raw signal in E2 induced DHSs (**Figure 27B**), we found ER and KLF4 widely co-occur in E2-induced DHSs. We performed ChIP-qPCR on several randomly selected E2 induced enhancers that contain both ER and KLF4 binding sites in cells treated with E2 or vehicle for 45 min. The recruitment of both ER and KLF4 on those enhancers was rapidly induced by E2 treatment (**Figure 28**), indicating that KLF4 may play regulatory role as well as ER by interacting with those cis-regulatory elements in early-stage E2-goverened transcriptional regulation. However, KLF4 knockdown only moderately decreased ER occupancy genome wide, and the reduction level of ER occupancy is indistinguishable between ER and KLF4 co-bound sites and ER only sites, suggesting that KLF4 is not required for the interaction between ER and chromatin (**Figure 29**).

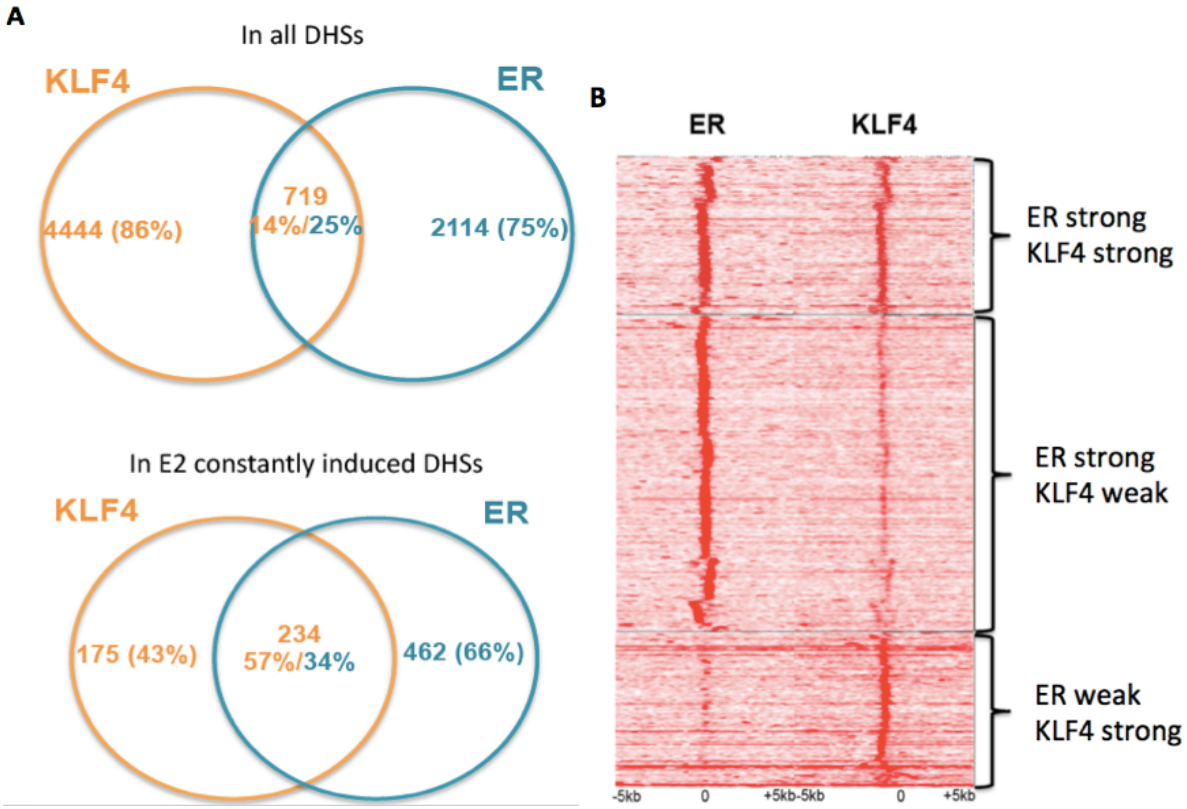

Figure 27. Co-occurrence of ER and KLF4 binding sites in E2 constantly induced DHSs
A) Overlap between DHSs containing ER and KLF4 binding sites. B) Heatmap of ER and KLF4 ChIP-seq signal in E2 constantly induced DHSs. ChIP-seq tag count was normalized to 10 million reads and plotted centered by DHS center. ER and KLF4 ChIP-seq peaks were called by MACS2 (FDR < 0.05) and overlapped with E2 induced DHSs. ER strong & KLF4 strong: E2 induced DHSs containing both ER and KLF4 high confident ChIP-seq peaks. ER strong & KLF4 weak: E2 induced DHSs containing high confident ER peaks but not KLF4 peaks. ER weak and KLF4 strong: E2 induced DHSs containing high confident KLF4 peaks but not ER peaks.

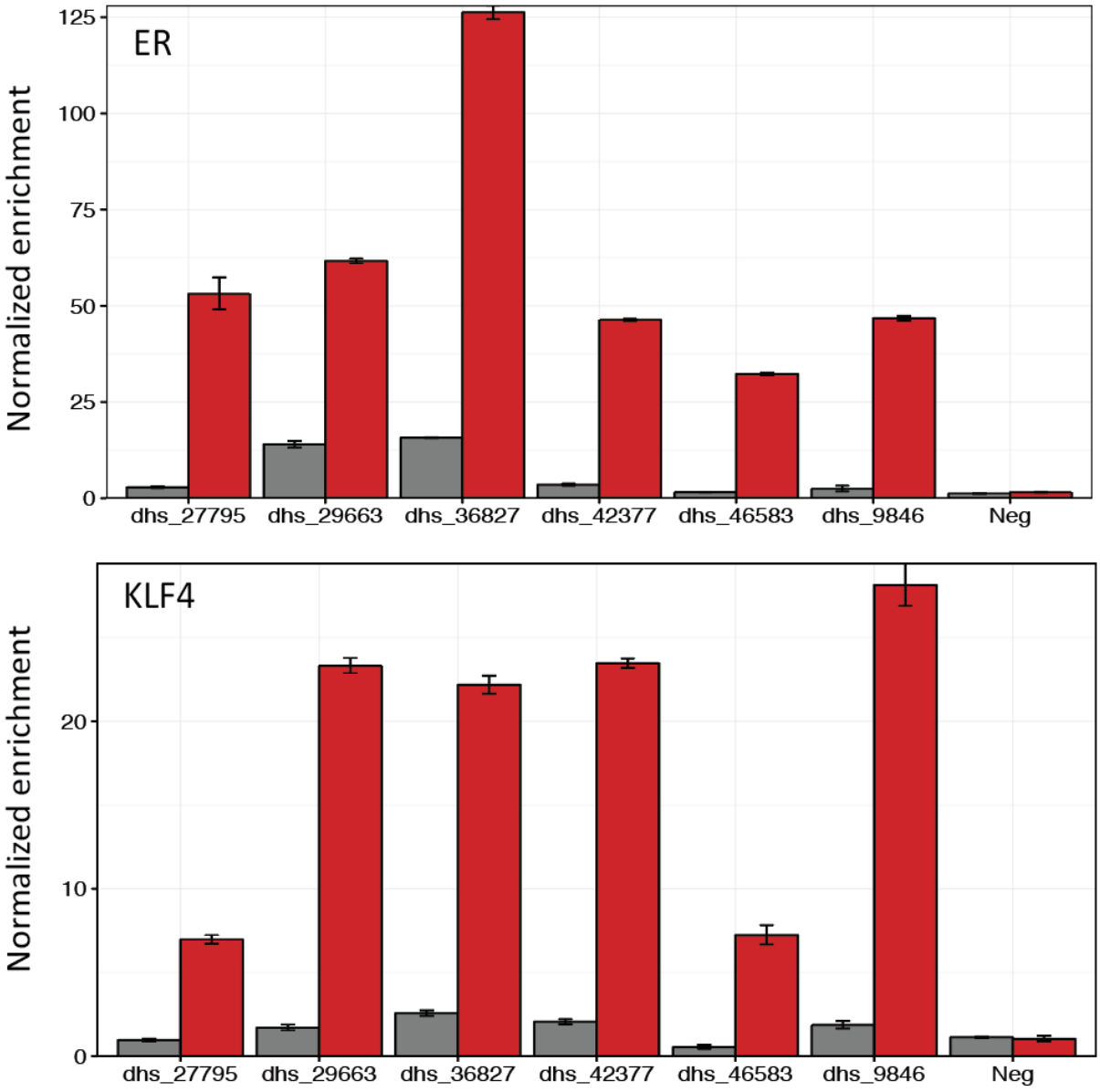

Figure 28. ChIP-qPCR on E2 induced enhancers
Red: E2 treatment for 45 min. Grey: vehicle treatment for 45 min. ChIP-qPCR data was normalized using comparative C_T method by input sample and negative control region. Error bars represent standard error from three biological replicates.

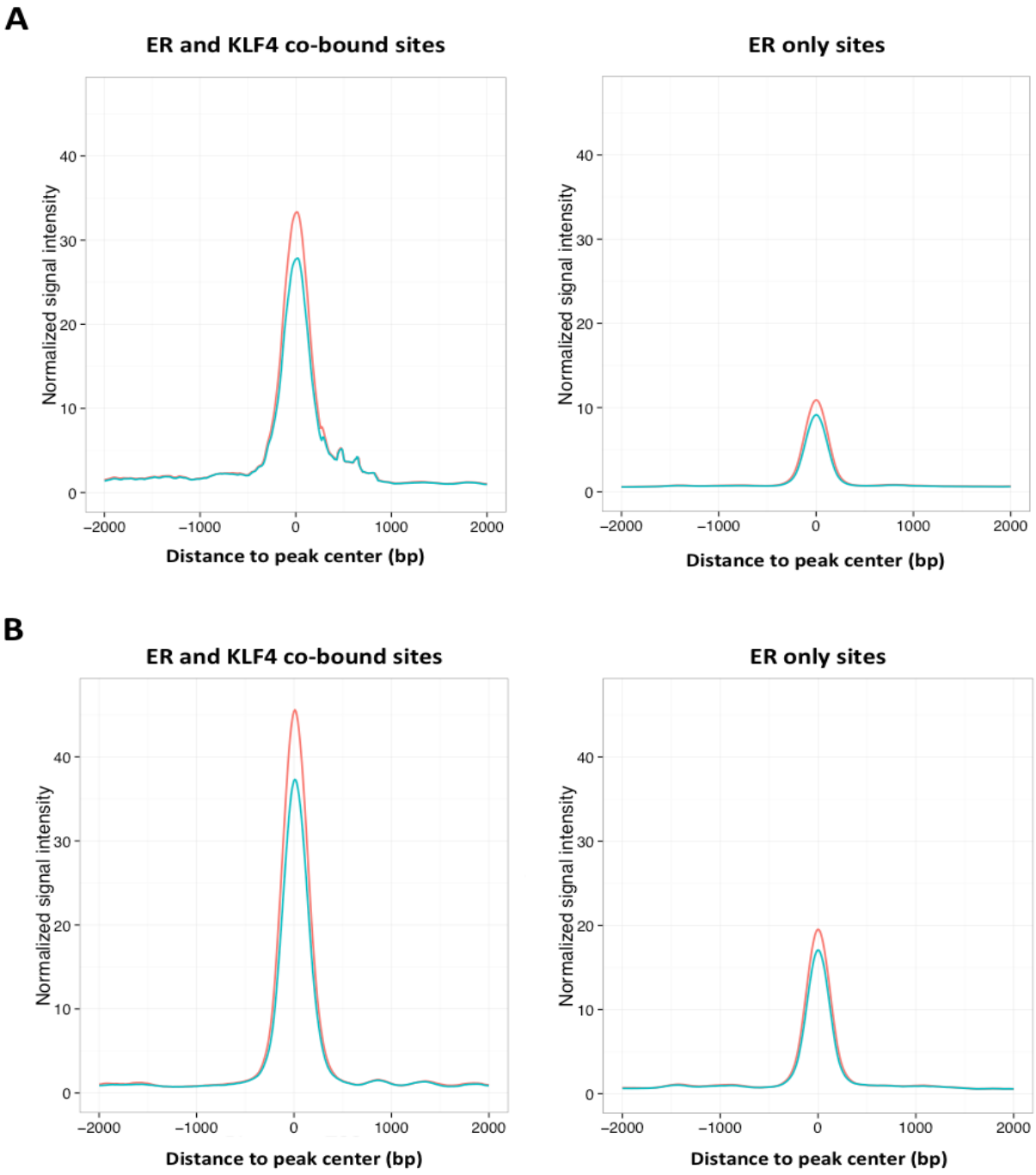

Figure 29. Meta-plot of ER ChIP-seq signal in KLF4 knockdown cells versus scrambled control cells

A) Average ER ChIP-seq signal across genome. B) Average ER ChIP-seq signal in E2 constantly induced DHSs. Red: shScramble. Blue: shKLF4. ChIP-seq tag count was normalized to 10 million reads and average tag count was calculated and plotted centered by ER binding site centers.

3.2.5 Necessity of KLF4 motif in enhancer activity of E2 induced enhancers

We have shown that KLF4 and ER are both recruited to E2 induced enhancers upon E2 treatment. Next, we investigated whether KLF4 motif is important for E2 responsive enhancer activity of E2 induced enhancers. We randomly selected four E2 induced enhancers that are co-bound by ER and KLF4, and we cloned WT and mutated enhancers (in which the KLF4 motif is deleted) into luciferase reporter constructs. We transfected them side-by-side into MCF7 cells and treated cells with E2 or vehicle for 12 hours. The WT enhancers all showed strong luciferase activity which were more than 50 fold higher than enhancer-less pGL4.23[minP/luc2] negative controls (**Figure 30A**). After deleting the KLF4 motif, the overall luciferase activity of all four enhancers was significantly reduced. E2 induced luciferase activity of two enhancers, dhs_29667 and dhs_42377, was significantly attenuated in KLF4 motif deleted mutants. The other two enhancers did not show significant change of E2 responsive enhancer activity (**Figure 30B**). Our data demonstrates that the KLF4 motif is essential for maintaining overall enhancer activity of E2-induced enhancers and is also important for E2-dependent enhancer activity of some E2-induced enhancers.

