MOLECULAR MECHANISMS AND FUNCTIONS OF ESTROGEN RECEPTOR ENHANCERS IN HORMONE-DEPENDENT GENE EXPRESSION

APPROVED BY SUPERVISORY COMMITTEE

W. Lee Kraus, Ph.D.
Taekyung Kim, Ph.D.
Steven Kliewer, Ph.D.
Chun-Li Zhang, Ph.D.

MOLECULAR MECHANISMS AND FUNCTIONS OF ESTROGEN RECEPTOR ENHANCERS IN HORMONE-DEPENDENT GENE EXPRESSION

by

SHINO MURAKAMI

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

December, 2017

ACKNOWLEDGMEN

I would like to express my sincere appreciation to my supervisor, Dr. W. Lee Kraus, for his continuous encouragement and support throughout my training in the graduate school. My journey in the past several years in the Kraus lab was truly invaluable. I am deeply grateful for him to grant me independency in great extent toward my work and to be always there patiently for the moments when I need his guidance. He has provided me with a variety of additional training opportunities including many pieces of career advise, attendance at international conferences, meetings with leading scientists, reviewing manuscripts for prominent journals, and leaning scientific communication and writing skills. Every single opportunity like above taught me vital lessens from which I leaned how to be a better scientist and why it is important to be responsible for my science and to contribute to the field in many ways.

I also would like to thank to the members of my thesis committee, Dr. Taekyung Kim, Dr. Steven Kliewer, and Dr. Chun-Li Zhang for the oversight on my progress, insightful scientific comments, and career advice. It is also important for me to express my gratitude to Dr. L. Rogelio Cruz-Vera, Dr. Jacqueline Moreno, Dr. Carmen Scholz, Dr. Lynn Boyd, and Dr. Roy Magnuson at the University of Alabama in Huntsville, where I obtained my B.S. degree, for seeding the amazing opportunities for my graduate education. Without their support, my experiences in graduate school would not have been the same.

The Kraus lab was foremost the central part of my graduate training. I would not be able to express enough appreciation to each of the fellow trainees in the Kraus lab to shape

up my experiences while I was there. I am grateful that they always provided kind companionship and cooperation to make the lab a wonderful place to work at. In addition, I leaned countless lessens from each one of them. Many critical, thoughtful, and constructive discussions with them over various scientific topics and our own projects were a crucial part of my experiences in my graduate training. Among the lab members, I would like to express special thanks to Anusha Nagari, M.S., Green Center computational biologist, for our long-term collaboration over many projects. Since we joined the Kraus lab around the same time, I believe we learned a lot from each other. I appreciate her expertise on data analysis and commitment to her responsibility and leadership. I am also particularly thankful to Drs. Hector L. Franco and Tim Hou for working closely, sharing scientific interests with me, and being wonderful colleges to work with. Numbers of discussions, conferences, and seminars that we had together had kept my life in the lab exciting. Additionally, I would like to thank Drs. Shrikanth Gadad, Yasmin Vasquez, and Sudeshna Tripathy for helpful discussions and sharing common interests in hormone-dependent gene regulation and physiology. Lastly, but not least, I would also like to thank the technical and administrative support teams of the Kraus lab, especially Ms. Constance Grafer and Ms. Deborah Miller for their continuous support with tremendous patience. I wish all and each of them the best for their future. I am very much looking forward to seeing them again somewhere in the future.

BIOGRAPHICAL SKETCH

Shino Murakami was born and raised in the prefecture of Ehime, Japan. After graduation from high school, she started her undergraduate education at the University of Alabama in Huntsville, where she learned fundamentals in chemistry and biology covering a range of topics including inorganic and organic chemistry, analytical chemistry, biochemistry and biophysics, molecular and cellular biology, and microbiology and immunology. Beside her course work, she joined the laboratory of Dr. L. Rogelio Cruz-Vera to study the amino acid-mediated regulation on prokaryotic translation. During this time, she also gained experiences in the laboratory of Dr. Carmen Scholz as an undergraduate research assistance to synthesize and study physical properties of clickable polymer polyhydroxyoctanoate and its derivatives. In 2009, she received a B.S. degree with a double major in chemistry and biology with magna cum laude. The following year, she started her graduate study in Genes, Development, and Disease program at the University of Texas Southwestern Medical Center at Dallas. In 2011, she joined the laboratory of Dr. W. Lee Kraus for her dissertation study on the hormone-dependent gene regulation. Using genomic, molecular, and cell-based approaches, her work in the Kraus lab provided a better understanding on the molecular mechanisms and functions of hormone-dependent transcription enhancers for gene regulation. In 2017, she completed her dissertation study and received Ph.D. in biomedical sciences.

MOLECULAR MECHANISMS AND FUNCTIONS OF ESTROGEN RECEPTOR ENHANCERS IN HORMONE-DEPENDENT GENE EXPRESSION

Shino Murakami, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2017

Supervising Professor: W. Lee Kraus, Ph.D.

Transcription is a fundamental regulatory mechanism of biological processes in a range of physiological and pathological conditions. Transcription enhancers are DNA regulatory elements that regulate the expression of the target genes by accommodating transcription factor (TF) binding through sequence specificity. Estrogen receptor alpha (ERα) belongs to ligand-dependent nuclear receptor superfamily. Upon activation by estrogenic ligands, ERα binds to specific sites on chromatin, and assembles and activates enhancer complexes, which in turn lead to the transcription of target genes. Various molecular events have been associated with enhancer function, including coregulator

recruitment, induction of enhancer-enriched histone modifications, nucleosome remodeling, enhancer-promoter chromatin interactions, and transcription activation at the enhancer, as well as the target gene promoter. However, we lack a clear understanding of the order of events, the specific roles of each coregulator and enhancer-enriched chromatin features, and the functional relationships among them. Using ERα in estrogen (E2)-regulated gene transcription as a model in combination with molecular and cellular biology, as well as genomic and computational approaches, my dissertation herein describes a series of studies elucidating the molecular mechanisms and functions of these evens that lead to ERα enhancer activation. Collectively, it demonstrates that (1) ERα enhancer assembly and activation is a dynamic process, (2) the temporally-defined recruitment and activation of key coregulators are required for successful activation of ERα enhancers, and (3) enhancer transcripts (eRNA) mark active enhancers.

Lastly, I delineate the development of a new technology, single-cell Global Run-on Sequencing (scGRO-seq), to uncover the link between enhancer activity and target gene transcription at the single-cell level. Single-cell imaging and sequencing technologies have demonstrated the heterogeneous nature of gene expression and enhancer activity in a wide range of biological systems, including clonally-expanded populations of cultured cells. However, our understanding on the molecular basis of heterogeneous gene expression is limited because of a lack of technologies that allow us to simultaneously examine enhancer activity and target gene transcription at the single-cell level. scGRO-seq will overcome this problem by capturing active transcription at the enhancers, which is an indicative of enhancer activity, and at the target gene in the same cells.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Martinez, A.K.*, Shirole, N.H.*, Murakami, S., Benedik, M.J., Sachs, M.S., and Cruz-Vera, L.R. (2012). Crucial elements that maintain the interactions between the regulatory TnaC peptide and the ribosome exit tunnel responsible for Trp inhibition of ribosome function. Nucleic Acids Res 40, 2247-2257.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res 23, 1210-1223.

Murakami, S., Gadad, S.S., and Kraus, W.L. (2015). A PreSTIGEous use of LncRNAs to predict enhancers. Cell Cycle 14, 1619-1620. [Preview]

Nagari, A., Murakami, S., Malladi, V.S., and Kraus, W.L. (2017). Computational approaches for mining GRO-Seq data to identify and characterize active enhancers. Methods Mol Biol 1468, 121-138.

Raj, G.V., Sareddy, G.R., Ma, S., Lee, T.K., Viswanadhapalli, S., Li, R., Liu, X., Murakami, S., Chen, C.C., Lee, W.R., *et al*. (2017). Estrogen receptor coregulator binding modulators (ERXs) effectively target estrogen receptor positive human breast cancers. Elife 6.

Murakami, S., Nagari, A., and Kraus, W.L. (2017). Dynamic assembly and activation of estrogen receptor α enhancers through coregulator switching. Genes Dev 31.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

3C – Chromosome conformation capture

AF-1/2 – Activation domain 1/2

CE – Core enhancer

ChIA-PET – Chromatin interaction analysis by paired-end tag sequencing

ChIP-exo –Chromatinimmunoprecipitaion with exonuclease treatment

ChIP-seq – Chromatinimmunoprecipitation and sequencing

DBD –DNA binding domain

DRRs – Distal regulatory regions

DNase-seq –DNase I hypersensitive sites sequencing

E2 – 17β-estradiol

ERα – Estrogen receptor alpha

ERβ – Estrogen receptor beta

ERBS – Estrogen receptor binding site

eRNA – Enhancer transcript, enhancer RNA

FP – Flavopiridol

GRO-seq – Global nuclear run-on sequencing

GTF – General transcription factor

H3.3 – Histone 3 variant 3

H3K4me1 – Histone 3 lysine 4 monomethylation

H3K4me2 – Histone 3 lysine 4 dimethylation

H3K9me3 – Histone 3 lysine 9 trimethylation

H3K27ac – Histone 3 lysine 27 acetylation

H3K27me3 – Histone 3 lysine 27 trimethylation

H12 – α-helix 12 of AF-2 in ERα

Hi-C – High-throughput genome-wide chromosome conformation capture

kb – Kilobases pairs of nucleotides

LBD – Ligand binding domain

LCR – locus control region

Mb – Megabases pairs of nucleotides

NELF – Negative elongation factor

NFR – Nucleosome free region

PID – p300-interaction domain of SRC2

Pol II – RNA polymerases II

qPCR – quantitative PCR

RNA-a – activating RNA

RNA FISH – RNA fluorescent in situ hybridization

RRD –RNA recognition domain

scGRO-seq – Single-cell global run-con sequencing

SERM – Selective estrogen receptor modulator

TCGA – The cancer genome atlas

TF – Transcription factor

TSS – Transcription start site

CHAPTER ONE

Introduction

1.1 Summary

Transcription is a tightly regulated biological process that is critical in a wide range of physiological conditions. Transcription factor (TF) binding to the genome triggers formation of the active enhancer complex that involves numbers of coregulators. Dynamic enhancer complex formation causes various changes at the enhancer as well as the target gene promoters, including alteration in the chromatin environment and histone modifications and formation of enhancer-promoter chromatin interactions. Although many features and coregulators in the enhancer complex have been identified, the order of assembly and functions of dynamic enhancer complex formation is not well characterized. Using estrogen receptor alpha (ERα) in estrogen (E2)-regulated gene transcription as a model system, I ought to investigate the molecular mechanisms and functions of enhancer complex formation.

1.2 Transcription Enhancers

Gene transcription is a fundamental regulatory step in a wide array of biological processes. Various extracellular as well as intracellular stimuli provoke transcription responses by controlling multi-layered transcription regulatory steps involving numbers of coregulators including activators and repressors on the chromatin (Lee and Young, 2013). For transcription activation, transcription factors bind to the DNA regulatory elements called enhancers across the genome to nucleate de novo formation of enhancer complexes (a.k.a. enhanceosomes) by recruiting numbers of coregulators (Carey, 1998). TF and coregulators collectively cause changes in histone codes, nucleosome occupancies, and chromatin accessibility. To this end, the newly formed enhancer complex induces chromatin interaction between the enhancer and the target gene promoter, and activates transcription both at the enhancer and at the target gene prompter (Ong and Corces, 2011). The active transcription occurring at the enhancer generates non-coding RNAs, named enhancer transcripts or eRNAs (Hah et al., 2011; Kim et al., 2010; Lam et al., 2013; Li et al., 2013). Through genome-wide TF binding and enhancer formation, a set of thousands of target gene expression is induced by a given TF under a given stimulus (Hah et al., 2011; Lam et al., 2013; Mahat et al., 2016). Many key players required between TF binding to the productive transcription activation have been identified; however, the precise mechanisms and the functions of each activation step are still not well understood. Although various principles are often conserved across eukaryotes, there are also distinct aspects in transcription regulation across species. This chapter will describe the transcription regulation in mammalian cells unless otherwise stated.

1.2.1 Chromatin Environment at Enhancers

Beside TF binding, enhancers exhibit several unique features. First, they tend to adopt open chromatin states by bearing fragile nucleosomes by incorporation of unique histone variants such as H3.3 and H2A.Z (Calo and Wysocka, 2013; Struhl and Segal, 2013). In addition, nucleosomes at enhancers are well positions as compared to the non-regulatory intergenic regions in the genome (Gaffney et al., 2012; Ho et al., 2006). While well positioned nucleosomes are biased on the transcription directionality at promoters, nucleosomes are similarly well positioned on the both sides of enhancers, which might be due to bidirectional nature of enhancer transcription as mentioned below. When TFs bind to enhancers, they recruit nucleosome remodelers along with other coregulators that further contribute to the TF-dependent changes of the chromatin environment at enhancers that eventually modulate activator-dependent gene transcription (He et al., 2010; Voss and Hager, 2014).

In addition to nucleosome occupancy and remodeling, enhancers are enriched with many specific histone modifications. Vigorous studies using chromatin immunoprecipitation and sequencing (ChIP-seq) assays for histone modifications in the last decade have provided opportunities to discover the genomic nature of enhancers. Based on these observations, histone modification ChIP-seq data became widely used to annotate enhancers genome-wide and to identify and group different types of enhancers (Heintzman et al., 2009; Heintzman et al., 2007; Rada-Iglesias et al., 2011). For instance, enhancer marks such as H3 lysine 27 trimethylation (H3K27me3) and H3 lysine 9 trimethylation (H3K9me3) are enriched for inactive enhancers, while other modifications, including many acetylation marks such as H3

lsyine 4 mono- and dimethylation (H3K4me1/2) and H3 lysine 27 acetylation (H3K27ac), are associated with active enhancers (Calo and Wysocka, 2013; Ong and Corces, 2011). Furthermore H3K4me1 is often enriched at closed and poised enhancers prior to TF binding, which are enhancers that are not currently active but awaiting for future activation. On the other hand, some modifications including H3K27ac and many acetylation on other histone lysines are dynamically controlled for some extent after TF binding and enhancer activation (Creyghton et al., 2010; Heintzman et al., 2009; Pradeepa et al., 2016; Taylor et al., 2013). Despite of our knowledge on the large repertoire of enhancer-enriched histone modifications, we have relatively limited knowledge regarding their mechanisms and functions; what enzymes are responsible for deposition and erase those modifications? How is the information as histone modification transmitted to regulate transcription? Further mechanistic studies beyond genomic correlations will be needed to fully understand the nature of enhancer in relation to histone codes.

1.2.2 Transcription Coregulators

Upon TF binding to the chromatin, it recruits numbers of transcription coregulators for formation of enhancer complexes to orchestrate proper transcription responses. Functions of these coregulators include regulation of the chromatin environment, TF activity, and activities of coregulators each other as well as transcription machineries (Carey, 1998; Cosma, 2002). Mediator is a multicomponent coregulator complex that is evolutionally conserved from yeast to human and is required for virtually all transcription driven by RNA polymerase II (Pol II) in eukaryotes. For activator-dependent transcription, Mediator is

considered to bridge between TF and the general transcription factors (GTFs) and to stabilize Pol II to assist transcription activation (Kagey et al., 2010; Malik and Roeder, 2010; Sainsbury et al., 2015). Beside Mediator, a given TF requires numbers of additional coregulators to induce transcription (Carey, 1995; Lewis and Reinberg, 2003; Li et al., 2007; Li et al., 2016; Xu and Li, 2003). Although many coregulators are shared by various TFs, it requires systematic approaches to fully understand the commonality and specificity of each coregulator in a repertoire of components in enhancer complexes nucleated by each TF. In addition, our knowledge on the order of assembly and operation among coregulators in dynamic enhancer complex formation is limited. Effects of each coregulator to other coregulators within a enhancer complex leave another layer of complexity to fully understand the mechanisms and functions of transcription enhancer complexes.

1.2.3 Enhancer-promoter Chromatin Looping

Transcription enhancers function independent of its location and orientation respect to their target genes. In some cases, enhancers are located more than Mb away or on other chromosome from their target genes (Lam et al., 2014; Ong and Corces, 2011; Zhao et al., 2006). In some cases, enhancers control their target genes that are located beyond non-target genes, implying an active mechanism for target gene selection and specificity. The distally located enhancers are considered to communicate with their target genes through enhancer-promoter chromatin looping (Fullwood et al., 2009; Levine et al., 2014). Development of chromatin conformation capture technique (3C) and its derivative methods combined with deep sequencing (e.g. HiC, 4C, and ChIA-PET) enabled us to take a snapshot of global or

locus-specific enhancer-promoter chromatin looping (Bonev and Cavalli, 2016; Nagano et al., 2013; Simonis et al., 2006). A series of studies using these technologies provided the evidence of highly dynamic and specific nature of enhancer-promoter chromatin interactions. In addition, parallel studies have identified few proteins, including Mediator and CTCF-independent Cohesin, that are involved in enhancer-promoter chromatin looping formation (Apostolou et al., 2013; Kagey et al., 2010; Phillips-Cremins et al., 2013; Schmidt et al., 2010; Zuin et al., 2014). Furthermore, Lai *et al.* identified a non-coding transcript that is required for the neighboring gene expression, namely activating RNA or RNA-a, as a link between Mediator and enhancer-promoter chromatin looping formation (Lai et al., 2013). In their study, RNA-a was shown to interact with a Mediator component Med12, which in tern stimulates the kinase activity of the Mediator and induces chromatin interactions (Lai et al., 2013). All together, accumulative reports in the literature have provided interesting aspects of enhancer-promoter interactions. However, there are many unanswered questions still remained to be uncovered. What is the mechanisms of enhancer-promoter chromatin looping formation? Is enhancer-promoter communication unidirectional from the enhancer to the promoter or bidirectional between them? How is the target gene assigned to a given enhancer without affecting non-target gene expression in the same locus? For what extent does each enhancer contribute to the target gene expression when multiple enhancers control a given gene.