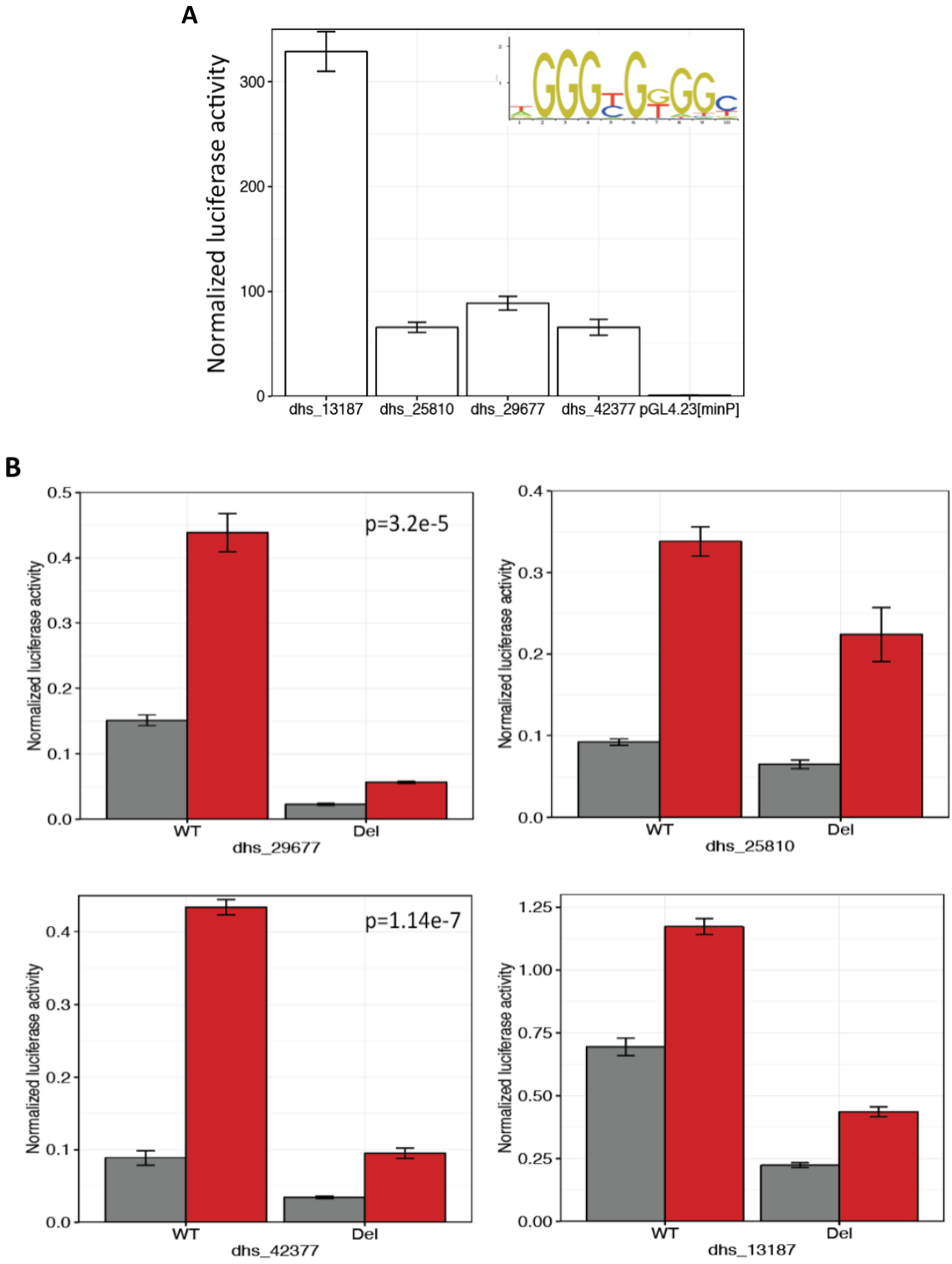

Figure 30. Luciferase assay of E2-induced enhancers with KLF4 motif deletion
A) Normalized luciferase activity of WT E2-induced enhancers compared to pGL4.23[minP/luc2] control. Firefly luciferase signal was normalized by co-transfected NanoLuc luciferase signal from

Figure 30. Luciferase assay of E2-induced enhancers with KLF4 motif deletion (continued) same population of cells. Error bars represent standard error of three biological replicates. B) Relative luciferase activity of WT versus KLF4 motif deleted enhancers with E2 or vehicle treatment for 12 hours. Red: E2. Grey: vehicle. Error bars represent standard error of three biological replicates. Pvalues indicate the significance level of interaction between genotype and treatment in linear model.

3.3 Discussion

In this chapter, we have investigated the role of KLF4 in E2-governed transcriptional regulation in ER-positive MCF7 cells. Complementing previous studies focusing on proteolysis of KLF4 protein regulated by E2 signaling (Hu et al. 2012), we have identified KLF4 as a primary transcriptional target of E2 signaling which mRNA and protein levels are rapidly accumulated upon E2 stimulation. By profiling KLF4 target genes using knockdown experiments, we revealed a substantial fraction of E2 up-regulated genes overlap with KLF4 target genes, and that the overlapped genes are extremely enriched for genes involved in cell cycle, DNA metabolism and replication. These results provide an explanation for attenuated E2-induced cell proliferation in KLF4 loss-of-function cells. Co-occurrence of ER and KLF4 binding sites associated with genes regulated by both TFs and in E2-induced DHSs suggests widespread crosstalk between the two TFs at transcriptional regulation level. We further tested the necessity of KLF4 binding motifs in E2-induced enhancers by reporter assays and demonstrated that KLF4 motif is not only essential for overall regulatory activity of E2-induced enhancers but also important for E2-responsive regulatory activity of specific E2-induced enhancers. Taken together, our data identifies KLF4 as an E2 early effector TF important for E2-dependent cell proliferation, and our results indicate that KLF4 is a cooperative TF participating in the E2-goverened gene regulation network.

Recently, more and more studies on KLF family members in NR signaling have uncovered KLFs as key effectors in NR-mediated transcriptional regulation. In general, several mechanisms have been proposed for the roles of KLFs in NR signaling (Knoedler and Denver 2014). Our results show that KLF4 is a primary and direct transcriptional target of ER that mediates transcription of subsequent E2 responsive genes with and without ER. Since KLF4 abundance is highly regulated by

a variety of stimuli including interferon-γ (Chen, Shie, and Tseng 2000), DNA damage (Yoon, Chen, and Yang 2003) and cAMP (Birsoy, Chen, and Friedman 2008), it can act as a convergence point integrating different signaling pathways to regulate endocrine responses. Secondly, KLF4 is an accessory TF that cooperatively regulates activity of cis-regulatory elements with ER. Though KLF4 is transcriptionally induced by E2 signaling, chromatin immunoprecipitation data indicates that KLF4 not only functions in secondary transcriptional regulation but also participates in primary gene regulation, since E2 induced rapid recruitment of KLF4 in cis-regulatory elements which may occur too fast to be secondary effect. Thirdly, KLF4 regulates expression of many key regulators in E2-governed transcriptional regulation such as JUN and FOS, which in turn are known to modulate E2 response of transcriptome.

In the earlier chapter, we have identified that SP/KLF TF family motifs (e.g. KLF4, KLF5, SP1, etc.) are over-represented in E2-induced DHSs and E2-induced enhancers. Though it is already recognized that ER is often tethered to Sp1 binding sites through protein-protein interaction to modulate transcription, little is known about the role of other SP/KLF family TFs in E2-goverend transcriptional regulation. KLF family members share highly conserved DNA binding domain consists of three Cys2-His2 zinc fingers which are similar to Sp family members, so their divergent function is mainly dependent on the highly variable N-terminal transactivation domains. From chromatin immunoprecipitation, we have confirmed that KLF4 frequently co-occupies in E2-induced DHSs with ER, and this association is quickly triggered by E2. Furthermore, disruption of KLF4 motifs in E2-induced enhancers effectively abrogated E2 response of several tested enhancers, indicating that ER requires KLF4 to fully function on certain cis-regulatory elements. Studies led by other groups have also proved that KLF4 physically interacts with ER in MCF7 cells by immunoprecipitation (Akaogi et al. 2009; Mohammed et al. 2013). Collectively, this evidence strongly indicates that ER and KLF4 may have direct interaction on chromatin. However, after

knocking down KLF4, ER genomic binding was only moderately affected globally. Several hypotheses can possibly explain this phenomenon: a) the knockdown efficiency of KLF4 in the study is not enough to affect ER binding; b) there is functional redundancy among SP/KLF family members in mediating ER interaction with chromatin; c) ER and KLF4 interacts with each other in multiple ways, so that despite of the physical interaction between the two TFs in the cells, they independently bind to their own binding motifs that are located proximal to each other on chromatin and coordinately regulate gene transcription. In order to test the above hypotheses and figure out the exact mechanism of crosstalk between KLF4 and ER on chromatin, further experiments will need to be performed. For example, in addition to shRNA-mediated knockdown, other loss-of-function and gain-of-function models such as complete knockout of KLF4 mediated by CRISPR/Cas9 technology and overexpression of exogenous KLF4 can be built to examine the consequence of manipulating KLF4 expression level to ER binding and gene regulation outcome.

While we have not observed significant influence on ER loading onto chromatin after knocking down KLF4, we have found KLF4 to be crucial for E2-dependent cell proliferation because of numerous cell cycle related genes regulated by both E2 and KLF4, including PLK1, CDC6, CDK2 and POLA1. KLF4 knockdown also significantly abrogated the pro-proliferative effect of E2. E2 is a strong mitogen for ER+ cells because among the hundreds to thousands of genes directly or indirectly regulated by ER, a subset of key genes are involved in cell cycle modulation. Since the genes regulated by both E2 and KLF4 are even more enriched for cell cycle regulators than genes only regulated by E2, we postulate that E2-triggered transcriptional regulation on a substantial proportion of cell cycle related genes is mediated by KLF4. On one hand, KLF4 seems to work as E2 early effector TF to "amplify" gene expression. For example, lower levels of KLF4 expression resulted in decreased overall expression of PLK1 and POLA1 but did not significantly confer E2 response of gene expression, suggesting an additive effect of E2 and KLF4.

On the other hand, for CDC45, CCNB1 and CDC25B, KLF4 knockdown not only led to reduced gene expression level but also attenuated E2-dependent gene expression, indicating E2 and KLF4 can synergistically regulate gene expression. According to prior studies, KLF4 regulates gene transcription through complex mechanisms that depend on cellular context. For activating gene transcription, it can either act alone by directly fusing its three zinc fingers with its N-terminal transactivation domain (Geiman et al. 2000), or it can interact with the co-activator p300/CBP complex (Evans et al. 2007). For achieving transcriptional repression, KLF4 can compete with other activators like Sp1 to passively repress target gene expression (Zhang et al. 1998), or it can actively repress genes through its own repressive domain by recruiting HDACs (Noti, Johnson, and Dillon 2005). It has also been reported that KLF4 is able to interact with KAT5, a bi-functional cofactor, to either activate or repress gene expression depending on cell context (Evans and Liu 2008). Therefore, in order to further dissect the additive and synergistic crosstalk on gene regulation between KLF4 and ER, we will need to associate their co-regulated genes with KLF4 and ER binding sites and perform in-depth analyses gene by gene.

Though further molecular studies are needed to clearly elucidate the mechanisms of KLF4 mediated gene transcription and the relationship between KLF4 and ER when interacting with chromatin in the context of E2 signaling in more ER positive breast cancer models, the results from our functional studies and those of other groups have identified KLF4 as a key player in breast cancer tumorigenesis and tumor progression. Many datasets based on primary samples have discovered elevated expression of KLF4 protein in breast cancer (Foster et al. 2000; Pandya et al. 2004; Hu et al. 2015), and some datasets have further suggested the positive association between KLF4 expression level and ER status (Akaogi et al. 2009; Hu et al. 2015). In agreement with these data, we have also found that KLF4 tends to be co-expressed with ER in breast cancer using large-scale TCGA data (log odds ratio = 1.484, p-value = 0.002, sample size = 960). It maybe useful to

stratify ER positive breast cancer based on KLF4 expression in evaluating clinical outcome including patient response to endocrine therapy.

3.4 Materials and methods

3.4.1 Cell culture and hormone treatment conditions

MCF7 cells (ATCC: HTB-22) and lentivirus transduced stable cell lines were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1% antibiotic-antimycotic. For hormone starvation, cells were cultured in phenol red free DMEM supplemented with 10% charcoal:dextran stripped FBS and 1% antibiotic-antimycotic. 10 nM of 17β-estradiol (Sigma) was used to stimulate cells after starvation, and ethanol was used as vehicle control.

3.4.2 Generation of KLF4 shRNA knockdown stable cell line

shRNA vector construction and lentiviral packaging were manufactured by VectorBuilder. Scrambled shRNA or shRNA targeting human KLF4 mRNA was cloned into lentiviral vector driven by U6 promoter. An eGFP:Puro fusion gene driven by PGK promoter was placed downstream of shRNA expression cassette as selective marker gene. Scrambled control and hKLF4 shRNA target sequence is listed in **Table 8**. 2.0×10^5 cells were seeded in 6-well plate the day before transduction. For transduction, lentiviral particles were mixed with Polybrene (5ug/ml) and transduced into cells (at 30~50% confluency) at multiplicity of infection (MOI) of 10. At 48 hours post-transduction, cells were cultured in hormone starvation medium supplemented with puromycin (2ug/ml) for 3 days to achieve hormone deprivation and selection for positively transduced cells prior to E2/vehicle treatment.

Oligo name	Sequence (5'->3')	Target	Citation
scrambled	CCTAAGGTTAAGTCGCCCTCG	luciferase	(Cai et al. 2006)
#1 shRNA	GGACGGCTGTGGATGGAAA	hKLF4 last exon	(Rowland, Bernards, and Peeper 2005)
#2 shRNA	TTGTGGATATCAGGGTATAAA	hKLF4 3'UTR	
#3 shRNA	GCTCCATTACCAAGAGCTCAT	hKLF4 spanning 3rd and 4th exons	(Lai et al. 2012)

Table 8. Information of shRNA targeted sequence

3.4.3 Western blot

At least 1.0×10^6 cells were harvested and lysed in 1×RIPA buffer (Pierce) supplied with cOmplete protease inhibitor (Roche). Total protein concentration was determined using BCA Protein Assay Kit (Pierce) and 40 ug protein was loaded. Western blotting was performed using anti-KLF4 (Santa Cruz: sc-20691), anti-ERα (Santa Cruz: HC-20) and anti-β-actin (Cell Signaling: 3700) as primary antibodies. Signal was detected on Odyssey imaging platform (LI-COR) using corresponding IRDye secondary antibodies (LI-COR).

3.4.4 Gene expression profiling and differential gene detection

After 72 hours of hormone starvation and puromycin selection, 1.0×10^6 shKLF4 or shScramble cells (at ~80% confluency) were treated with E2 or vehicle for 24 hours. Two biological replicates were generated under each condition. Total RNA was extracted from E2/vehicle treated cells using MagNA Pure Compact RNA Isolation Kit (Roche). RNA-seq library preparation was conducted using TruSeq Stranded mRNA Library Prep Kit (Illumina). RNA-seq libraries were sequenced on Illumina HiSeq2000/2500 platform with 50 bp single-end format. RNA-seq raw reads were mapped to human genome hg19 using TopHat. Transcript assembly (RefSeq annotation) and gene FPKM estimation were done by Cufflinks. Read counts of genes were summarized by HTSeq, and differentially expressed genes between different conditions were detected by edgeR (Robinson, McCarthy, and Smyth 2010) using exact negative binomial test (FDR<0.001 and average FPKM > 1).

3.4.5 ChIP-seq, ChIP-qPCR and peak calling

2.0×10^7 cells (at ~80% confluency) were treated with E2/vehicle for 45 min after 72 hours of hormone starvation. ChIP experiment was done as previously described (Hua et al. 2008). Chromatin was lysed in 0.1% SDS lysis buffer and sheared by sonication (Covaris S2, intensity=4, duty cycle=5%, cycles per burst=200, time=7min). Chromatin immunoprecipitatino was performed using anti-ERα (Santa Cruz: HC-20) antibody. ChIP-seq library preparation was performed as previously described (Hua et al. 2008) using adaptors and primers listed in Table 6. ChIP-seq libraries were sequenced on Illumina HiSeq2000/2500 platform in 50 bp single-end format. Raw sequencing reads were aligned to human genome hg19 by BWA using default parameters. Only uniquely mapped reads were used in downstream analysis. ChIP-seq peaks were called using MACS2 (Zhang et al. 2008) with default parameters (FDR<0.05).

For ChIP-qPCR, the fold enrichment of ChIP-ed DNA relative to input for a given testing site was determined by comparative C_T method using LightCycler 480 SYBR Green I Master Mix (Roche). The sequence of primers used in qPCR is described in **Table 9**.

Oligo name	Sequence (5'->3')	Genomic coordinates (hg19)
dhs_27795_f	GAGAACACGGGCACTTTGA	chr16:85499721-85499803
dhs_27795_r	TTACCACGACCTGGCTTTG	
dhs_29663_f	TCTGGCTGTTCCCTCTCC	chr17:38478670-38478754
dhs_29663_r	TCACGGTGACACTGCCT	
dhs_36827_f	GTGACACTTGGCGTCCTG	chr2:11638827-11638907
dhs_36827_r	AGTCATGCCCGTGCCTA	
dhs_42377_f	GAGTTGCTAGCCAACCTAGAG	chr2:238467741-238467854
dhs_42377_r	ACTGTGTGGGAGTGTTGTG	
dhs_46583_f	GTCACAGGCTGCCAAGTAT	chr22:29209858-29209953
dhs_46583_r	GTCACAGGTCTGTGTGCTC	
dhs_9846_f	TGACCTGAATCTCTGTCATCAATAA	chr10:95194970-95195056
dhs_9846_r	GCTGGCCCTCTTTCTACTTC	
Neg1_f	GTCTACTCTCGGGTCTCCTA	chr12:117864219-117864329
Neg1_r	CTGCGCTTGGCAGAATG	
Neg2_f	CACCAAGACCCAGTTGTTACT	chr19:35731477-35731578
Neg2_r	TTGCTGGAAGAGACTTCACTG	

Table 9. Primer information of ChIP-qPCR

3.4.6 EdU proliferation assay

After 3 days of hormone deprivation and puromycin selection, 3.0×10^5 cells (at ~20% confluency) were treated with E2 or vehicle. 10 uM EdU was applied to medium at 24 hour and 48 hour after hormone treatment, and incubated with cells for 1.5 hours at 37°C CO_2 incubator. Cells were collected after being trypsinized, fixed, permeabilized and stained using Click-iT Plus EdU Flow Cytometry Assay kit (ThermoFisher Scientific). Total cells and EdU positive cells were analyzed on BD LSR II flow cytometer. Three biological replicates were generated under each condition.

3.4.7 Dual luciferase reporter assay

Luciferase reporter constructs were manufactured by VectorBuilder. KLF4 motifs in E2 induced enhancers were identified by FIMO (Grant, Bailey, and Noble 2011) using default parameters. WT and motif deleted mutant enhancers were cloned into pGL4.23[luc2/minP] (Promega) multiple cloning site which is upstream of the minP promoter. On the day prior to transfection, 1.0×10^4 cells were plated into 96-well plates in starvation medium. During transfection, Firefly luciferase reporter plasmids were co-transfected with NanoLuc luciferase control plasmid pNL1.1.TK[Nluc/TK] (Promega) using ViaFect Transfection Reagent (Promega). For each well, 100 ng total DNA (Firefly plasmid : NanoLuc plasmid = 99:1) was mixed with 0.3 ul transfection reagent in 10 ul serum-free medium and incubated for 10 min at room temperature before being added into cells. 24 hours after transfection, cells were treated with E2 or vehicle. Luciferase assays were performed using Nano-Glo Dual-Luciferase Reporter Assay Kit (Promega) at desired time points post E2/vehicle treatment on Synergy HT plate reader (BioTek) following manufacturer's protocol.

3.4.8 RT-qPCR

shRNA stable knockdown cells were collected after trypsinization. Total RNA was extracted using MagNA Pure Compact RNA Isolation Kit (Roche). Reverse transcription was performed on total RNA to generate cDNA primed by oligo dT using SuperScript III Reverse Transcriptase (ThermoFisher Scientific). Real-time PCR was performed using LightCycler 480 Mater Mix (Roche) and relative expression level of KLF4 was estimated by comparative C_T method using RPLP0 house keeping gene as internal control (**Table 10**).

Oligo name	Sequence (5'->3')
KLF4_f	GGGCCCAATTACCCATCCTT
KLF4_r	GGCATGAGCTCTTGGTAATGG
RPLP0_f	ACAATGGCAGCATCTACAACC
RPLP0_r	GGTGTAATCCGTCTCCACAGA

Table 10. Primers used in RT-qPCR

3.4.9 Gene ontology enrichment analysis

We divided E2 up-regulated genes into two categories: genes up-regulated by both E2 and KLF4, and genes up-regulated only by E2. Genes with FPKM > 1 in MCF7 cells after excluding E2 up-regulated genes were chosen as background gene list. Gene ontology annotation of "Biological Process" from DAVID Bioinformatics Database (Dennis et al. 2003) was obtained for gene enrichment analysis. Fisher's exact tests were conducted to compare gene category representation between genes of interest and background gene list. A cut-off of 0.05 for Bonferroni corrected p values was used.

3.4.10 Heat map and meta-plot of ChIP-seq signal in regions of interest

To make heat map of ChIP-seq signal in E2 induced DHSs, 25 bp windows were generated in 5 kb flanking regions of the centers of E2 induced DHSs. ChIP-seq tag count in each window of each DHSs was computed and normalized by library size equivalent to 10 million reads using HOMER (Heinz et al. 2010). Heatmap data matrix was clustered using "centered correlation" method in Cluster 3.0 (de Hoon et al. 2004) and visualized in Java Tree View (Saldanha 2004). To make meta-plot, 10 bp windows were created in 2 kb flanking regions of centers of ER binding sites, and average ChIP-seq tag count in each window was computed and normalized by library size equivalent to 10 million reads using HOMER, then plotted as histogram.

3.5 Contributions

Grace Yu and Kevin White designed the project. Grace Yu generated all experimental data, performed data analyses and composed figures and tables. Grace Yu and Kevin White wrote the manuscript.

CHAPTER 4

Summary and perspectives

We have presented studies on gene transcriptional regulation of E2 signaling in ER+ breast carcinoma cell line MCF7. In Chapter 2, we have identified thousands of E2-responsive DHSs in a time-course manner within 24 hours of E2 stimulation. By integrating a series of genomic data, we uncovered chromatin signatures of E2-dependent transcriptional regulation at E2-responsive DHSs, and dynamic association between E2-responsive DHSs and E2-regulated genes. Functional assessment of regulatory activity of ~10,000 candidate E2 gene regulation related DHSs via CapSTARR-seq has revealed cis-regulatory activity of more than 3000 regions and E2-dependent cis-regulatory activity of 800 regions. We have proposed various TFs to coordinately interplay with E2-responsive DHSs and cis-regulatory elements associated and not associated with ER by motif analysis and DNase I footprint analysis. Among them, AP-1, SP/KLF and AP-2 family TFs are indicated to be important E2 early responsive TFs and play key roles in primary and secondary transcriptional regulation of E2 signaling. In Chapter 3, we have preliminarily characterized the involvement of KLF4 in E2-governed gene regulation. We provided strong evidence suggesting that KLF4 is ER primary and direct transcriptional target induced quickly by E2 stimulation. We have revealed that knocking down of KLF4 in MCF7 abrogated E2-dependent cell proliferation, since KLF4 cooperatively regulates a substantial proportion of E2-regulated genes especially genes involved in cell cycle regulation. We further validated the frequent co-occurrence of KLF4 and ER at E2-induced DHSs triggered by E2, and demonstrated the necessity of KLF4 motif in contributing to overall and E2-responsive regulatory activity of E2-induced enhancers. We propose KLF4 as a

novel cooperative TF participating in E2-goverened gene regulation network and important for E2-dependent tumor growth.

As a driving pathway in ER+ breast cancer, E2 signaling extensively regulates transcriptome and interplays with various other signaling pathways to exert cellular functions controlling tumor growth. The genomic and non-genomic actions of E2 signaling converge onto chromatin, where E2-activated ER and other TFs interact with cis-regulatory elements to collectively modulate transcription of E2 target genes. Therefore, to elucidate and establish E2-goverened gene regulation network is fundamental for developing novel therapeutic strategies for ER+ breast cancer and other estrogen-responsive diseases. Previous studies have made numerous achievements in identifying E2 targets genes and ER-dependent cistrome. However, a big barrier still exists when trying to functionally evaluate E2-responsive cis-regulatory elements and corresponding TFs. Another difficulty of decoding E2-goverened gene regulation network is to identify the hierarchy of key regulators function in different time frames of transcriptional regulation network. We have built an experimental and computational pipeline to systematically discover E2-responsive cistrome and functionally interrogate E2-responsive cis-regulatory elements genome-wide. It is the first time to massively study functionality of E2-dependent cis-regulatory elements in MCF7 cells.

The first take-home message from this study is that chromatin accessible regions with most significant E2-induced changes are enriched for HOT regions, suggesting E2-governed transcriptional regulation is a complex action involving various TFs in addition to ER. Besides the numerous candidate functional TFs predicted by motif analysis and DNase I footprint analysis, a lot of in-depth experiments are needed to verify TFs actually interplay with given cis-regulatory elements and dissect the contribution of different TFs recruited in the same cis-region. A recent study using GAL4-DNA binding domain fusions in luciferase reporter assays has indicated that the

contributions and functions of different TFs to regulatory activity of enhancers differ a lot depending on enhancer contexts, and similar TFs can substitute or compete with each other (Stampfel et al. 2015). It would be important to investigate any cooperation, redundancy, competition and antagonism among different TFs in fine-tuning E2-dependent enhancer function, because it may have clinical implications such as patients may show different sensitivity to hormone therapy due to varying ratios of the abundance of two competitive TFs at E2-responsive cis-regulatory elements leading to different gene regulation outcomes.

Another important finding from our data is that many E2-induced DHSs and CapSTARR-seq validated E2-induced cis-regulatory elements are not associated with direct ER binding. Since the majority of previous studies mainly focus on ER-dependent transcriptional regulation, little is known about E2-responsive cis-regulatory elements that do not directly interact with ER. Besides the possible scenario that some of these cis-regions may be indirectly and weakly bound by ER through tethering or looping so that they fail to be cross-linked in ChIP-based assays, we assume that many of these cis-regulatory elements only rely on other non-ER TFs to exert gene regulation function in E2 signaling. Though the potential interacting TFs may still be ER-dependent as they may be activated either by post-translational modification via ER-initiated non-genomic actions, or as primary transcriptional targets of ER, they may act as important regulators controlling expression of critical downstream genes involving in tumor-related cellular processes. More in-depth functional studies on candidate TFs should be conducted to help explain regulation of genes not directly associated with ER binding sites.

As a novel approach that is still in development, CapSTARR-seq has shown big potential in quantitative studies on regulatory elements in large scale. It has provided more real-time assessment of regulatory activity of candidate DHSs in response to E2 compared to luciferase reporter assay, which is pivotal for establishing gene regulation network. Some TF motifs enriched in E2-induced

DHSs no longer show up in E2-induced enhancers, probably either because those motifs revealed by motif analysis on DHSs are false positive hits or because those TFs exert function by interplaying with chromatin machinery. In order to fully characterize E2-dependent regulatory function of cis-regions in endogenous environment and identify their target genes, CRISRP/Cas9-mediated genome editing should be performed to precisely create loss-of-function models of particular cis-regulatory elements or perturb regulatory activity by manipulating motif composition. Since the current CapSTARR-seq analysis is limited to a subset of DHSs, which only account for ~14% of total open chromatin regions in MCF7 cells, the next step would be to expand the experiment to all open chromatin regions or whole genome. It is known that ~98% of human genome is composed of noncoding regions and ~10% of the genome is under selection, indicating that functional roles for vast noncoding regions (Rands et al. 2014). ChIP-seq data generated from different labs has all revealed that a big fraction of ER binding sites are located outside of chromatin accessible regions, which is not normally seen for other TFs. It would be interesting to test whether ER binding sites in closed chromatin regions are really functional or not using whole genome STARR-seq. Furthermore, CRISPR/Cas9-mediated genetic screens on noncoding regions associated with E2-dependent tumor phenotypes such as cell proliferation and resistance to hormone-therapy drugs is a promising approach to detect cis-regulatory elements functionally important for ER+ breast cancers.