Collectively, successful enhancer complex formation eventually results in active transcription of the target genes. In addition, active transcription occurs at the vicinity of the TF binding sites, leading to the production of enhancer transcripts. The nature of sequence-

specific TF binding to DNA is well characterized by numerous studies involving structural, biochemical, biophysical, and genomic assays. In addition, number of coregulators have been identified and characterized in stabilized enhancer complexes with various TFs. However, there still are number of questions remain to be answered in relation to the dynamic mechanisms of enhancer complex formation involving coregulators, enhancer-enriched chromatin environment, enhancer-promoter chromatin looping, and enhancer transcripts.

1.3 Enhancer Transcription

1.3.1 Discoveries of Enhancer Transcripts

Few studies using loci-specific assays in early 2000s documented the recruitment of PNA polymerase II and GTFs not only at the activated promoter but also at the enhancer regions, where there was no annotated protein-coding or regulatory RNA genes (Johnson et al., 2001; Johnson et al., 2003; Spicuglia et al., 2002). Following studies provided evidence that those RNA polymerase at the enhancers are indeed engaged in active transcription to produce intergenic transcripts, which later named eRNAs (Ho et al., 2006; Kim et al., 2007). Finally, studies using genomic techniques uncovered genome-wide distribution of Pol II and active transcription in intergenic regions that overlap with intergenic enhancers (Brodsky et al., 2005; Carroll et al., 2006; De Santa et al., 2010 ; Kim et al., 2010). Together, these studies characterized eRNAs as unstable, non-polyadenylated and often bidirectionally

transcribed short non-coding transcripts (often < 1-2 kb) that arise from intergenic enhancers proceeding activation of the adjacent coding genes (Lai and Shiekhattar, 2014; Natoli and Andrau, 2012).

1.3.2 Transcription Regulation of eRNAs

Many enhancer features described above are also shared by active promoters. Both have relatively open chromatin states surrounded by well-positioned nucleosomes, enrichment of selective histone modifications (e.g. H3K4me2, H3K27ac, and H3K9ac) (Ernst and Kellis, 2012), enrichment of coregulators, GTFs, and RNA PolII, and active transcription (Andersson et al., 2015; Core et al., 2014; Kim and Shiekhattar, 2015). The correlation between the transcription levels at the enhancer and the target gene promoter with extensive enhancer-promoter chromatin looping brought up a question if and how those transcription events are coordinated (Hah et al., 2013).

Using transcription elongation inhibitors, Johnson *et al.* and we showed that the inhibition of enhancer transcription did not affect histone modifications or Pol II recruitment at the enhancers using β-globin and estrogen-regulated gene loci, respectively (Hah et al., 2013; Johnson et al., 2003). Johnson *et al.* further proposed that the role of the enhancer activated by p45/NF-E2 as a site of assembly for the transcription machinery, which in tern be translocated to the target gene promoter at the β-globin locus in erythroleukemia cells. Their model is developed from the observation where the deletion of p45/NF-E2 reduced histone acetylation and Pol II occupancies only at promoter without affecting their enrichment at the enhancer. Their results indicate the TF-independent Pol II recruitment at

the enhancer and the function of the TFs as a link between the Pol II loaded enhancer to the target gene promoter (Johnson et al., 2001). The Pol II translocation and tracking model from enhancers to the promoter was also supported by Wang *et al.* in androgen receptor-mediated gene transcription in androgen-sensitive prostate cancers using tiling ChIP-PCR assays (Wang et al., 2005). On the other hand, in the case of K^+-stimulated transcription in neuron, Kim *et al.* reported that the production of eRNA, but not Pol II recruitment to the enhancer, is dependent on the target gene promoter using the Arc prompter deleted mouse neuron (Kim et al., 2010). Theses studies highlighted the hierarchical recruitment of transcription machineries and transcription control at the enhancer and the target gene promoter. However, the complicated relationship between the Pol II recruitment and enhancer transcription, enhancer-promoter chromatin looping, and the target gene induction still remained unclear.

In contrast to the hierarchical or dependency models of the relationship between an enhancer and the target gene promoter, other studies proposed independency in transcription events at the two distinct genomic sites. For instance, Kim *et al.* introduced mutations at TATA box in the target gene promoter on an exogenous reporter DNA containing a DNase high sensitive enhancer (a.k.a. locus control region, LCR) and the target gene ε-globin to test the relationship between the enhancer and the target gene transcription. Using the reporter construct, they demonstrated that the disruption of TATA box at the target gene promoter reduces Pol II recruitment and transcription only at the target gene without affecting either histone modifications or transcription at the enhancer, suggesting that enhancer transcription as an intrinsic nature of the enhancer. In addition, they pointed out that the TF-dependent

increase of histone acetylation and transcription observed at the enhancer is not sufficient for induction of the gene transcription, leading to the econclusion that the enhancer transcription and the target gene transcription occur as independent entities (Kim et al., 2007).

Core *et al.* further supported the notion of transcription control at enhancers and the target gene promoters using genomic strategies. Based on ChIP-exo datasets, a modified chromatin immunoprecipitation method combined with exonuclease treatment to precisely map the location of DAN binding proteins at nucleotide resolution, they showed bimodal distribution of GTFs and Pol II within the nucleosome free regions (NFRs) at distally located enhancers similar to coding gene promoters, indicating enhancer-intrinsic transcription initiation independent of, but in similar mechanisms as compared to, the target gene promoters (Core et al., 2014). Using lipopolysaccharide (LPS)-stimulated macrophages, promoter-independent recruitment of GTFs and Pol II at promoter-proximal enhancer-like regions, that are bound by context-specific TFs and enriched for H3K4me1 and H3K27ac, was also demonstrated in another independent study by Scruggs *et al*. (Scruggs et al., 2015). Rather, the study by Core *et al.* and others uncovered the drastic differences in transcript stability between enhancer transcripts and mRNAs (Andersson et al., 2014; Andersson et al., 2015; Core et al., 2014; Kim and Shiekhattar, 2015; Lubas et al., 2015; Pefanis et al., 2015; Sigova et al., 2015).

Vast majority of studies on transcription regulation using molecular biology and genomic assays including those stated above were conducted with populations of cells or nuclei. While these techniques have their own advantages, use of bulk cells may cloud key observations to understand regulatory mechanisms and functions of enhancer transcription,

including whether enhancer transcription occurs in conjunction with the target gene promoter-driven transcription. For instance, the transcription at the two sites must occur in the same cells, presumably at the same allele, if the transcription control at enhancer and the target gene is coordinated or transcription of one of the two sites influences the other in *cis*. Rahman *et al.* addressed this issue using single molecular RNA FISH assays for estrogen-induced enhancer transcripts with the nascent gene transcripts. Interestingly, they reported that only less than 30% of cells expressing E2-induced gene transcripts also express eRNAs from the enhancer on either strands. These results indicate that transcription from the enhancer and the target gene promoters controlled independently. In addition, the transcription output for mRNAs in the cells that co-express eRNA and mRNA was at the equivalent level to that in the cells express the target gene transcripts without eRNA expression (Rahman et al., 2016). Since they tested only few pairs of a single enhancer to a single target gene, we cannot disregard the possibly of other enhancers controlling the target gene under the same condition. However, their study provides an important aspect for exploration of eRNA regulation and function. With recent rapid development of single-cell sequencing technologies and high-resolution and high-throughput imaging technologies (Boettiger et al., 2016; Buenrostro et al., 2015; Coleman et al., 2015; Cusanovich et al., 2015; Klein et al., 2015; Macosko et al., 2015; Nagano et al., 2013; Rotem et al., 2015), application of new experimental methods may lead us to better understanding of the functions and mechanisms of enhancer-dependent transcription.

1.3.3 Mechanisms and Functions of Enhancer Transcripts

Once eRNA production was accepted as a general feature of mammalian transcription using a number of genomic datasets in various biological contexts, the focus of the exploration moved to the functions and the mechanisms of enhancer transcripts. The correlation between TF binding, eRNA production, and the nearby gene induction observed in genomic assays suggested the functions of eRNAs as an activator for the target gene transcription (Franco et al., 2015; Hah et al., 2013; Kim et al., 2010). In deed, several studies demonstrated the functions on eRNAs in the target gene expression in a range of biological settings (Hsieh et al., 2014; Lam et al., 2013; Li et al., 2013; Melo et al., 2013; Mousavi et al., 2013; Schaukowitch et al., 2014; Sigova et al., 2015). However, the reported mechanisms vary in a great extent, with which conflict each other in some cases, including control of chromatin accessibilities, control of recruitment of TF or coregulator proteins, and induction of enhancer-promoter chromatin looping. Whether eRNAs as stable transcripts or the active transcription at the enhancers per se is critical for their functions was also debated to differentiate the modes of actions. It is also important to note that some hypothesized the enhancer transcription being noise or pervasive transcription (Koch et al., 2008; Natoli and Andrau, 2012). Few selected studies on functions and mechanisms of eRNAs are highlighted below.