Here, we would also like to summarize some limitations of the current study. Given the limitations of accuracy of DNase I footprint analysis and dynamic range of CapSTARR-seq, improvements in data generation and bioinformatics analyses are needed to get a higher resolution roadmap of transcriptional regulation network in E2 signaling. Within the timeframe of the experiments, we have not observed dramatic genome-wide reprogramming of chromatin accessibility landscape compared to the big differences of chromatin landscape usually observed between species, tissue types or cell types, indicating that more powerful and sensitive mathematical

models should be built to detect more DHSs responding to E2. The conventional tag count-based statistical models for detecting differential genes or differential ChIP-seq peaks like DESeq (Anders and Huber 2010) and edgeR (Robinson, McCarthy, and Smyth 2010) are not suitable for DNase-seq data, because based on the experimental mechanism of generating DNase-seq data, it is the distribution of the DNase I cleavage that matters but not the total count of cleavage within DNase I hypersensitivity sites. New approaches like WaveQTL (Shim and Stephens 2015) may be used for determining minor differences of DNase I cleavage due to changes of TF recruitment on cis-regulatory regions. An intriguing application of DNase I hypersensitivity data is to detect DNase I footprints that are indicative of TF binding at single base pair resolution. However, in spite of the several already published tools, accurately and robustly detecting DNase I footprints is a big challenge in the field. Common bias affecting the performance of footprint analysis include DNase I cleavage bias due to intrinsic DNA sequence, bias arisen from different library generation protocols and bias among different enzymes (Yardimci et al. 2014). Besides these biases, the performance of DNase I footprint analysis is highly dependent on the abundance and residence time on chromatin of interested TFs in given cell type (Sung et al. 2014). Therefore, it is generally hard to predict given TF footprints from DNase-seq data given high false positive rate and high false negative rate. Improvement of better computational tools is required for getting precise prediction of TF occupancy from DNase-seq data. For STARR-seq and CapSTARR-seq, besides the caveat that they measure regulatory activity of exogenous plasmid without chromatin context, other limitations include intensive experimental labor and narrow dynamic range are also important issues that need to be improved.

REFERENCES

Akaogi, K., Y. Nakajima, I. Ito, S. Kawasaki, S. H. Oie, A. Murayama, K. Kimura, and J. Yanagisawa. 2009. 'KLF4 suppresses estrogen-dependent breast cancer growth by inhibiting the transcriptional activity of ERalpha', Oncogene, 28: 2894-902.

Albanito, L., A. Madeo, R. Lappano, A. Vivacqua, V. Rago, A. Carpino, T. I. Oprea, E. R. Prossnitz, A. M. Musti, S. Ando, and M. Maggiolini. 2007. 'G protein-coupled receptor 30 (GPR30) mediates gene expression changes and growth response to 17beta-estradiol and selective GPR30 ligand G-1 in ovarian cancer cells', Cancer Res, 67: 1859-66.

Anders, S., and W. Huber. 2010. 'Differential expression analysis for sequence count data', Genome Biol, 11: R106.

Arnold, C. D., D. Gerlach, D. Spies, J. A. Matts, Y. A. Sytnikova, M. Pagani, N. C. Lau, and A. Stark. 2014. 'Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution', Nat Genet, 46: 685-92.

Arnold, C. D., D. Gerlach, C. Stelzer, L. M. Boryn, M. Rath, and A. Stark. 2013. 'Genome-wide quantitative enhancer activity maps identified by STARR-seq', Science, 339: 1074-7.

Arpino, G., L. Wiechmann, C. K. Osborne, and R. Schiff. 2008. 'Crosstalk between the estrogen receptor and the HER tyrosine kinase receptor family: molecular mechanism and clinical implications for endocrine therapy resistance', Endocr Rev, 29: 217-33.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. 'High-resolution profiling of histone methylations in the human genome', Cell, 129: 823-37.

Belandia, B., R. L. Orford, H. C. Hurst, and M. G. Parker. 2002. 'Targeting of SWI/SNF chromatin remodelling complexes to estrogen-responsive genes', EMBO J, 21: 4094-103.

Birsoy, K., Z. Chen, and J. Friedman. 2008. 'Transcriptional regulation of adipogenesis by KLF4', Cell Metab, 7: 339-47.

Bjornstrom, L., and M. Sjoberg. 2004. 'Estrogen receptor-dependent activation of AP-1 via non-genomic signalling', Nucl Recept, 2: 3.

———. 2005. 'Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes', Mol Endocrinol, 19: 833-42.

Boyce, B. F., L. Xing, G. Franzoso, and U. Siebenlist. 1999. 'Required and nonessential functions of nuclear factor-kappa B in bone cells', Bone, 25: 137-9.

Butt, A. J., C. M. McNeil, E. A. Musgrove, and R. L. Sutherland. 2005. 'Downstream targets of growth factor and oestrogen signalling and endocrine resistance: the potential roles of c-Myc, cyclin D1 and cyclin E', Endocr Relat Cancer, 12 Suppl 1: S47-59.

Cai, D., V. M. Latham, Jr., X. Zhang, and G. I. Shapiro. 2006. 'Combined depletion of cell cycle and transcriptional cyclin-dependent kinase activities induces apoptosis in cancer cells', Cancer Res, 66: 9270-80.

Carmeci, C., D. A. Thompson, H. Z. Ring, U. Francke, and R. J. Weigel. 1997. 'Identification of a gene (GPR30) with homology to the G-protein-coupled receptor superfamily associated with estrogen receptor expression in breast cancer', Genomics, 45: 607-17.

Carroll, J. S., X. S. Liu, A. S. Brodsky, W. Li, C. A. Meyer, A. J. Szary, J. Eeckhoute, W. Shao, E. V. Hestermann, T. R. Geistlinger, E. A. Fox, P. A. Silver, and M. Brown. 2005. 'Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1', Cell, 122: 33-43.

Castet, A., A. Boulahtouf, G. Versini, S. Bonnet, P. Augereau, F. Vignon, S. Khochbin, S. Jalaguier, and V. Cavailles. 2004. 'Multiple domains of the Receptor-Interacting Protein 140 contribute to transcription inhibition', Nucleic Acids Res, 32: 1957-66.

Castoria, G., A. Migliaccio, A. Bilancio, M. Di Domenico, A. de Falco, M. Lombardi, R. Fiorentino, L. Varricchio, M. V. Barone, and F. Auricchio. 2001. 'PI3-kinase in concert with Src promotes the S-phase entry of oestradiol-stimulated MCF-7 cells', EMBO J, 20: 6050-9.

Castro-Rivera, E., I. Samudio, and S. Safe. 2001. 'Estrogen regulation of cyclin D1 gene expression in ZR-75 breast cancer cells involves multiple enhancer elements', J Biol Chem, 276: 30853-61.

Catoe, H. W., and Z. Nawaz. 2011. 'E6-AP facilitates efficient transcription at estrogen responsive promoters through recruitment of chromatin modifiers', Steroids, 76: 897-902.

Chakraborty, A. K., A. Welsh, and M. P. Digiovanna. 2010. 'Co-targeting the insulin-like growth factor I receptor enhances growth-inhibitory and pro-apoptotic effects of anti-estrogens in human breast cancer cell lines', Breast Cancer Res Treat, 120: 327-35.

Chen, D., S. M. Huang, and M. R. Stallcup. 2000. 'Synergistic, p160 coactivator-dependent enhancement of estrogen receptor function by CARM1 and p300', J Biol Chem, 275: 40810-6.

Chen, J. D., and R. M. Evans. 1995. 'A transcriptional co-repressor that interacts with nuclear hormone receptors', Nature, 377: 454-7.

Chen, S., N. E. Sanjana, K. Zheng, O. Shalem, K. Lee, X. Shi, D. A. Scott, J. Song, J. Q. Pan, R. Weissleder, H. Lee, F. Zhang, and P. A. Sharp. 2015. 'Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis', Cell, 160: 1246-60.

Chen, X., D. C. Johns, D. E. Geiman, E. Marban, D. T. Dang, G. Hamlin, R. Sun, and V. W. Yang. 2001. 'Kruppel-like factor 4 (gut-enriched Kruppel-like factor) inhibits cell proliferation by blocking G1/S progression of the cell cycle', J Biol Chem, 276: 30423-8.

Chen, Z. Y., J. Shie, and C. Tseng. 2000. 'Up-regulation of gut-enriched kruppel-like factor by interferon-gamma in human colon carcinoma cells', FEBS Lett, 477: 67-72.

Chu, I. M., L. Hengst, and J. M. Slingerland. 2008. 'The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy', Nat Rev Cancer, 8: 253-67.

Cirillo, L. A., C. E. McPherson, P. Bossard, K. Stevens, S. Cherian, E. Y. Shim, K. L. Clark, S. K. Burley, and K. S. Zaret. 1998. 'Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome', EMBO J, 17: 244-54.

Clarke, R., J. J. Tyson, and J. M. Dixon. 2015. 'Endocrine resistance in breast cancer - An overview and update', Mol Cell Endocrinol, 418 Pt 3: 220-34.

Consortium, Encode Project. 2012. 'An integrated encyclopedia of DNA elements in the human genome', Nature, 489: 57-74.

Cui, K., C. Zang, T. Y. Roh, D. E. Schones, R. W. Childs, W. Peng, and K. Zhao. 2009. 'Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation', Cell Stem Cell, 4: 80-93.

Dahlman-Wright, K., and I. J. McEwan. 1996. 'Structural studies of mutant glucocorticoid receptor transactivation domains establish a link between transactivation activity in vivo and alpha-helix-forming potential in vitro', Biochemistry, 35: 1323-7.

Dang, D. T., C. S. Mahatan, L. H. Dang, I. A. Agboola, and V. W. Yang. 2001. 'Expression of the gut-enriched Kruppel-like factor (Kruppel-like factor 4) gene in the human colon cancer cell line RKO is dependent on CDX2', Oncogene, 20: 4884-90.

de Hoon, M. J., S. Imoto, J. Nolan, and S. Miyano. 2004. 'Open source clustering software', Bioinformatics, 20: 1453-4.

Delage-Mourroux, R., P. G. Martini, I. Choi, D. M. Kraichely, J. Hoeksema, and B. S. Katzenellenbogen. 2000. 'Analysis of estrogen receptor interaction with a repressor of estrogen receptor activity (REA) and the regulation of estrogen receptor transcriptional activity by REA', J Biol Chem, 275: 35848-56.

Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki. 2003. 'DAVID: Database for Annotation, Visualization, and Integrated Discovery', Genome Biol, 4: P3.

Dickel, D. E., Y. Zhu, A. S. Nord, J. N. Wylie, J. A. Akiyama, V. Afzal, I. Plajzer-Frick, A. Kirkpatrick, B. Gottgens, B. G. Bruneau, A. Visel, and L. A. Pennacchio. 2014. 'Function-based identification of mammalian enhancers using site-specific integration', Nat Methods, 11: 566-71.

Dowsett, M., J. Cuzick, J. Ingle, A. Coates, J. Forbes, J. Bliss, M. Buyse, M. Baum, A. Buzdar, M. Colleoni, C. Coombes, C. Snowdon, M. Gnant, R. Jakesz, M. Kaufmann, F. Boccardo, J. Godwin, C. Davies, and R. Peto. 2010. 'Meta-analysis of breast cancer outcomes in adjuvant trials of aromatase inhibitors versus tamoxifen', J Clin Oncol, 28: 509-18.

Driggers, P. H., and J. H. Segars. 2002. 'Estrogen action and cytoplasmic signaling pathways. Part II: the role of growth factors and phosphorylation in estrogen signaling', Trends Endocrinol Metab, 13: 422-7.

Duan, R., W. Porter, and S. Safe. 1998. 'Estrogen-induced c-fos protooncogene expression in MCF-7 human breast cancer cells: role of estrogen receptor Sp1 complex formation', Endocrinology, 139: 1981-90.

Dutertre, M., and C. L. Smith. 2000. 'Molecular mechanisms of selective estrogen receptor modulator (SERM) action', J Pharmacol Exp Ther, 295: 431-7.

Early Breast Cancer Trialists' Collaborative, Group. 1998. 'Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group', Lancet, 351: 1451-67.

———. 2005. 'Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials', Lancet, 365: 1687-717.

Easton, D. F., K. A. Pooley, A. M. Dunning, P. D. Pharoah, D. Thompson, D. G. Ballinger, J. P. Struewing, J. Morrison, H. Field, R. Luben, N. Wareham, S. Ahmed, C. S. Healey, R. Bowman, Search collaborators, K. B. Meyer, C. A. Haiman, L. K. Kolonel, B. E. Henderson, L. Le Marchand, P. Brennan, S. Sangrajrang, V. Gaborieau, F. Odefrey, C. Y. Shen, P. E. Wu, H. C. Wang, D. Eccles, D. G. Evans, J. Peto, O. Fletcher, N. Johnson, S. Seal, M. R. Stratton, N. Rahman, G. Chenevix-Trench, S. E. Bojesen, B. G. Nordestgaard, C. K. Axelsson, M. Garcia-Closas, L. Brinton, S. Chanock, J. Lissowska, B. Peplonska, H. Nevanlinna, R. Fagerholm, H. Eerola, D. Kang, K. Y. Yoo, D. Y. Noh, S. H. Ahn, D. J. Hunter, S. E. Hankinson, D. G. Cox, P. Hall, S. Wedren, J. Liu, Y. L. Low, N. Bogdanova, P. Schurmann, T. Dork, R. A. Tollenaar, C. E. Jacobi, P. Devilee, J. G. Klijn, A. J. Sigurdson, M. M. Doody, B. H. Alexander, J. Zhang, A. Cox, I. W. Brock, G. MacPherson, M. W. Reed, F. J. Couch, E. L. Goode, J. E. Olson, H. Meijers-Heijboer, A. van den Ouweland, A. Uitterlinden, F. Rivadeneira, R. L. Milne, G. Ribas, A. Gonzalez-Neira, J. Benitez, J. L. Hopper, M. McCredie, M. Southey, G. G. Giles, C. Schroen, C. Justenhoven, H. Brauch, U. Hamann, Y. D. Ko, A. B. Spurdle, J. Beesley, X. Chen, kConFab, Aocs Management Group, A. Mannermaa, V. M. Kosma, V. Kataja, J. Hartikainen, N. E. Day, D. R. Cox, and B. A. Ponder. 2007. 'Genome-wide association study identifies novel breast cancer susceptibility loci', Nature, 447: 1087-93.

el-Deiry, W. S., T. Tokino, V. E. Velculescu, D. B. Levy, R. Parsons, J. M. Trent, D. Lin, W. E. Mercer, K. W. Kinzler, and B. Vogelstein. 1993. 'WAF1, a potential mediator of p53 tumor suppression', Cell, 75: 817-25.

Encarnacion, C. A., D. R. Ciocca, W. L. McGuire, G. M. Clark, S. A. Fuqua, and C. K. Osborne. 1993. 'Measurement of steroid hormone receptors in breast cancer patients on tamoxifen', Breast Cancer Res Treat, 26: 237-46.

Eroles, P., A. Bosch, J. A. Perez-Fidalgo, and A. Lluch. 2012. 'Molecular biology in breast cancer: intrinsic subtypes and signaling pathways', Cancer Treat Rev, 38: 698-707.

Evans, P. M., and C. Liu. 2008. 'Roles of Krupel-like factor 4 in normal homeostasis, cancer and stem cells', Acta Biochim Biophys Sin (Shanghai), 40: 554-64.

Evans, P. M., W. Zhang, X. Chen, J. Yang, K. K. Bhakat, and C. Liu. 2007. 'Kruppel-like factor 4 is acetylated by p300 and regulates gene transcription via modulation of histone acetylation', J Biol Chem, 282: 33994-4002.

Fan, S., Y. X. Ma, C. Wang, R. Q. Yuan, Q. Meng, J. A. Wang, M. Erdos, I. D. Goldberg, P. Webb, P. J. Kushner, R. G. Pestell, and E. M. Rosen. 2001. 'Role of direct interaction in BRCA1 inhibition of estrogen receptor activity', Oncogene, 20: 77-87.

———. 2002. 'p300 Modulates the BRCA1 inhibition of estrogen receptor activity', Cancer Res, 62: 141-51.

Farmer, H., N. McCabe, C. J. Lord, A. N. Tutt, D. A. Johnson, T. B. Richardson, M. Santarosa, K. J. Dillon, I. Hickson, C. Knights, N. M. Martin, S. P. Jackson, G. C. Smith, and A. Ashworth. 2005. 'Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy', Nature, 434: 917-21.

Filardo, E. J., C. T. Graeber, J. A. Quinn, M. B. Resnick, D. Giri, R. A. DeLellis, M. M. Steinhoff, and E. Sabo. 2006. 'Distribution of GPR30, a seven membrane-spanning estrogen receptor, in primary breast cancer and its association with clinicopathologic determinants of tumor progression', Clin Cancer Res, 12: 6359-66.

Filardo, E. J., J. A. Quinn, K. I. Bland, and A. R. Frackelton, Jr. 2000. 'Estrogen-induced activation of Erk-1 and Erk-2 requires the G protein-coupled receptor homolog, GPR30, and occurs via trans-activation of the epidermal growth factor receptor through release of HB-EGF', Mol Endocrinol, 14: 1649-60.

Filardo, E. J., J. A. Quinn, A. R. Frackelton, Jr., and K. I. Bland. 2002. 'Estrogen action via the G protein-coupled receptor, GPR30: stimulation of adenylyl cyclase and cAMP-mediated attenuation of the epidermal growth factor receptor-to-MAPK signaling axis', Mol Endocrinol, 16: 70-84.

Flouriot, G., H. Brand, S. Denger, R. Metivier, M. Kos, G. Reid, V. Sonntag-Buck, and F. Gannon. 2000. 'Identification of a new isoform of the human estrogen receptor-alpha (hER-alpha) that is encoded by distinct transcripts and that is able to repress hER-alpha activation function 1', EMBO J, 19: 4688-700.

Fong, P. C., D. S. Boss, T. A. Yap, A. Tutt, P. Wu, M. Mergui-Roelvink, P. Mortimer, H. Swaisland, A. Lau, M. J. O'Connor, A. Ashworth, J. Carmichael, S. B. Kaye, J. H. Schellens, and J. S. de Bono. 2009. 'Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers', N Engl J Med, 361: 123-34.

Foster, K. W., A. R. Frost, P. McKie-Bell, C. Y. Lin, J. A. Engler, W. E. Grizzle, and J. M. Ruppert. 2000. 'Increase of GKLF messenger RNA and protein expression during progression of breast cancer', Cancer Res, 60: 6488-95.

Foster, K. W., S. Ren, I. D. Louro, S. M. Lobo-Ruppert, P. McKie-Bell, W. Grizzle, M. R. Hayes, T. R. Broker, L. T. Chow, and J. M. Ruppert. 1999. 'Oncogene expression cloning by retroviral transduction of adenovirus E1A-immortalized rat kidney RK3E cells: transformation of a host with epithelial features by c-MYC and the zinc finger protein GKLF', Cell Growth Differ, 10: 423-34.

Foulkes, W. D., I. E. Smith, and J. S. Reis-Filho. 2010. 'Triple-negative breast cancer', N Engl J Med, 363: 1938-48.

Frasor, J., A. Weaver, M. Pradhan, Y. Dai, L. D. Miller, C. Y. Lin, and A. Stanculescu. 2009. 'Positive cross-talk between estrogen receptor and NF-kappaB in breast cancer', Cancer Res, 69: 8918-25.

Freedman, M. L., A. N. Monteiro, S. A. Gayther, G. A. Coetzee, A. Risch, C. Plass, G. Casey, M. De Biasi, C. Carlson, D. Duggan, M. James, P. Liu, J. W. Tichelaar, H. G. Vikis, M. You, and I. G. Mills. 2011. 'Principles for the post-GWAS functional characterization of cancer risk loci', Nat Genet, 43: 513-8.

Fullwood, M. J., M. H. Liu, Y. F. Pan, J. Liu, H. Xu, Y. B. Mohamed, Y. L. Orlov, S. Velkov, A. Ho, P. H. Mei, E. G. Chew, P. Y. Huang, W. J. Welboren, Y. Han, H. S. Ooi, P. N. Ariyaratne, V. B. Vega, Y. Luo, P. Y. Tan, P. Y. Choy, K. D. Wansa, B. Zhao, K. S. Lim, S. C. Leow, J. S. Yow, R. Joseph, H. Li, K. V. Desai, J. S. Thomsen, Y. K. Lee, R. K. Karuturi, T. Herve, G. Bourque, H. G. Stunnenberg, X. Ruan, V. Cacheux-Rataboul, W. K. Sung, E. T. Liu, C. L. Wei, E. Cheung, and Y. Ruan. 2009. 'An oestrogen-receptor-alpha-bound human chromatin interactome', Nature, 462: 58-64.

Gao, X., S. K. Mohsin, Z. Gatalica, G. Fu, P. Sharma, and Z. Nawaz. 2005. 'Decreased expression of e6-associated protein in breast and prostate carcinomas', Endocrinology, 146: 1707-12.

Gardiner-Garden, M., and M. Frommer. 1987. 'CpG islands in vertebrate genomes', J Mol Biol, 196: 261-82.

Geiman, D. E., H. Ton-That, J. M. Johnson, and V. W. Yang. 2000. 'Transactivation and growth suppression by the gut-enriched Kruppel-like factor (Kruppel-like factor 4) are dependent on acidic amino acid residues and protein-protein interaction', Nucleic Acids Res, 28: 1106-13.

Generali, D., A. Berruti, M. P. Brizzi, L. Campo, S. Bonardi, S. Wigfield, A. Bersiga, G. Allevi, M. Milani, S. Aguggini, V. Gandolfi, L. Dogliotti, A. Bottini, A. L. Harris, and S. B. Fox. 2006. 'Hypoxia-inducible factor-1alpha expression predicts a poor response to primary chemoendocrine therapy and disease-free survival in primary human breast cancer', Clin Cancer Res, 12: 4562-8.

Gisselbrecht, S. S., L. A. Barrera, M. Porsch, A. Aboukhalil, P. W. Estep, 3rd, A. Vedenko, A. Palagi, Y. Kim, X. Zhu, B. W. Busser, C. E. Gamble, A. Iagovitina, A. Singhania, A. M. Michelson, and M. L. Bulyk. 2013. 'Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos', Nat Methods, 10: 774-80.

Gold, B., T. Kirchhoff, S. Stefanov, J. Lautenberger, A. Viale, J. Garber, E. Friedman, S. Narod, A. B. Olshen, P. Gregersen, K. Kosarin, A. Olsh, J. Bergeron, N. A. Ellis, R. J. Klein, A. G. Clark, L. Norton, M. Dean, J. Boyd, and K. Offit. 2008. 'Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33', Proc Natl Acad Sci U S A, 105: 4340-5.

Goss, P. E., J. N. Ingle, S. Martino, N. J. Robert, H. B. Muss, M. J. Piccart, M. Castiglione, D. Tu, L. E. Shepherd, K. I. Pritchard, R. B. Livingston, N. E. Davidson, L. Norton, E. A. Perez, J. S. Abrams, P. Therasse, M. J. Palmer, and J. L. Pater. 2003. 'A randomized trial of letrozole in

postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer', N Engl J Med, 349: 1793-802.

Grant, C. E., T. L. Bailey, and W. S. Noble. 2011. 'FIMO: scanning for occurrences of a given motif', Bioinformatics, 27: 1017-8.

Gronemeyer, H., J. A. Gustafsson, and V. Laudet. 2004. 'Principles for modulation of the nuclear receptor superfamily', Nat Rev Drug Discov, 3: 950-64.

Gruber, C. J., W. Tschugguel, C. Schneeberger, and J. C. Huber. 2002. 'Production and actions of estrogens', N Engl J Med, 346: 340-52.

Guarneri, V., and P. Conte. 2009. 'Metastatic breast cancer: therapeutic options according to molecular subtypes and prior adjuvant therapy', Oncologist, 14: 645-56.

Guertin, M. J., X. Zhang, S. A. Coonrod, and G. L. Hager. 2014. 'Transient estrogen receptor binding and p300 redistribution support a squelching mechanism for estradiol-repressed genes', Mol Endocrinol, 28: 1522-33.

Hah, N., C. G. Danko, L. Core, J. J. Waterfall, A. Siepel, J. T. Lis, and W. L. Kraus. 2011. 'A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells', Cell, 145: 622-34.

Hah, N., and W. L. Kraus. 2014. 'Hormone-regulated transcriptomes: lessons learned from estrogen signaling pathways in breast cancer cells', Mol Cell Endocrinol, 382: 652-64.

Hah, N., S. Murakami, A. Nagari, C. G. Danko, and W. L. Kraus. 2013. 'Enhancer transcripts mark active estrogen receptor binding sites', Genome Res, 23: 1210-23.

Han, M. R., J. Long, J. Y. Choi, S. K. Low, S. S. Kweon, Y. Zheng, Q. Cai, J. Shi, X. Guo, K. Matsuo, M. Iwasaki, C. Y. Shen, M. K. Kim, W. Wen, B. Li, A. Takahashi, M. H. Shin, Y. B. Xiang, H. Ito, Y. Kasuga, D. Y. Noh, K. Matsuda, M. H. Park, Y. T. Gao, H. Iwata, S. Tsugane, S. K. Park, M. Kubo, X. O. Shu, D. Kang, and W. Zheng. 2016. 'Genome-wide association study in East Asians identifies two novel breast cancer susceptibility loci', Hum Mol Genet.

Harris, M. B., J. Mostecki, and P. B. Rothman. 2005. 'Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function', J Biol Chem, 280: 13114-21.

Hayashi, S., and Y. Yamaguchi. 2008. 'Estrogen signaling pathway and hormonal therapy', Breast Cancer, 15: 256-61.

He, B., Q. Feng, A. Mukherjee, D. M. Lonard, F. J. DeMayo, B. S. Katzenellenbogen, J. P. Lydon, and B. W. O'Malley. 2008. 'A repressive role for prohibitin in estrogen signaling', Mol Endocrinol, 22: 344-60.

He, H. H., C. A. Meyer, H. Shin, S. T. Bailey, G. Wei, Q. Wang, Y. Zhang, K. Xu, M. Ni, M. Lupien, P. Mieczkowski, J. D. Lieb, K. Zhao, M. Brown, and X. S. Liu. 2010. 'Nucleosome dynamics define transcriptional enhancers', Nat Genet, 42: 343-7.

Heery, D. M., E. Kalkhoven, S. Hoare, and M. G. Parker. 1997. 'A signature motif in transcriptional co-activators mediates binding to nuclear receptors', Nature, 387: 733-6.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren. 2009. 'Histone modifications at human enhancers reflect global cell-type-specific gene expression', Nature, 459: 108-12.

Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren. 2007. 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', Nat Genet, 39: 311-8.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. 2010. 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', Mol Cell, 38: 576-89.

Heldring, N., A. Pike, S. Andersson, J. Matthews, G. Cheng, J. Hartman, M. Tujague, A. Strom, E. Treuter, M. Warner, and J. A. Gustafsson. 2007. 'Estrogen receptors: how do they signal and what are their targets', Physiol Rev, 87: 905-31.

Helleman, J., M. P. Jansen, K. Ruigrok-Ritstier, I. L. van Staveren, M. P. Look, M. E. Meijer-van Gelder, A. M. Sieuwerts, J. G. Klijn, S. Sleijfer, J. A. Foekens, and E. M. Berns. 2008. 'Association of an extracellular matrix gene cluster with breast cancer prognosis and endocrine therapy response', Clin Cancer Res, 14: 5555-64.

Holmes, K. A., A. Hurtado, G. D. Brown, R. Launchbury, C. S. Ross-Innes, J. Hadfield, D. T. Odom, and J. S. Carroll. 2012. 'Transducin-like enhancer protein 1 mediates estrogen receptor binding and transcriptional activity in breast cancer cells', Proc Natl Acad Sci U S A, 109: 2748-53.

Horlein, A. J., A. M. Naar, T. Heinzel, J. Torchia, B. Gloss, R. Kurokawa, A. Ryan, Y. Kamei, M. Soderstrom, C. K. Glass, and et al. 1995. 'Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor', Nature, 377: 397-404.