Using siRNA-mediated knockdown of eRNAs, Mousavi *et al.* studied the function of myogenic transcription factor MyoD and MyoG-induced eRNAs in the myogenic target gene expression. In particular, they focused on two enhancers, the distal regulatory regions (DRRs) and the core enhancer (CE) upstream of MyoD gene. They reported that eRNA

arising from CE region acts in *cis* to regulate MyoD gene downstream of the two enhancers. Interestingly, on the other hand, eRNA transcribed from DRR is required to upregulate the expression of MyoG, which is located on another chromosome by functioning as a *trans* regulatory element, without affecting the expression of MyoD which is in the same locus with the two enhancers. In this case, eRNAs mediate the chromatin accessibility and Pol II recruitment at the genes that they act on regardless of the *cis* or *trans* mode of action (Mousavi et al., 2013). The molecular mechanisms that determine the mode of actions in *cis* or *trans* and the specificity of eRNAs to the target gene promoters, especially located in *trans*, might involve chromatin looping since interchromosomal interactions are rear but observed in other biological settings (Fullwood et al., 2009). It is also important to conduct follow-up studies to evaluate if the proposed mechanism is generalized as an eRNA function in broad biological contexts.

Others studies identified the role or eRNAs as a modulator of chromatin looping. Li *et al.* and Hsieh *et al.* investigated the role of estrogen and androgen-dependent eRNAs on estrogen and androgen-regulated gene induction. Under siRNA-mediated eRNA depletion, they showed impaired enhancer-promoter looping formation that results in the reduced hormone-dependent nearby gene expression. In these studies, eRNA functioned on the gene expression only in the same loci (Hsieh et al., 2014; Li et al., 2013). Along with these lines, Lai *et al.* exhibited the role of a non-codling RNA (a.k.a.RNA-a) in chromatin looping formation and expression of the neighboring gene by interacting and recruiting Mediator at the RAN-a transcribing loci and the target gene promoter (Lai et al., 2013). On the contrary, in other studies by Schaukowitch *et al.* and us, eRNA depletion using siRNA or transcription

inhibitor, respectively, did not present significant changes in enhancer-promoter chromatin looping or the recruitment of transcription machineries at the enhancers. (Hah et al., 2013; Schaukowitch et al., 2014). Comparisons of our previous study and the study by Li *et al.* are in particular interest as both use some overlapping genes and enhancers in the same cell line in similar conditions. One important difference in these two studies is that, while we depleted eRNAs using a transcription inhibitor, Li *et al.* used siRNA-mediated eRNA knockdown (Hah et al., 2013; Li et al., 2013). A possible explanation of the differences in the two findings may rise from a possibility that enhancer transcription as an act of transcription and eRNAs as a stable transcript may take two distinctive roles in some biological contexts such as estrogen-dependent gene expression.

Some suggested mechanisms of eRNA functions beyond the stage of transcription initiation. It has been well recognized that transcription is tightly controlled by multiple regulatory steps ranging from initiation, elongation, termination, as well as co-transcription RNA processing, which in tern regulates transcription. Thus, it may not be surprising for eRNAs to act on any of these regulatory steps. One study by Schaukowitch *et al.* using K^+-induced transcription responses in mouse neuron demonstrated that eRNAs decoys NELF-E, a component of negative elongation factor (NELF), from the target gene promoter to allow PolII to escape from promoter-proximal pausing to proceed productive transcription. In deed, NELF-E is also know to interact with RNA *in vivo* through its RNA recognition domain (RRD) (Pagano et al., 2014; Rao et al., 2006; Rao et al., 2008; Vos et al., 2016; Yamaguchi et al., 2002). Their study expanded the boundaries of eRNA functions in gene expression.

Despite that the majority of studies on eRNA functions have reported its function in the neighboring gene regulation (Lam et al., 2013; Li et al., 2013; Schaukowitch et al., 2014), at least one eRNA regulated by myogenic TF was shown to act in *trans (Mousavi et al., 2013)*. In either case, further studies are required to answer numbers of additional questions. If they have any function to control target gene transcription, how does an eRNA determine the specificity to the target gene? In addition, as exemplified above, the modes of action in each study are distinctive. We will come to see whether the mechanisms of eRNA function are highly context-specific or there is a unifying model that could apply to wide range of biological settings.

1.3.4 Enhancer Transcription Marks Active Enhancers

Although functions and mechanisms of enhancer transcripts are still in an active debate, our previous study and others show that the production of eRNAs differentiate active enhancers from unfunctional TF binding events (Hah et al., 2013; Lai and Shiekhattar, 2014; Shlyueva et al., 2014; Wu et al., 2014; Zhu et al., 2013). In our previous study, we examined the functional relationship between enhancer transcription and the target gene expression controlled by ERα enhancers. Using ChIP-seq datasets for ERα, coregulators, and histone marks combined with datasets of GRO-seq, global run-on and sequencing which captures snapshots of the genomic localization and orientation of actively transcribing RNA polymerases, we grouped ERα binding sites (ERBSs) based on the levels of enhancer transcription in an E2-treated ER-positive breast cancer cell line. To our surprise, we found that the only ~50% of ERBSs produce enhancer transcripts. Furthermore, the group of

ERBSs with eRNA production is significantly more enriched for other active enhancer marks including the coregulator recruitment, the H3K27ac levels, and the target gene induction compared to the group of ERBSs without eRNA production regardless of ERα binding intensities. Despite numerous studies using TF ChIP-seq and ChIP-on-chip describes TF binding sites as interchangeable to active enhancer, our results highlight that TF binding does not necessarily result in active transcription. Although we have limited understanding on the mechanisms on TF-depended enhancer transcription and the target gene expression, perhaps eRNAs are produced as the end product of enhancer activation, thus distinguish active enhancer compared to inactive TF binding sites (Hah et al., 2013). Furthermore, our study underscores the importance of steps that occurs after TF binding to enhancer transcription. Further studies are required to fully understand the functions of each steps, which eventually allow us to understand how enhancer transcription discriminate active enhancers from unfunctional TF binding sites.

With the recognition that eRNAs mark active enhancers and they often exhibit unique shot bidirectional transcription units, we also demonstrated in our previous study the use of eRNAs to annotate active enhancers without knowing the TFs that drive the transcription of the enhancer. Once active enhancers were identified based on enhancer transcripts, we determined the TFs that drive the transcription of eRNAs by performing motif search at those enhancers (Hah et al., 2013). Now several studies are underway to annotate active enhancers and key TFs using this principle in various biological systems where key TFs that control biological processes are unknown (Danko et al., 2015; Fang et al., 2014; Magnuson et al.,

2015; Nagari et al., 2017, Franco et al., unpublished; Zhu et al., 2013). As such, enhancer transcripts provide a vigorous means to annotate active enhancers.

1.4 Estrogen Receptor-dependent Enhancers

1.4.1 Estrogen Receptors

Estrogen signaling presents a broad range of physiological roles in various reproductive and non-reproductive tissues including regulation of estrous cycles, the balance between osteoclast and osteoblast, and metabolism (Burns and Korach, 2012; Simpson and Santen, 2015). Also, its pro-proliferative effect plays a mutagenic role in pathological conditions including hormone-dependent breast and uterus cancers. The vast majority of the effects of estrogen in the mammary grand and the uterus are mediated by nuclear signaling and gene regulation through evolutionally conserved hormone-dependent DNA binding TF, estrogen receptor alpha (ERα) (Couse et al., 1997; Nilsson et al., 2001). Once activated, the receptors bind on the chromatin to rapidly and transiently regulate the expression of thousands of genes to exert proper physiological reactions (Hah et al., 2011; Shao and Brown, 2004).

Estrogen receptors bind to various ligands in a relatively wide structural range that is not strictly limited to steroids (Nilsson et al., 2001). The structural variation of the ligands suggests considerable flexibility in its ligand biding domain (Figure 1.1) (Shiau et al., 1998). 17-β-estradiol (E2) is a naturally circulating estrogen in mammals with high potency to ERs

(Turner et al., 2007). In addition, numbers of synthetic ligands, namely selective estrogen receptor modulators (SERMs), bind to the receptors to modulate the receptor activities. Widely employed SERMs in the clinics include tamoxifen and raloxifene, which are partial agonist for ERα, and a full agonist fulvestrant (Katzenellenbogen and Katzenellenbogen, 2000). It is important to note that the term "selective" in SERMs represent the context-specific effects of these compounds. The determinant factor of their context-specific effects are still not clear, however, it is speculated that the combination of coregulators that interact with ERs in context-specific manners at least in part paly a role in excreting the tissue-specific outcomes of SERMs (Katzenellenbogen and Katzenellenbogen, 2002).