Howell, A., J. F. Robertson, J. Quaresma Albano, A. Aschermannova, L. Mauriac, U. R. Kleeberg, I. Vergote, B. Erikstein, A. Webster, and C. Morris. 2002. 'Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment', J Clin Oncol, 20: 3396-403.

Hu, D., M. Gur, Z. Zhou, A. Gamper, M. C. Hung, N. Fujita, L. Lan, I. Bahar, and Y. Wan. 2015. 'Interplay between arginine methylation and ubiquitylation regulates KLF4-mediated genome stability and carcinogenesis', Nat Commun, 6: 8419.

Hu, D., Z. Zhou, N. E. Davidson, Y. Huang, and Y. Wan. 2012. 'Novel insight into KLF4 proteolytic regulation in estrogen receptor signaling and breast carcinogenesis', J Biol Chem, 287: 13584-97.

Hua, S., C. B. Kallen, R. Dhar, M. T. Baquero, C. E. Mason, B. A. Russell, P. K. Shah, J. Liu, A. Khramtsov, M. S. Tretiakova, T. N. Krausz, O. I. Olopade, D. L. Rimm, and K. P. White. 2008. 'Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression', Mol Syst Biol, 4: 188.

Hua, S., R. Kittler, and K. P. White. 2009. 'Genomic antagonism between retinoic acid and estrogen signaling in breast cancer', Cell, 137: 1259-71.

Hurtado, A., K. A. Holmes, C. S. Ross-Innes, D. Schmidt, and J. S. Carroll. 2011. 'FOXA1 is a key determinant of estrogen receptor function and endocrine response', Nat Genet, 43: 27-33.

Iguchi-Ariga, S. M., and W. Schaffner. 1989. 'CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation', Genes Dev, 3: 612-9.

Ingraham, H. A., and M. R. Redinbo. 2005. 'Orphan nuclear receptors adopted by crystallography', Curr Opin Struct Biol, 15: 708-15.

Isakoff, S. J., B. Overmoyer, N. M. Tung, R. S. Gelman, V. L. Giranda, K. M. Bernhard, K. R. Habin, L. W. Ellisen, E. P. Winer, and P. E. Goss. 2010. 'A phase II trial of the PARP inhibitor veliparib (ABT888) and temozolomide for metastatic breast cancer.', Journal of Clinical Oncology, 28.

Iwase, H., Y. Omoto, H. Iwata, Y. Hara, Y. Ando, and S. Kobayashi. 1998. 'Genetic and epigenetic alterations of the estrogen receptor gene and hormone independence in human breast cancer', Oncology, 55 Suppl 1: 11-6.

Iwayanagi, Y., T. Takada, and H. Suzuki. 2008. 'HNF4alpha is a crucial modulator of the cholesterol-dependent regulation of NPC1L1', Pharm Res, 25: 1134-41.

Jagannathan, V., and M. Robinson-Rechavi. 2011. 'The challenge of modeling nuclear receptor regulatory networks in mammalian cells', Mol Cell Endocrinol, 334: 91-7.

Jeziorska, D. M., K. W. Jordan, and K. W. Vance. 2009. 'A systems biology approach to understanding cis-regulatory module function', Semin Cell Dev Biol, 20: 856-62.

Jin, C., C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao, and G. Felsenfeld. 2009. 'H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions', Nat Genet, 41: 941-5.

Joseph, R., Y. L. Orlov, M. Huss, W. Sun, S. L. Kong, L. Ukil, Y. F. Pan, G. Li, M. Lim, J. S. Thomsen, Y. Ruan, N. D. Clarke, S. Prabhakar, E. Cheung, and E. T. Liu. 2010. 'Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha', Mol Syst Biol, 6: 456.

Jump, D. B. 2004. 'Fatty acid regulation of gene transcription', Crit Rev Clin Lab Sci, 41: 41-78.

Juven-Gershon, T., and J. T. Kadonaga. 2010. 'Regulation of gene expression via the core promoter and the basal transcriptional machinery', Dev Biol, 339: 225-9.

Kaczynski, J., T. Cook, and R. Urrutia. 2003. 'Sp1- and Kruppel-like transcription factors', Genome Biol, 4: 206.

Kadmiel, M., and J. A. Cidlowski. 2013. 'Glucocorticoid receptor signaling in health and disease', Trends Pharmacol Sci, 34: 518-30.

Kalaany, N. Y., and D. J. Mangelsdorf. 2006. 'LXRS and FXR: the yin and yang of cholesterol and fat metabolism', Annu Rev Physiol, 68: 159-91.

Kalaitzidis, D., and T. D. Gilmore. 2005. 'Transcription factor cross-talk: the estrogen receptor and NF-kappaB', Trends Endocrinol Metab, 16: 46-52.

Karin, M. 1995. 'The regulation of AP-1 activity by mitogen-activated protein kinases', J Biol Chem, 270: 16483-6.

Kaufman, B., R. Shapira-Frommer, R. K. Schmutzler, M. W. Audeh, M. Friedlander, J. Balmana, G. Mitchell, G. Fried, S. M. Stemmer, A. Hubert, O. Rosengarten, M. Steiner, N. Loman, K. Bowen, A. Fielding, and S. M. Domchek. 2015. 'Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation', J Clin Oncol, 33: 244-50.

Kennecke, H., R. Yerushalmi, R. Woods, M. C. Cheang, D. Voduc, C. H. Speers, T. O. Nielsen, and K. Gelmon. 2010. 'Metastatic behavior of breast cancer subtypes', J Clin Oncol, 28: 3271-7.

Kheradpour, P., J. Ernst, A. Melnikov, P. Rogov, L. Wang, X. Zhang, J. Alston, T. S. Mikkelsen, and M. Kellis. 2013. 'Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay', Genome Res, 23: 800-11.

Kininis, M., and W. L. Kraus. 2008. 'A global view of transcriptional regulation by nuclear receptors: gene expression, factor localization, and DNA sequence analysis', Nucl Recept Signal, 6: e005.

Kittler, R., J. Zhou, S. Hua, L. Ma, Y. Liu, E. Pendleton, C. Cheng, M. Gerstein, and K. P. White. 2013. 'A comprehensive nuclear receptor network for breast cancer cells', Cell Rep, 3: 538-51.

Klinge, C. M. 2000. 'Estrogen receptor interaction with co-activators and co-repressors', Steroids, 65: 227-51.

Klinge, C. M., K. A. Riggs, N. S. Wickramasinghe, C. G. Emberts, D. B. McConda, P. N. Barry, and J. E. Magnusen. 2010. 'Estrogen receptor alpha 46 is reduced in tamoxifen resistant breast cancer cells and re-expression inhibits cell proliferation and estrogen receptor alpha 66-regulated target gene transcription', Mol Cell Endocrinol, 323: 268-76.

Knoedler, J. R., and R. J. Denver. 2014. 'Kruppel-like factors are effectors of nuclear receptor signaling', Gen Comp Endocrinol, 203: 49-59.

Kong, S. L., G. Li, S. L. Loh, W. K. Sung, and E. T. Liu. 2011. 'Cellular reprogramming by the conjoint action of ERalpha, FOXA1, and GATA3 to a ligand-inducible growth state', Mol Syst Biol, 7: 526.

Kumar, V., and P. Chambon. 1988. 'The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer', Cell, 55: 145-56.

Kurokawa, R., M. Soderstrom, A. Horlein, S. Halachmi, M. Brown, M. G. Rosenfeld, and C. K. Glass. 1995. 'Polarity-specific activities of retinoic acid receptors determined by a co-repressor', Nature, 377: 451-4.

Kushner, P. J., D. A. Agard, G. L. Greene, T. S. Scanlan, A. K. Shiau, R. M. Uht, and P. Webb. 2000. 'Estrogen receptor pathways to AP-1', J Steroid Biochem Mol Biol, 74: 311-7.

Kwasnieski, J. C., C. Fiore, H. G. Chaudhari, and B. A. Cohen. 2014. 'High-throughput functional testing of ENCODE segmentation predictions', Genome Res, 24: 1595-602.

Lai, J. K., H. C. Wu, Y. C. Shen, H. Y. Hsieh, S. Y. Yang, and C. C. Chang. 2012. 'Kruppel-like factor 4 is involved in cell scattering induced by hepatocyte growth factor', J Cell Sci, 125: 4853-64.

Lanz, R. B., N. J. McKenna, S. A. Onate, U. Albrecht, J. Wong, S. Y. Tsai, M. J. Tsai, and B. W. O'Malley. 1999. 'A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex', Cell, 97: 17-27.

Lavinsky, R. M., K. Jepsen, T. Heinzel, J. Torchia, T. M. Mullen, R. Schiff, A. L. Del-Rio, M. Ricote, S. Ngo, J. Gemsch, S. G. Hilsenbeck, C. K. Osborne, C. K. Glass, M. G. Rosenfeld, and D. W. Rose. 1998. 'Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes', Proc Natl Acad Sci U S A, 95: 2920-5.

Lee, J. M., J. L. Hays, C. M. Annunziata, A. M. Noonan, L. Minasian, J. A. Zujewski, M. Yu, N. Gordon, J. Ji, T. M. Sissung, W. D. Figg, N. Azad, B. J. Wood, J. Doroshow, and E. C. Kohn. 2014. 'Phase I/Ib study of olaparib and carboplatin in BRCA1 or BRCA2 mutation-associated breast or ovarian cancer with biomarker analyses', J Natl Cancer Inst, 106: dju089.

Lee, T. I., and R. A. Young. 2013. 'Transcriptional regulation and its misregulation in disease', Cell, 152: 1237-51.

Levin, E. R. 2003. 'Bidirectional signaling between the estrogen receptor and the epidermal growth factor receptor', Mol Endocrinol, 17: 309-17.

Li, Q., J. H. Seo, B. Stranger, A. McKenna, I. Pe'er, T. Laframboise, M. Brown, S. Tyekucheva, and M. L. Freedman. 2013. 'Integrative eQTL-based analyses reveal the biology of breast cancer risk loci', Cell, 152: 633-41.

Li, W., D. Notani, Q. Ma, B. Tanasa, E. Nunez, A. Y. Chen, D. Merkurjev, J. Zhang, K. Ohgi, X. Song, S. Oh, H. S. Kim, C. K. Glass, and M. G. Rosenfeld. 2013. 'Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation', Nature, 498: 516-20.

Li, Y., M. H. Lambert, and H. E. Xu. 2003. 'Activation of nuclear receptors: a perspective from structural genomics', Structure, 11: 741-6.

Lin, C. Y., A. Strom, V. B. Vega, S. L. Kong, A. L. Yeo, J. S. Thomsen, W. C. Chan, B. Doray, D. K. Bangarusamy, A. Ramasamy, L. A. Vergara, S. Tang, A. Chong, V. B. Bajic, L. D. Miller, J. A. Gustafsson, and E. T. Liu. 2004. 'Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells', Genome Biol, 5: R66.

Liu, X. F., and M. K. Bagchi. 2004. 'Recruitment of distinct chromatin-modifying complexes by tamoxifen-complexed estrogen receptor at natural target gene promoters in vivo', J Biol Chem, 279: 15050-8.

Livraghi, L., and J. E. Garber. 2015. 'PARP inhibitors in the management of breast cancer: current data and future prospects', BMC Med, 13: 188.

Lonard, D. M., and B. W. O'Malley. 2012. 'Nuclear receptor coregulators: modulators of pathology and therapeutic targets', Nat Rev Endocrinol, 8: 598-604.

Long, J., Q. Cai, H. Sung, J. Shi, B. Zhang, J. Y. Choi, W. Wen, R. J. Delahanty, W. Lu, Y. T. Gao, H. Shen, S. K. Park, K. Chen, C. Y. Shen, Z. Ren, C. A. Haiman, K. Matsuo, M. K. Kim, U. S. Khoo, M. Iwasaki, Y. Zheng, Y. B. Xiang, K. Gu, N. Rothman, W. Wang, Z. Hu, Y. Liu, K. Y. Yoo, D. Y. Noh, B. G. Han, M. H. Lee, H. Zheng, L. Zhang, P. E. Wu, Y. L. Shieh, S. Y. Chan, S. Wang, X. Xie, S. W. Kim, B. E. Henderson, L. Le Marchand, H. Ito, Y. Kasuga, S. H. Ahn, H. S. Kang, K. Y. Chan, H. Iwata, S. Tsugane, C. Li, X. O. Shu, D. H. Kang, and W. Zheng. 2012. 'Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer', PLoS Genet, 8: e1002532.

Lu, Q., H. Ebling, J. Mittler, W. E. Baur, and R. H. Karas. 2002. 'MAP kinase mediates growth factor-induced nuclear translocation of estrogen receptor alpha', FEBS Lett, 516: 1-8.

Magnani, L., E. B. Ballantyne, X. Zhang, and M. Lupien. 2011. 'PBX1 genomic pioneer function drives ERalpha signaling underlying progression in breast cancer', PLoS Genet, 7: e1002368.

Malik, S., and R. G. Roeder. 2010. 'The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation', Nat Rev Genet, 11: 761-72.

Manavathi, B., V. S. Samanthapudi, and V. N. Gajulapalli. 2014. 'Estrogen receptor coregulators and pioneer factors: the orchestrators of mammary gland cell fate and development', Front Cell Dev Biol, 2: 34.

Mangelsdorf, D. J., C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and R. M. Evans. 1995. 'The nuclear receptor superfamily: the second decade', Cell, 83: 835-9.

Marino, M., P. Galluzzo, and P. Ascenzi. 2006. 'Estrogen signaling multiple pathways to impact gene transcription', Curr Genomics, 7: 497-508.

Maston, G. A., S. K. Evans, and M. R. Green. 2006. 'Transcriptional regulatory elements in the human genome', Annu Rev Genomics Hum Genet, 7: 29-59.

Mathelier, A., O. Fornes, D. J. Arenillas, C. Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, and W. W. Wasserman. 2016. 'JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles', Nucleic Acids Res, 44: D110-5.

McNally, J. G., W. G. Muller, D. Walker, R. Wolford, and G. L. Hager. 2000. 'The glucocorticoid receptor: rapid exchange with regulatory sites in living cells', Science, 287: 1262-5.

Michailidou, K., P. Hall, A. Gonzalez-Neira, M. Ghoussaini, J. Dennis, R. L. Milne, M. K. Schmidt, J. Chang-Claude, S. E. Bojesen, M. K. Bolla, Q. Wang, E. Dicks, A. Lee, C. Turnbull, N. Rahman, Breast, Collaboration Ovarian Cancer Susceptibility, O. Fletcher, J. Peto, L. Gibson, I. Dos Santos Silva, H. Nevanlinna, T. A. Muranen, K. Aittomaki, C. Blomqvist, K. Czene, A. Irwanto, J. Liu, Q. Waisfisz, H. Meijers-Heijboer, M. Adank, Breast Hereditary, Netherlands Ovarian Cancer Research Group, R. B. van der Luijt, R. Hein, N. Dahmen, L. Beckman, A. Meindl, R. K. Schmutzler, B. Muller-Myhsok, P. Lichtner, J. L. Hopper, M. C. Southey, E. Makalic, D. F. Schmidt, A. G. Uitterlinden, A. Hofman, D. J. Hunter, S. J. Chanock, D. Vincent, F. Bacot, D. C. Tessier, S. Canisius, L. F. Wessels, C. A. Haiman, M. Shah, R. Luben, J. Brown, C. Luccarini, N. Schoof, K. Humphreys, J. Li, B. G. Nordestgaard, S. F. Nielsen, H. Flyger, F. J. Couch, X. Wang, C. Vachon, K. N. Stevens, D. Lambrechts, M. Moisse, R. Paridaens, M. R. Christiaens, A. Rudolph, S. Nickels, D. Flesch-Janys, N. Johnson, Z. Aitken, K. Aaltonen, T. Heikkinen, A. Broeks, L. J. Veer, C. E. van der Schoot, P. Guenel, T. Truong, P. Laurent-Puig, F. Menegaux, F. Marme, A. Schneeweiss, C. Sohn, B. Burwinkel, M. P. Zamora, J. I. Perez, G. Pita, M. R. Alonso, A. Cox, I. W. Brock, S. S. Cross, M. W. Reed, E. J. Sawyer, I. Tomlinson, M. J. Kerin, N. Miller, B. E. Henderson, F. Schumacher, L. Le Marchand, I. L. Andrulis, J. A. Knight, G. Glendon, A. M. Mulligan, Investigators kConFab, Group Australian Ovarian Cancer Study, A. Lindblom, S. Margolin, M. J. Hooning, A. Hollestelle, A. M. van den Ouweland, A. Jager, Q. M. Bui, J. Stone, G. S. Dite, C. Apicella, H. Tsimiklis, G. G. Giles, G. Severi, L. Baglietto, P. A. Fasching, L. Haeberle, A. B. Ekici, M. W. Beckmann, H. Brenner, H. Muller, V. Arndt, C. Stegmaier, A. Swerdlow, A. Ashworth, N. Orr, M. Jones, J. Figueroa, J. Lissowska, L. Brinton, M. S. Goldberg, F. Labreche, M. Dumont, R. Winqvist, K. Pylkas, A. Jukkola-Vuorinen, M. Grip, H. Brauch, U. Hamann, T. Bruning, Genica Network, P. Radice, P. Peterlongo, S. Manoukian, B. Bonanni, P. Devilee, R. A. Tollenaar, C. Seynaeve, C. J. van Asperen, A. Jakubowska, J. Lubinski, K. Jaworska, K. Durda, A. Mannermaa, V. Kataja, V. M. Kosma, J. M. Hartikainen, N. V. Bogdanova, N. N. Antonenkova, T. Dork, V. N. Kristensen, H. Anton-Culver, S. Slager, A. E. Toland, S. Edge, F. Fostira, D. Kang, K. Y. Yoo, D. Y. Noh, K. Matsuo, H. Ito, H. Iwata, A. Sueta, A. H. Wu, C. C. Tseng, D. Van Den Berg, D. O. Stram, X. O. Shu, W. Lu, Y. T. Gao, H. Cai, S. H. Teo, C. H. Yip, S. Y. Phuah, B. K. Cornes, M. Hartman, H. Miao, W. Y. Lim, J. H. Sng, K. Muir, A. Lophatananon, S. Stewart-Brown, P. Siriwanarangsan, C. Y. Shen, C. N. Hsiung, P. E. Wu, S. L. Ding, S. Sangrajrang, V. Gaborieau, P. Brennan, J. McKay, W. J. Blot, L. B. Signorello, Q. Cai, W. Zheng, S. Deming-Halverson, M. Shrubsole, J. Long, J. Simard, M. Garcia-Closas, P. D. Pharoah, G. Chenevix-Trench, A. M. Dunning, J. Benitez, and D. F. Easton. 2013. 'Large-scale genotyping identifies 41 new loci associated with breast cancer risk', Nat Genet, 45: 353-61, 61e1-2.

Migliaccio, A., M. Di Domenico, G. Castoria, A. de Falco, P. Bontempo, E. Nola, and F. Auricchio. 1996. 'Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells', EMBO J, 15: 1292-300.

Miller, W. R., A. Larionov, L. Renshaw, T. J. Anderson, J. R. Walker, A. Krause, T. Sing, D. B. Evans, and J. M. Dixon. 2009. 'Gene expression profiles differentiating between breast cancers clinically responsive or resistant to letrozole', J Clin Oncol, 27: 1382-7.

Miranda, T. B., T. C. Voss, M. H. Sung, S. Baek, S. John, M. Hawkins, L. Grontved, R. L. Schiltz, and G. L. Hager. 2013. 'Reprogramming the chromatin landscape: interplay of the estrogen and glucocorticoid receptors at the genomic level', Cancer Res, 73: 5130-9.

Miura, A., K. Yamagata, M. Kakei, H. Hatakeyama, N. Takahashi, K. Fukui, T. Nammo, K. Yoneda, Y. Inoue, F. M. Sladek, M. A. Magnuson, H. Kasai, J. Miyagawa, F. J. Gonzalez, and I. Shimomura. 2006. 'Hepatocyte nuclear factor-4alpha is essential for glucose-stimulated insulin secretion by pancreatic beta-cells', J Biol Chem, 281: 5246-57.

Mohammed, H., C. D'Santos, A. A. Serandour, H. R. Ali, G. D. Brown, A. Atkins, O. M. Rueda, K. A. Holmes, V. Theodorou, J. L. Robinson, W. Zwart, A. Saadi, C. S. Ross-Innes, S. F. Chin, S. Menon, J. Stingl, C. Palmieri, C. Caldas, and J. S. Carroll. 2013. 'Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor', Cell Rep, 3: 342-9.

Mohammed, H., I. A. Russell, R. Stark, O. M. Rueda, T. E. Hickey, G. A. Tarulli, A. A. Serandour, S. N. Birrell, A. Bruna, A. Saadi, S. Menon, J. Hadfield, M. Pugh, G. V. Raj, G. D. Brown, C. D'Santos, J. L. Robinson, G. Silva, R. Launchbury, C. M. Perou, J. Stingl, C. Caldas, W. D. Tilley, and J. S. Carroll. 2015. 'Progesterone receptor modulates ERalpha action in breast cancer', Nature, 523: 313-7.

Morgan, L., J. Gee, S. Pumford, L. Farrow, P. Finlay, J. Robertson, I. Ellis, H. Kawakatsu, R. Nicholson, and S. Hiscox. 2009. 'Elevated Src kinase activity attenuates Tamoxifen response in vitro and is associated with poor prognosis clinically', Cancer Biol Ther, 8: 1550-8.

Mosselman, S., J. Polman, and R. Dijkema. 1996. 'ER beta: identification and characterization of a novel human estrogen receptor', FEBS Lett, 392: 49-53.

Muerdter, F., L. M. Boryn, and C. D. Arnold. 2015. 'STARR-seq - principles and applications', Genomics, 106: 145-50.

Murtha, M., Z. Tokcaer-Keskin, Z. Tang, F. Strino, X. Chen, Y. Wang, X. Xi, C. Basilico, S. Brown, R. Bonneau, Y. Kluger, and L. Dailey. 2014. 'FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells', Nat Methods, 11: 559-65.

Musgrove, E. A., and R. L. Sutherland. 2009. 'Biological determinants of endocrine resistance in breast cancer', Nat Rev Cancer, 9: 631-43.

Nagy, L., and J. W. Schwabe. 2004. 'Mechanism of the nuclear receptor molecular switch', Trends Biochem Sci, 29: 317-24.

Noti, J. D., A. K. Johnson, and J. D. Dillon. 2005. 'The leukocyte integrin gene CD11d is repressed by gut-enriched Kruppel-like factor 4 in myeloid cells', J Biol Chem, 280: 3449-57.

Novac, N., and T. Heinzel. 2004. 'Nuclear receptors: overview and classification', Curr Drug Targets Inflamm Allergy, 3: 335-46.

O'Lone, R., M. C. Frith, E. K. Karlsson, and U. Hansen. 2004. 'Genomic targets of nuclear estrogen receptors', Mol Endocrinol, 18: 1859-75.

Onate, S. A., V. Boonyaratanakornkit, T. E. Spencer, S. Y. Tsai, M. J. Tsai, D. P. Edwards, and B. W. O'Malley. 1998. 'The steroid receptor coactivator-1 contains multiple receptor interacting and activation domains that cooperatively enhance the activation function 1 (AF1) and AF2 domains of steroid receptors', J Biol Chem, 273: 12101-8.

Onate, S. A., S. Y. Tsai, M. J. Tsai, and B. W. O'Malley. 1995. 'Sequence and characterization of a coactivator for the steroid hormone receptor superfamily', Science, 270: 1354-7.

Osborne, C. K., V. Bardou, T. A. Hopp, G. C. Chamness, S. G. Hilsenbeck, S. A. Fuqua, J. Wong, D. C. Allred, G. M. Clark, and R. Schiff. 2003. 'Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer', J Natl Cancer Inst, 95: 353-61.

Osborne, C. K., and R. Schiff. 2011. 'Mechanisms of endocrine resistance in breast cancer', Annu Rev Med, 62: 233-47.

Pandya, A. Y., L. I. Talley, A. R. Frost, T. J. Fitzgerald, V. Trivedi, M. Chakravarthy, D. C. Chhieng, W. E. Grizzle, J. A. Engler, H. Krontiras, K. I. Bland, A. F. LoBuglio, S. M. Lobo-Ruppert, and J. M. Ruppert. 2004. 'Nuclear localization of KLF4 is associated with an aggressive phenotype in early-stage breast cancer', Clin Cancer Res, 10: 2709-19.

Patwardhan, R. P., J. B. Hiatt, D. M. Witten, M. J. Kim, R. P. Smith, D. May, C. Lee, J. M. Andrie, S. I. Lee, G. M. Cooper, N. Ahituv, L. A. Pennacchio, and J. Shendure. 2012. 'Massively parallel functional dissection of mammalian enhancers in vivo', Nat Biotechnol, 30: 265-70.

Perez-Tenorio, G., F. Berglund, A. Esguerra Merca, B. Nordenskjold, L. E. Rutqvist, L. Skoog, and O. Stal. 2006. 'Cytoplasmic p21WAF1/CIP1 correlates with Akt activation and poor response to tamoxifen in breast cancer', Int J Oncol, 28: 1031-42.

Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown, and D. Botstein. 2000. 'Molecular portraits of human breast tumours', Nature, 406: 747-52.

Piccart-Gebhart, M. J., M. Procter, B. Leyland-Jones, A. Goldhirsch, M. Untch, I. Smith, L. Gianni, J. Baselga, R. Bell, C. Jackisch, D. Cameron, M. Dowsett, C. H. Barrios, G. Steger, C. S. Huang, M. Andersson, M. Inbar, M. Lichinitser, I. Lang, U. Nitz, H. Iwata, C. Thomssen, C. Lohrisch, T. M. Suter, J. Ruschoff, T. Suto, V. Greatorex, C. Ward, C. Straehle, E. McFadden, M. S. Dolci, R. D. Gelber, and Team Herceptin Adjuvant Trial Study. 2005. 'Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer', N Engl J Med, 353: 1659-72.

Pratt, W. B., and D. O. Toft. 1997. 'Steroid receptor interactions with heat shock protein and immunophilin chaperones', Endocr Rev, 18: 306-60.

Privalsky, M. L. 2004. 'The role of corepressors in transcriptional regulation by nuclear hormone receptors', Annu Rev Physiol, 66: 315-60.

Prossnitz, E. R., J. B. Arterburn, H. O. Smith, T. I. Oprea, L. A. Sklar, and H. J. Hathaway. 2008. 'Estrogen signaling through the transmembrane G protein-coupled receptor GPR30', Annu Rev Physiol, 70: 165-90.

Prossnitz, E. R., L. A. Sklar, T. I. Oprea, and J. B. Arterburn. 2008. 'GPR30: a novel therapeutic target in estrogen-related disease', Trends Pharmacol Sci, 29: 116-23.

Ramamoorthy, S., and Z. Nawaz. 2008. 'E6-associated protein (E6-AP) is a dual function coactivator of steroid hormone receptors', Nucl Recept Signal, 6: e006.

Rands, C. M., S. Meader, C. P. Ponting, and G. Lunter. 2014. '8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage', PLoS Genet, 10: e1004525.

Reis-Filho, J. S., and A. N. Tutt. 2008. 'Triple negative tumours: a critical review', Histopathology, 52: 108-18.

Revankar, C. M., D. F. Cimino, L. A. Sklar, J. B. Arterburn, and E. R. Prossnitz. 2005. 'A transmembrane intracellular estrogen receptor mediates rapid cell signaling', Science, 307: 1625-30.

Roadmap Epigenomics, Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, and M. Kellis. 2015. 'Integrative analysis of 111 reference human epigenomes', Nature, 518: 317-30.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', Bioinformatics, 26: 139-40.

Ross, J. S., E. A. Slodkowska, W. F. Symmans, L. Pusztai, P. M. Ravdin, and G. N. Hortobagyi. 2009. 'The HER-2 receptor and breast cancer: ten years of targeted anti-HER-2 therapy and personalized medicine', Oncologist, 14: 320-68.

Rowland, B. D., R. Bernards, and D. S. Peeper. 2005. 'The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene', Nat Cell Biol, 7: 1074-82.