ERα adopts a classical structure of steroid hormone nuclear receptors (Figure 1.1). The N-terminal domain contain activation domain 1 (AF-1), which exhibits weak ligand-independent transactivation function. Following to the N-terminal domain, ERα harbors DNA binding domain (DBD) and hinge domain. DBD of ERα with two zinc-finger binding structures adopts the domain structure suitable for binding to sequence-specific double-stranded DNA (Nilsson et al., 2001). The C-terminal of ERα consists of the ligand binding domain (LBD) and the main activation domain, AF-2, which is secluded within LBD. LBD is composed of 12 α-helices with two β-sheets connected by linker coils. 11 α-helices (H1-11) together form a hydrophobic cliff to provide a ligand binding pocket, and the two β-sheets sit on one side of the ligand binding pocket. Full agonistic ligand such as E2 binding to this domain causes structural changes where the last α-helix H12 shifts over to cover the ligand binding pocket on the opposite site from the two β-sheets. The positional shift of H12 then creates another hydrophobic cliff consist of H3-5 and H12, which serves as an

interaction surfaces for coregulators. On the other hand, partial agonist such as tamoxifen or raloxifene displace the H12 from the position adjacent to the ligand binding pocket, disrupting the canonical hydrophobic surfaces required for coregulator interactions (Brzozowski et al., 1997; Shiau et al., 1998).

Upon ligand binding, ERα homo-dimerizes and bind to chromatin across numerous genomic locations. ERα binding sties are often predetermined by other chromatin-binding proteins including FoxA1 and AP2γ (Jozwik and Carroll, 2012). Also, like enhancers controlled by other TFs, the pre-programmed ERα binding sites have other characteristics including relatively open chromatin states and enhancer-enriched histone marks such as H3K4me1 and H3K27ac (Calo and Wysocka, 2013; Hon et al., 2013). Chromatin bound ERα in turn recruits various coregulators to regulate the chromatin environment, ERα activity, and the activity of coregulators each other. Formation of the active enhancer complex stimulates establishment of additional active enhancer features including the elevation of the H3K27ac level, enhancer-promoter chromatin looping formation, and the active transcription at the enhancer as well as the target genes (Shao and Brown, 2004) (Figure 1.2).

1.4.2 Estrogen Receptor Coregulators

For ERα to execute the proper gene activation in response to estrogen, they require a number of transcription coregualators (Lonard and O'Malley, 2005; Nilsson et al., 2001). Mediator and SRCs are the two major coregulators for ERα, both of which directly interact with ERα through α-helix 12 in the main transactivation domain AF-2 of the LBD (Kang et

al., 2002; Malik and Roeder, 2005, 2010; Shiau et al., 1998). Mediator complex, interacting with ERα through its subunit Med1, is considered to bridge ERα to the general transcription factors to stabilize their interaction with Pol II and to assist transcription initiation (Chen and Roeder, 2011). On the other hand, SRCs serve as scaffold proteins to recruit many additional coregulators to ERα binding sites including bromodomain-containing lysine acetyltranscferase p300/CBP (Gojis et al., 2010; Paul et al., 2007; Yi et al., 2015).

p300/CBP is required for transcription activation controlled by various TFs by acetylating histones, TFs, and other coregulators, and thus widely used as a mark of active enhancers (Visel et al., 2009). p300/CBP are highly homologous proteins and considered to paly almost, but not exactly, redundant roles in many biological contexts (Goodman and Smolik, 2000; Vo and Goodman, 2001). The mechanisms of their recruitment vary depending on transcription factors that they work with, but it has shown to form a stable complex with ERα through SRCs (Dancy and Cole, 2015; Kim et al., 2001; Yi et al., 2015). In addition, a recent study reported that CBP interacts with eRNAs to stimulate the catalytic activity *in vitro*, expanding its role in enhancer activation (Bose et al., 2017). Furthermore, aberrant functions of p300/CBP have been implicated to play important roles in several types of cancers. Indeed, p300 is one of the most commonly mutated genes across cancer types including breast cancers (Chang et al., 2016; Farria et al., 2015; Kandoth et al., 2013).

Lysine acetylation that catalyzed by various acetyltransferases including p300/CBP performs critical roles in transcription activation. Negatively charged acetyl group neutralize the positive charge on lysine, which may affect the structure of the acetylated protein or the interactions with negatively charged molecules such as DNA. Besides influencing charges

on amino acids, acetylated lysine serves as an interaction surface for bromodomains contained in 46 distinct genes in human genome (Filippakopoulos and Knapp, 2012). BET family is a bromodomain-containing protein family that consists of four family members, BRD2, BRD3, BRD4, and BRDt (Belkina and Denis, 2012; Filippakopoulos and Knapp, 2012). During transcription activation, BRDs are recruited to acetylated histones at promoters through their bromodomains. They in tern recruit pTEFb, a positive elongation factor, which then phosphorylates the negative elongation factor to allow paused polymerases to proceed to productive transcription (Filippakopoulos and Knapp, 2012; Kanno et al., 2014; Wang and Filippakopoulos, 2015; Zeng and Zhou, 2002). Using an estrogen-treated breast cancer cell line, Nagarajan *et al.* reported an important role of BRD4 in E2-depednet gene expression (Nagarajan et al., 2014). However, the mechanisms and functions of the BRD family members at ERα enhancer complex are still not fully understood.

ERα binding to chromatin nucleates *de novo* formation of the active enhancer complex, involving coregulator recruitment and establishment of unique chromatin environments (Lonard and O'Malley, 2005). A series of studies highlighted coregulator functions on E2-dependent enhancer formation and transcription activation by observing stably formed enhancer complexes (Voegel et al., 1998; Yang et al., 2000; Yi et al., 2015). However, our knowledge on the precise molecular mechanisms of the kinetics of the assembly and functions of ERα enhancer complex formation are yet to be fully elucidated.

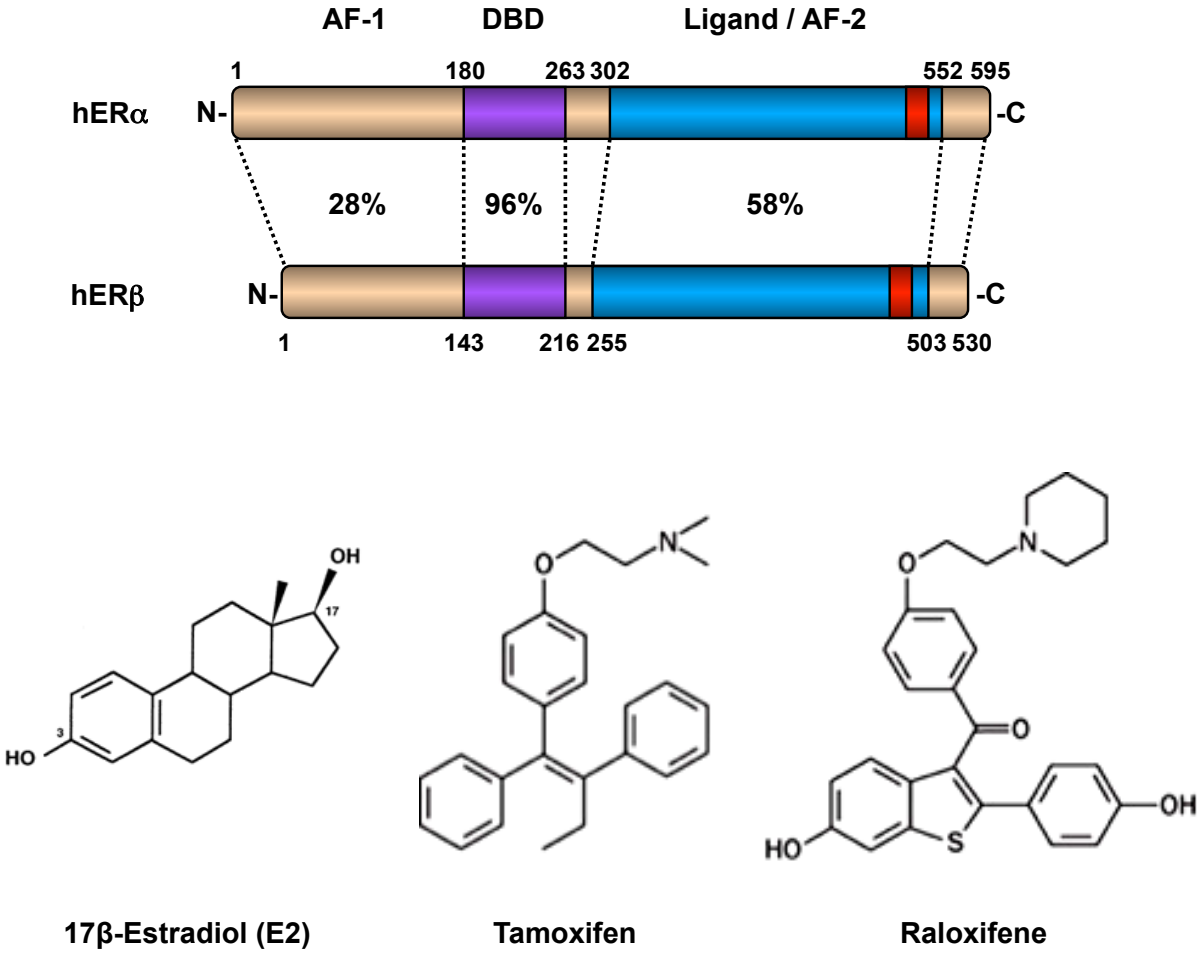


Figure 1.1. Domain structures of estrogen receptors.
Domain structures of ERα and ERβ (top) and examples of their ligands (bottom). Both receptors consist of the activation domain 1 (AF-1), DNA binding domain (DBD), followed by the ligand binding domain (LBD), which contains the second and predominant transactivation domain 2 (AF-2). The key α-helix, H12 in the AF-2 is indicated as a red box.