Sabbah, M., D. Courilleau, J. Mester, and G. Redeuilh. 1999. 'Estrogen induction of the cyclin D1 promoter: involvement of a cAMP response-like element', Proc Natl Acad Sci U S A, 96: 11217-22.

Safe, S. 2001. 'Transcriptional activation of genes by 17 beta-estradiol through estrogen receptor-Sp1 interactions', Vitam Horm, 62: 231-52.

Saldanha, A. J. 2004. 'Java Treeview--extensible visualization of microarray data', Bioinformatics, 20: 3246-8.

Sandhu, S. K., W. R. Schelman, G. Wilding, V. Moreno, R. D. Baird, S. Miranda, L. Hylands, R. Riisnaes, M. Forster, A. Omlin, N. Kreischer, K. Thway, H. Gevensleben, L. Sun, J. Loughney, M. Chatterjee, C. Toniatti, C. L. Carpenter, R. Iannone, S. B. Kaye, J. S. de Bono, and R. M. Wenham. 2013. 'The poly(ADP-ribose) polymerase inhibitor niraparib (MK4827) in BRCA mutation carriers and patients with sporadic cancer: a phase 1 dose-escalation trial', Lancet Oncol, 14: 882-92.

Shi, L., B. Dong, Z. Li, Y. Lu, T. Ouyang, J. Li, T. Wang, Z. Fan, T. Fan, B. Lin, Z. Wang, and Y. Xie. 2009. 'Expression of ER-{alpha}36, a novel variant of estrogen receptor {alpha}, and resistance to tamoxifen treatment in breast cancer', J Clin Oncol, 27: 3423-9.

Shlyueva, D., G. Stampfel, and A. Stark. 2014. 'Transcriptional enhancers: from properties to genome-wide predictions', Nat Rev Genet, 15: 272-86.

Shlyueva, D., C. Stelzer, D. Gerlach, J. O. Yanez-Cuna, M. Rath, L. M. Boryn, C. D. Arnold, and A. Stark. 2014. 'Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin', Mol Cell, 54: 180-92.

Siegel, R. L., K. D. Miller, and A. Jemal. 2016. 'Cancer statistics, 2016', CA Cancer J Clin, 66: 7-30.

Simpson, E. R., M. S. Mahendroo, G. D. Means, M. W. Kilgore, M. M. Hinshelwood, S. Graham-Lorence, B. Amarneh, Y. Ito, C. R. Fisher, M. D. Michael, and et al. 1994. 'Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis', Endocr Rev, 15: 342-55.

Slamon, D. J., B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, J. Baselga, and L. Norton. 2001. 'Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2', N Engl J Med, 344: 783-92.

Smith, C. L., and B. W. O'Malley. 2004. 'Coregulator function: a key to understanding tissue specificity of selective receptor modulators', Endocr Rev, 25: 45-71.

Smith, D. F. 1993. 'Dynamics of heat shock protein 90-progesterone receptor binding and the disactivation loop model for steroid receptor complexes', Mol Endocrinol, 7: 1418-29.

Smith, I. E., and M. Dowsett. 2003. 'Aromatase inhibitors in breast cancer', N Engl J Med, 348: 2431-42.

Smith, R. P., L. Taher, R. P. Patwardhan, M. J. Kim, F. Inoue, J. Shendure, I. Ovcharenko, and N. Ahituv. 2013. 'Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model', Nat Genet, 45: 1021-8.

Song, L., and G. E. Crawford. 2010. 'DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells', Cold Spring Harb Protoc, 2010: pdb prot5384.

Sonoda, J., L. Pei, and R. M. Evans. 2008. 'Nuclear receptors: decoding metabolic disease', FEBS Lett, 582: 2-9.

Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning, and A. L. Borresen-Dale. 2001. 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', Proc Natl Acad Sci U S A, 98: 10869-74.

Span, P. N., V. C. Tjan-Heijnen, P. Manders, L. V. Beex, and C. G. Sweep. 2003. 'Cyclin-E is a strong predictor of endocrine therapy failure in human breast cancer', Oncogene, 22: 4898-904.

Staaf, J., M. Ringner, J. Vallon-Christersson, G. Jonsson, P. O. Bendahl, K. Holm, A. Arason, H. Gunnarsson, C. Hegardt, B. A. Agnarsson, L. Luts, D. Grabau, M. Ferno, P. O. Malmstrom, O. T. Johannsson, N. Loman, R. B. Barkardottir, and A. Borg. 2010. 'Identification of subtypes in human epidermal growth factor receptor 2--positive breast cancer reveals a gene signature prognostic of outcome', J Clin Oncol, 28: 1813-20.

Stampfel, G., T. Kazmar, O. Frank, S. Wienerroither, F. Reiter, and A. Stark. 2015. 'Transcriptional regulators form diverse groups with context-dependent regulatory functions', Nature, 528: 147-51.

Stender, J. D., K. Kim, T. H. Charn, B. Komm, K. C. Chang, W. L. Kraus, C. Benner, C. K. Glass, and B. S. Katzenellenbogen. 2010. 'Genome-wide analysis of estrogen receptor alpha DNA binding and tethering mechanisms identifies Runx1 as a novel tethering factor in receptor-mediated transcriptional activation', Mol Cell Biol, 30: 3943-55.

Stunnenberg, H. G. 1993. 'Mechanisms of transactivation by retinoic acid receptors', Bioessays, 15: 309-15.

Sung, M. H., M. J. Guertin, S. Baek, and G. L. Hager. 2014. 'DNase footprint signatures are dictated by factor dynamics and DNA sequence', Mol Cell, 56: 275-85.

Takahashi, K., and S. Yamanaka. 2006. 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors', Cell, 126: 663-76.

Tamrazi, A., K. E. Carlson, J. R. Daniels, K. M. Hurth, and J. A. Katzenellenbogen. 2002. 'Estrogen receptor dimerization: ligand binding regulates dimer affinity and dimer dissociation rate', Mol Endocrinol, 16: 2706-19.

Tan, S. K., Z. H. Lin, C. W. Chang, V. Varang, K. R. Chng, Y. F. Pan, E. L. Yong, W. K. Sung, and E. Cheung. 2011. 'AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription', EMBO J, 30: 2569-81.

Tenbaum, S., and A. Baniahmad. 1997. 'Nuclear receptors: structure, function and involvement in disease', Int J Biochem Cell Biol, 29: 1325-41.

Theodorou, V., R. Stark, S. Menon, and J. S. Carroll. 2013. 'GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility', Genome Res, 23: 12-22.

Thomas, C., and J. A. Gustafsson. 2011. 'The different roles of ER subtypes in cancer biology and therapy', Nat Rev Cancer, 11: 597-608.

Thomas, P., Y. Pang, E. J. Filardo, and J. Dong. 2005. 'Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells', Endocrinology, 146: 624-32.

Tiwari, N., N. Meyer-Schaller, P. Arnold, H. Antoniadis, M. Pachkov, E. van Nimwegen, and G. Christofori. 2013. 'Klf4 is a transcriptional regulator of genes critical for EMT, including Jnk1 (Mapk8)', PLoS One, 8: e57329.

Tontonoz, P., and D. J. Mangelsdorf. 2003. 'Liver X receptor signaling pathways in cardiovascular disease', Mol Endocrinol, 17: 985-93.

Umayahara, Y., R. Kawamori, H. Watada, E. Imano, N. Iwama, T. Morishima, Y. Yamasaki, Y. Kajimoto, and T. Kamada. 1994. 'Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer', J Biol Chem, 269: 16433-42.

Vanhille, L., A. Griffon, M. A. Maqbool, J. Zacarias-Cabeza, L. T. Dao, N. Fernandez, B. Ballester, J. C. Andrau, and S. Spicuglia. 2015. 'High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq', Nat Commun, 6: 6905.

Vivacqua, A., R. Lappano, P. De Marco, D. Sisci, S. Aquila, F. De Amicis, S. A. Fuqua, S. Ando, and M. Maggiolini. 2009. 'G protein-coupled receptor 30 expression is up-regulated by EGF and TGF alpha in estrogen receptor alpha-positive cancer cells', Mol Endocrinol, 23: 1815-26.

Wakeling, A. E., and J. Bowler. 1992. 'ICI 182,780, a new antioestrogen with clinical potential', J Steroid Biochem Mol Biol, 43: 173-7.

Wallace, J. A., and G. Felsenfeld. 2007. 'We gather together: insulators and genome organization', Curr Opin Genet Dev, 17: 400-7.

Walter, P., S. Green, G. Greene, A. Krust, J. M. Bornert, J. M. Jeltsch, A. Staub, E. Jensen, G. Scrace, M. Waterfield, and et al. 1985. 'Cloning of the human estrogen receptor cDNA', Proc Natl Acad Sci U S A, 82: 7889-93.

Wang, L. H., X. Y. Yang, K. Mihalic, W. Xiao, D. Li, and W. L. Farrar. 2001. 'Activation of estrogen receptor blocks interleukin-6-inducible cell growth of human multiple myeloma involving molecular

cross-talk between estrogen receptor and STAT3 mediated by coregulator PIAS3', J Biol Chem, 276: 31839-44.

Wang, Y., and R. J. Miksicek. 1991. 'Identification of a dominant negative form of the human estrogen receptor', Mol Endocrinol, 5: 1707-15.

Wang, Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang, and K. Zhao. 2008. 'Combinatorial patterns of histone acetylations and methylations in the human genome', Nat Genet, 40: 897-903.

Wang, Z., X. Zhang, P. Shen, B. W. Loggie, Y. Chang, and T. F. Deuel. 2005. 'Identification, cloning, and expression of human estrogen receptor-alpha36, a novel variant of human estrogen receptor-alpha66', Biochem Biophys Res Commun, 336: 1023-7.

Warnmark, A., E. Treuter, A. P. Wright, and J. A. Gustafsson. 2003. 'Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation', Mol Endocrinol, 17: 1901-9.

Webb, P., P. Nguyen, J. Shinsako, C. Anderson, W. Feng, M. P. Nguyen, D. Chen, S. M. Huang, S. Subramanian, E. McKinerney, B. S. Katzenellenbogen, M. R. Stallcup, and P. J. Kushner. 1998. 'Estrogen receptor activation function 1 works by binding p160 coactivator proteins', Mol Endocrinol, 12: 1605-18.

Wei, D., W. Gong, M. Kanai, C. Schlunk, L. Wang, J. C. Yao, T. T. Wu, S. Huang, and K. Xie. 2005. 'Drastic down-regulation of Kruppel-like factor 4 expression is critical in human gastric cancer development and progression', Cancer Res, 65: 2746-54.

Weinberger, C., S. M. Hollenberg, E. S. Ong, J. M. Harmon, S. T. Brower, J. Cidlowski, E. B. Thompson, M. G. Rosenfeld, and R. M. Evans. 1985. 'Identification of human glucocorticoid receptor complementary DNA clones by epitope selection', Science, 228: 740-2.

Welboren, W. J., M. A. van Driel, E. M. Janssen-Megens, S. J. van Heeringen, F. C. Sweep, P. N. Span, and H. G. Stunnenberg. 2009. 'ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands', EMBO J, 28: 1418-28.

Williams, C., K. Edvardsson, S. A. Lewandowski, A. Strom, and J. A. Gustafsson. 2008. 'A genome-wide study of the repressive effects of estrogen receptor beta on estrogen receptor alpha signaling in breast cancer cells', Oncogene, 27: 1019-32.

Woodfield, G. W., A. D. Horan, Y. Chen, and R. J. Weigel. 2007. 'TFAP2C controls hormone response in breast cancer cells through multiple pathways of estrogen signaling', Cancer Res, 67: 8439-43.

Wright, J. B., and N. E. Sanjana. 2016. 'CRISPR Screens to Discover Functional Noncoding Elements', Trends Genet.

Wurtz, J. M., W. Bourguet, J. P. Renaud, V. Vivat, P. Chambon, D. Moras, and H. Gronemeyer. 1996. 'A canonical structure for the ligand-binding domain of nuclear receptors', Nat Struct Biol, 3: 206.

Xu, H. E. 2015. 'Family reunion of nuclear hormone receptors: structures, diseases, and drug discovery', Acta Pharmacol Sin, 36: 1-2.

Yardimci, G. G., C. L. Frank, G. E. Crawford, and U. Ohler. 2014. 'Explicit DNase sequence bias modeling enables high-resolution transcription factor footprint detection', Nucleic Acids Res, 42: 11865-78.

Yoon, H. S., X. Chen, and V. W. Yang. 2003. 'Kruppel-like factor 4 mediates p53-dependent G1/S cell cycle arrest in response to DNA damage', J Biol Chem, 278: 2101-5.

Yu, F., J. Li, H. Chen, J. Fu, S. Ray, S. Huang, H. Zheng, and W. Ai. 2011. 'Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion', Oncogene, 30: 2161-72.

Yuh, C. H., H. Bolouri, and E. H. Davidson. 1998. 'Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene', Science, 279: 1896-902.

Zabidi, M. A., C. D. Arnold, K. Schernhuber, M. Pagani, M. Rath, O. Frank, and A. Stark. 2015. 'Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation', Nature, 518: 556-9.

Zaret, K. S., and J. S. Carroll. 2011. 'Pioneer transcription factors: establishing competence for gene expression', Genes Dev, 25: 2227-41.

Zhang, Q. X., A. Borg, D. M. Wolf, S. Oesterreich, and S. A. Fuqua. 1997. 'An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer', Cancer Res, 57: 1244-9.

Zhang, W., J. M. Shields, K. Sogawa, Y. Fujii-Kuriyama, and V. W. Yang. 1998. 'The gut-enriched Kruppel-like factor suppresses the activity of the CYP1A1 promoter in an Sp1-dependent fashion', J Biol Chem, 273: 17917-25.

Zhang, X., and Z. Y. Wang. 2013. 'Estrogen receptor-alpha variant, ER-alpha36, is involved in tamoxifen resistance and estrogen hypersensitivity', Endocrinology, 154: 1990-8.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. 2008. 'Model-based analysis of ChIP-Seq (MACS)', Genome Biol, 9: R137.

Zhou, Y., C. Yau, J. W. Gray, K. Chew, S. H. Dairkee, D. H. Moore, U. Eppenberger, S. Eppenberger-Castori, and C. C. Benz. 2007. 'Enhanced NF kappa B and AP-1 transcriptional activity associated with antiestrogen resistant breast cancer', BMC Cancer, 7: 59.