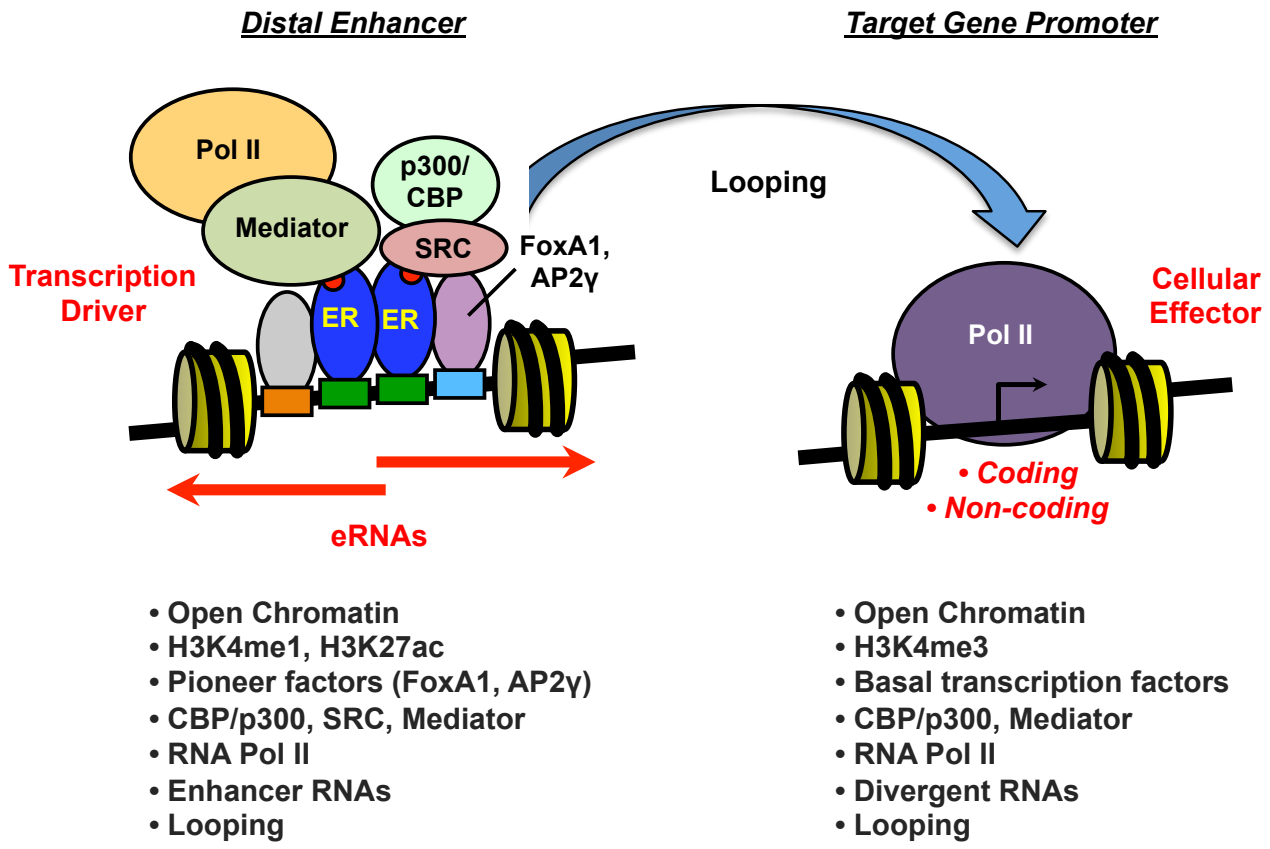


Figure 1.2. ERα enhancers are associated with active enhancer features.
Upon activation, ERα binds to thousands of sites across the genome. These sites are often bound by pioneering factors such as FoxA1 and AP2γ prior to ERα binding. The chromatin bound ERα recruits various coregulators and RNA polymerases. Enhancer activation leads to enhancer RNA (eRNA) production and enhancer-promoter chromatin interaction, that collectively regulate the target gene transcription.

1.5 References

Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. Nat Commun *5*, 5336.

Andersson, R., Sandelin, A., and Danko, C.G. (2015). A unified architecture of transcriptional regulatory elements. Trends Genet *31*, 426-433.

Apostolou, E., Ferrari, F., Walsh, R.M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., Stuart, H.T., Polo, J.M., Ohsumi, T.K., Borowsky, M.L., *et al.* (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. Cell Stem Cell *12*, 699-712.

Belkina, A.C., and Denis, G.V. (2012). BET domain co-regulators in obesity, inflammation and cancer. Nat Rev Cancer *12*, 465-477.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature *529*, 418-422.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat Rev Genet *17*, 661-678.

Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell 168, 135-149 e122.

Brodsky, A.S., Meyer, C.A., Swinburne, I.A., Hall, G., Keenan, B.J., Liu, X.S., Fox, E.A., and Silver, P.A. (2005). Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. Genome Biol *6*, R64.

Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.A., and Carlquist, M. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. Nature *389*, 753-758.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486-490.

Burns, K.A., and Korach, K.S. (2012). Estrogen receptors and human disease: an update. Arch Toxicol *86*, 1491-1504.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

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

Carey, M.F. (1995). Transcriptional activation. A holistic view of the complex. Curr Biol *5*, 1003-1005.

Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F., *et al.* (2006). Genome-wide analysis of estrogen receptor binding sites. Nat Genet *38*, 1289-1297.

Chang, M.T., Asthana, S., Gao, S.P., Lee, B.H., Chapman, J.S., Kandoth, C., Gao, J., Socci, N.D., Solit, D.B., Olshen, A.B., *et al.* (2016). Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol *34*, 155-163.

Chen, W., and Roeder, R.G. (2011). Mediator-dependent nuclear receptor function. Semin Cell Dev Biol *22*, 749-758.

Coleman, R.A., Liu, Z., Darzacq, X., Tjian, R., Singer, R.H., and Lionnet, T. (2015). Imaging Transcription: Past, Present, and Future. Cold Spring Harb Symp Quant Biol *80*, 1-8.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311-1320.

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

Couse, J.F., Lindzey, J., Grandien, K., Gustafsson, J.A., and Korach, K.S. (1997). Tissue distribution and quantitative analysis of estrogen receptor-alpha (ERalpha) and estrogen receptor-beta (ERbeta) messenger ribonucleic acid in the wild-type and ERalpha-knockout mouse. Endocrinology *138*, 4613-4621.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910-914.

Dancy, B.M., and Cole, P.A. (2015). Protein lysine acetylation by p300/CBP. Chem Rev *115*, 2419-2452.

Danko, C.G., Hyland, S.L., Core, L.J., Martins, A.L., Waters, C.T., Lee, H.W., Cheung, V.G., Kraus, W.L., Lis, J.T., and Siepel, A. (2015). Identification of active transcriptional regulatory elements from GRO-seq data. Nat Methods *12*, 433-438.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.

Fang, B., Everett, L.J., Jager, J., Briggs, E., Armour, S.M., Feng, D., Roy, A., Gerhart-Hines, Z., Sun, Z., and Lazar, M.A. (2014). Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. Cell *159*, 1140-1152.

Farria, A., Li, W., and Dent, S.Y. (2015). KATs in cancer: functions and therapies. Oncogene *34*, 4901-4913.

Filippakopoulos, P., and Knapp, S. (2012). The bromodomain interaction module. FEBS Lett *586*, 2692-2704.

Franco, H.L., Nagari, A., and Kraus, W.L. (2015). TNFalpha signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. Mol Cell *58*, 21-34.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H.*, et al.* (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58-64.

Gaffney, D.J., McVicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J.K. (2012). Controls of nucleosome positioning in the human genome. PLoS Genet *8*, e1003036.

Gojis, O., Rudraraju, B., Gudi, M., Hogben, K., Sousha, S., Coombes, R.C., Cleator, S., and Palmieri, C. (2010). The role of SRC-3 in human breast cancer. Nat Rev Clin Oncol *7*, 83-89.

Goodman, R.H., and Smolik, S. (2000). CBP/p300 in cell growth, transformation, and development. Genes Dev *14*, 1553-1577.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res *23*, 1210-1223.

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M.*, et al.* (2010). Nucleosome dynamics define transcriptional enhancers. Nat Genet *42*, 343-347.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Ho, Y., Elefant, F., Liebhaber, S.A., and Cooke, N.E. (2006). Locus control region transcription plays an active role in long-range gene activation. Mol Cell *23*, 365-375.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet *45*, 1198-1206.

Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S.*, et al.* (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc Natl Acad Sci U S A *111*, 7319-7324.

Johnson, K.D., Christensen, H.M., Zhao, B., and Bresnick, E.H. (2001). Distinct mechanisms control RNA polymerase II recruitment to a tissue-specific locus control region and a downstream promoter. Mol Cell *8*, 465-471.

Johnson, K.D., Grass, J.A., Park, C., Im, H., Choi, K., and Bresnick, E.H. (2003). Highly restricted localization of RNA polymerase II within a locus control region of a tissue-specific chromatin domain. Mol Cell Biol *23*, 6484-6493.

Jozwik, K.M., and Carroll, J.S. (2012). Pioneer factors in hormone-dependent cancers. Nat Rev Cancer *12*, 381-385.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A.*, et al.* (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333-339.

Kang, Y.K., Guermah, M., Yuan, C.X., and Roeder, R.G. (2002). The TRAP/Mediator coactivator complex interacts directly with estrogen receptors alpha and beta through the TRAP220 subunit and directly enhances estrogen receptor function in vitro. Proc Natl Acad Sci U S A *99*, 2642-2647.

Kanno, T., Kanno, Y., LeRoy, G., Campos, E., Sun, H.W., Brooks, S.R., Vahedi, G., Heightman, T.D., Garcia, B.A., Reinberg, D.*, et al.* (2014). BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. Nat Struct Mol Biol *21*, 1047-1057.

Katzenellenbogen, B.S., and Katzenellenbogen, J.A. (2000). Estrogen receptor transcription and transactivation: Estrogen receptor alpha and estrogen receptor beta: regulation by selective estrogen receptor modulators and importance in breast cancer. Breast Cancer Res *2*, 335-344.

Katzenellenbogen, B.S., and Katzenellenbogen, J.A. (2002). Biomedicine. Defining the "S" in SERMs. Science *295*, 2380-2381.

Kim, A., Zhao, H., Ifrim, I., and Dean, A. (2007). Beta-globin intergenic transcription and histone acetylation dependent on an enhancer. Mol Cell Biol *27*, 2980-2986.

Kim, M.Y., Hsiao, S.J., and Kraus, W.L. (2001). A role for coactivators and histone acetylation in estrogen receptor alpha-mediated transcription initiation. EMBO J *20*, 6084-6094.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, T.K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. Cell *162*, 948-959.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Koch, F., Jourquin, F., Ferrier, P., and Andrau, J.C. (2008). Genome-wide RNA polymerase II: not genes only! Trends Biochem Sci *33*, 265-273.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497-501.

Lai, F., and Shiekhattar, R. (2014). Enhancer RNAs: the new molecules of transcription. Curr Opin Genet Dev *25*, 38-42.

Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M.*, et al.* (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature *498*, 511-515.

Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci *39*, 170-182.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237-1251.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13-25.

Lewis, B.A., and Reinberg, D. (2003). The mediator coactivator complex: functional and physical roles in transcriptional regulation. J Cell Sci *116*, 3667-3675.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X.*, et al.* (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516-520.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet *17*, 207-223.

Lonard, D.M., and O'Malley, B.W. (2005). Expanding functional diversity of the coactivators. Trends in biochemical sciences *30*, 126-132.

Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. Cell Rep *10*, 178-192.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M.*, et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Magnuson, B., Veloso, A., Kirkconnell, K.S., de Andrade Lima, L.C., Paulsen, M.T., Ljungman, E.A., Bedi, K., Prasad, J., Wilson, T.E., and Ljungman, M. (2015). Identifying transcription start sites and active enhancer elements using BruUV-seq. Sci Rep *5*, 17978.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Mol Cell *62*, 63-78.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends in biochemical sciences *30*, 256-263.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet *11*, 761-772.

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Leveille, N., Kalluri, R.*, et al.* (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell *49*, 524-535.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell *51*, 606-617.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59-64.

Nagarajan, S., Hossan, T., Alawi, M., Najafova, Z., Indenbirken, D., Bedi, U., Taipaleenmaki, H., Ben-Batalla, I., Scheller, M., Loges, S.*, et al.* (2014). Bromodomain protein BRD4 is required for estrogen receptor-dependent enhancer activation and gene transcription. Cell Rep *8*, 460-469.

Nagari, A., Murakami, S., Malladi, V.S., and Kraus, W.L. (2017). Computational Approaches for Mining GRO-Seq Data to Identify and Characterize Active Enhancers. Methods Mol Biol *1468*, 121-138.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annu Rev Genet *46*, 1-19.

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

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283-293.

Pagano, J.M., Kwak, H., Waters, C.T., Sprouse, R.O., White, B.S., Ozer, A., Szeto, K., Shalloway, D., Craighead, H.G., and Lis, J.T. (2014). Defining NELF-E RNA binding in HIV-1 and promoter-proximal pause regions. PLoS Genet *10*, e1004090.

Paul, B.D., Buchholz, D.R., Fu, L., and Shi, Y.B. (2007). SRC-p300 coactivator complex is required for thyroid hormone-induced amphibian metamorphosis. J Biol Chem *282*, 7472-7481.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P.*, et al.* (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell *161*, 774-789.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y.*, et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281-1295.

Pradeepa, M.M., Grimes, G.R., Kumar, Y., Olley, G., Taylor, G.C., Schneider, R., and Bickmore, W.A. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. Nat Genet *48*, 681-686.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Rahman, S., Zorca, C.E., Traboulsi, T., Noutahi, E., Krause, M.R., Mader, S., and Zenklusen, D. (2016). Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. Nucleic Acids Res.

Rao, J.N., Neumann, L., Wenzel, S., Schweimer, K., Rosch, P., and Wohrl, B.M. (2006). Structural studies on the RNA-recognition motif of NELF E, a cellular negative transcription elongation factor involved in the regulation of HIV transcription. Biochem J *400*, 449-456.

Rao, J.N., Schweimer, K., Wenzel, S., Wohrl, B.M., and Rosch, P. (2008). NELF-E RRM undergoes major structural changes in flexible protein regions on target RNA binding. Biochemistry *47*, 3756-3761.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotechnol *33*, 1165-1172.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 129-143.

Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell *56*, 29-42.

Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., and Odom, D.T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res *20*, 578-588.

Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Mol Cell *58*, 1101-1112.

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

Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., and Greene, G.L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell *95*, 927-937.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978-981.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet *38*, 1348-1354.

Simpson, E., and Santen, R.J. (2015). Celebrating 75 years of oestradiol. J Mol Endocrinol *55*, T1-20.

Spicuglia, S., Kumar, S., Yeh, J.H., Vachez, E., Chasson, L., Gorbatch, S., Cautres, J., and Ferrier, P. (2002). Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. Mol Cell *10*, 1479-1487.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat Struct Mol Biol *20*, 267-273.

Taylor, G.C., Eskeland, R., Hekimoglu-Balkan, B., Pradeepa, M.M., and Bickmore, W.A. (2013). H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res *23*, 2053-2065.

Turner, J.V., Agatonovic-Kustrin, S., and Glass, B.D. (2007). Molecular aspects of phytoestrogen selective binding at estrogen receptors. J Pharm Sci *96*, 1879-1885.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Vo, N., and Goodman, R.H. (2001). CREB-binding protein and p300 in transcriptional regulation. J Biol Chem *276*, 13505-13508.

Voegel, J.J., Heine, M.J., Tini, M., Vivat, V., Chambon, P., and Gronemeyer, H. (1998). The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. EMBO J *17*, 507-519.

Vos, S.M., Pollmann, D., Caizzi, L., Hofmann, K.B., Rombaut, P., Zimniak, T., Herzog, F., and Cramer, P. (2016). Architecture and RNA binding of the human negative elongation factor. Elife *5*.

Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet *15*, 69-81.

Wang, C.Y., and Filippakopoulos, P. (2015). Beating the odds: BETs in disease. Trends Biochem Sci *40*, 468-479.

Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol Cell *19*, 631-642.

Wu, H., Nord, A.S., Akiyama, J.A., Shoukry, M., Afzal, V., Rubin, E.M., Pennacchio, L.A., and Visel, A. (2014). Tissue-specific RNA expression marks distant-acting developmental enhancers. PLoS Genet *10*, e1004610.

Xu, J., and Li, Q. (2003). Review of the in vivo functions of the p160 steroid receptor coactivator family. Mol Endocrinol *17*, 1681-1692.

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., and Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol *22*, 2918-2927.

Yang, W., Rachez, C., and Freedman, L.P. (2000). Discrete roles for peroxisome proliferator-activated receptor gamma and retinoid X receptor in recruiting nuclear receptor coactivators. Mol Cell Biol *20*, 8008-8017.

Yi, P., Wang, Z., Feng, Q., Pintilie, G.D., Foulds, C.E., Lanz, R.B., Ludtke, S.J., Schmid, M.F., Chiu, W., and O'Malley, B.W. (2015). Structure of a biologically active estrogen receptor-coactivator complex on DNA. Mol Cell *57*, 1047-1058.

Zeng, L., and Zhou, M.M. (2002). Bromodomain: an acetyl-lysine binding domain. FEBS Lett *513*, 124-128.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U.*, et al.* (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet *38*, 1341-1347.

Zhu, Y., Sun, L., Chen, Z., Whitaker, J.W., Wang, T., and Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. Nucleic Acids Res *41*, 10032-10043.

Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F*., et al.* (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A *111*, 996-1001.

CHAPTER TWO

Enhancer Transcripts Mark Active Estrogen Receptor Binding Sites

The study described in this chapter is selected sections from the prior publication in Hah *et al.*, Genome Research, 2013. This chapter highlights the study with the sections with my contributions. The study was conducted in collaboration with Nasun Hah, Ph.D., Anusha Nagari, M.S., and Charles Dnako, Ph.D.. N.H. initiated the study, performed ChIP-qPCR, and prepared the manuscript. I designed and performed molecular biology assays except ChIP-qPCR. A.N. performed bioinformatics analysis on the GRO-seq dataset prepared by N.H. for previous publications and other genomic datasets obtained from public domains. C.D. performed preliminary genomic analysis.

2.1 Summary

We have integrated and analyzed a large number of data sets from a variety of genomic assays (e.g., GRO-seq, ChIP-seq, and DNase-seq) using a novel computational pipeline to provide a comprehensive and global view of estrogen receptor alpha (ERα) enhancers in MCF-7 human breast cancer cells. Using this approach, we have defined a class of primary transcripts (eRNAs) that are transcribed uni- or bidirectionally from ERα binding sites (ERBSs) with an average transcription unit length of ~3 to 5 kb. The majority is upregulated by short treatments with estradiol (i.e., 10, 25, or 40 min.) with kinetics that generally precede or match the induction of the target genes. The production of eRNAs at ERBSs is strongly correlated with the enrichment of a number of genomic features that have been shown to be associated with enhancers (e.g., H3K4me1, p300/CBP, and an open chromatin architecture). In the absence of eRNA production, strong enrichment of these features is not observed, even though ERα binding is evident. We find that flavopiridol, a cdk9 inhibitor that blocks transcription elongation, inhibits eRNA production, but does not affect other molecular indicators of enhancer activity (e.g., RNA pol II binding, H3K4me1 levels, enhancer looping). These results indicate that the assembly of enhancer complexes can be dissociated from eRNA production, suggesting that eRNA production occurs after the assembly of active enhancers. Together, our studies have shed new light on the activity of ERα at its enhancer sites and provide new insights about enhancer function in general.

2.2 Introduction

The steroid hormone estrogen plays critical roles in a variety of developmental and physiological processes, as well as many disease states (Couse and Korach, 1999; Deroo and Korach, 2006). The actions of estrogen are mediated through two estrogen receptor (ER) proteins, ERα and ERβ, which display distinct, tissue-specific biological functions (Couse and Korach, 1999; Deroo and Korach, 2006; Nilsson et al., 2001). ERs function as ligand-dependent nuclear transcription factors that bind to cis-acting DNA regulatory elements (i.e., enhancers) to regulate gene expression programs upon binding 17β-estradiol (E2), the predominant naturally occurring ligand (Nilsson et al., 2001). The catalog of ERα binding sites ("cistrome") has been mapped across the genome in a number of cell lines and mouse tissues (reviewed in (Cheung and Kraus, 2010)), with estimates ranging from ~10,000 to ~30,000 sites (depending on the methods and cutoffs used) in the ERα-positive MCF-7 human breast cancer cell line, a commonly used model of estrogen action (Cheung and Kraus, 2010; Hurtado et al., 2011; Lin et al., 2007; Welboren et al., 2009). The majority of these ERα binding sites are located in genomic regions distal from transcription start sites (TSSs) (Carroll et al., 2006; Lin et al., 2007; Welboren et al., 2009). Although these studies have been provided many new insights about ERα action at enhancers across the genome, many key questions about the functions and mechanisms of action of ERα enhancers remain.

Enhancers were first characterized as regulatory elements that (1) carry sequence information for transcription factor binding, (2) are generally located far from TSSs, (3) regulate gene expression regardless of location and orientation, and (4) play key roles in

controlling tissue-specific gene expression (Bulger and Groudine, 2011; Ong and Corces, 2011). Current models posit that enhancers function by promoting communication with target gene promoters through chromatin loops or by tracking of enhancer-bound transcription factors through intervening chromatin to target gene promoters (Bulger and Groudine, 2011; Kolovos et al., 2012; Ong and Corces, 2011). Recent studies have focused intense interest on the properties of enhancers, beyond the binding of sequence-specific transcription factors, which might give clues to their mechanisms of action and aid in their identification. In this regard, histone modifications (e.g., H3 lysine 4 monomethyl, H3K4me1; H3 lysine 27 acetyl, H3K27ac), histone variants (e.g., H2A.Z), coactivators (e.g., p300, CBP, Mediator), and an open chromatin architecture (e.g., DNase I hypersensitivity) have been identified as genomic features that mark or identify enhancers (Melgar et al., 2011; Natoli and Andrau, 2012). Differential association of these features with enhancers in a given cell may define distinct classes of enhancers that specify distinct gene regulatory mechanisms and biological outcomes (Creyghton et al., 2010; Ghisletti et al., 2010; Pham et al., 2012; Rada-Iglesias et al., 2011; Shen et al., 2012; Wang et al., 2011; Whyte et al., 2012; Zentner et al., 2011). Enhancer profiles may even provide useful clinical signatures for cancer diagnosis and prognosis (Akhtar-Zaidi et al., 2012; Ross-Innes et al., 2012).

More recently, a number of studies have shown that many enhancers overlap with sites of RNA pol II binding, active RNA pol II transcription, and the production of enhancer RNAs ("eRNAs") (De Santa et al., 2010; Djebali et al., 2012; Hah et al., 2011; Kim et al., 2010; Wang et al., 2011). A common signature of enhancer transcription is the production of short (i.e., ~1 to 2 kb) eRNAs that are transcribed bidirectionally (Kim et al., 2010). We and

others have recently shown that the genomic binding sites for ERα and other steroid receptors overlap with sites of transcription (Hah et al., 2011; Wang et al., 2011). The role of transcription in enhancer function is unknown, but the act of transcription may help to create an open chromatin environment that promotes enhancer function (Natoli and Andrau, 2012). Alternatively, the stable accumulation of eRNAs may play a functional, perhaps even structural, role and may facilitate gene looping (Natoli and Andrau, 2012; Orom et al., 2010; Orom and Shiekhattar, 2011).

eRNA production, as well as the other enhancer features described above, has been used in a number of studies to identify or predict enhancers on a genome-wide basis (Fernandez and Miranda-Saavedra, 2012; Maston et al., 2012; Melgar et al., 2011; Pennacchio et al., 2007; Visel et al., 2009; Won et al., 2008). In spite of our ability to identify enhancers based on chromatin features, detect steady-state eRNA production, and monitor gene looping events, we still know little about the molecular mechanisms of enhancer function. In addition, we lack a comprehensive and integrated view of enhancer chromatin features, eRNA production, and gene looping events in a signal-regulated gene regulatory system. This information is needed to understand the molecular mechanisms of action of signal-regulated enhancers and their cell type-specific functions.

In the studies described herein, we used Global Run-on Sequencing (GRO-seq), a method that assays the location and orientation of all active RNA polymerases genome-wide (Core et al., 2008), to generate a global profile of active transcription at ERα binding sites (ERBSs) in MCF-7 human breast cancer cells in response to a short time course of E2 treatment. GRO-seq allows the detection of active or ongoing transcription at enhancers, in

contrast to the detection of steady state accumulation of eRNAs by RNA-seq. We integrated the data from our GRO-seq assays with data from a variety of other genomic assays (e.g., ChIP-seq and DNase-seq) using a novel computational pipeline to provide a comprehensive and global view of ERα enhancers and their regulation by E2 in MCF-7 cells. In addition, we used a series of locus-specific molecular assays to reveal new insights about the molecular mechanisms of ERα enhancer function. Together, our studies have shed new light on the activity of ERα at its enhancer sites and provide new insights about enhancer function in general, including the potential roles of enhancer transcription.

2.3 Results

ERα Enhancers Are Sites of Estrogen-induced Transcription

In a previous study using GRO-seq to characterize the estrogen-regulated transcriptome in MCF-7 cells, we identified hundreds of transcribed regions in the genome generating primary transcripts that overlap estrogen receptor alpha (ERα) binding sites (ERBSs) (Hah et al., 2011). In this paper, we have undertaken a comprehensive identification and analysis of ERα enhancer transcription in MCF-7 human breast cancer cells by integrating a wide array of genomic data sets with locus-specific molecular analyses. Multiple examples of transcribed ERα enhancers located upstream of estrogen-regulated target genes are shown in browser track representation in Fig. 2.1A and Fig. 2.2. These include an ERBS, which we refer to as ERBS1, located ~20 kb upstream of the promoter of

the *P2RY2* gene, as well as these additional enhancer/gene pairs: ERBS2/*GREB1*, ERBS3/*SBNO2*, ERBS4/*SMAD7*, and ERBS5/*PGR*. As shown in the GRO-seq browser tracks in Fig. 2.1A, transcription of the *P2RY2* gene and a region around ERBS1 is upregulated rapidly in a short time course of treatment with 17β-estradiol. The transcripts from ERBS1 (Fig. 2.1A), as well as ERBSs 2 through 5 (Fig. 2.2), are produced bidirectionally from both strands of DNA, reminiscent of the enhancer RNAs (eRNAs) described previously (Kim et al., 2010), and the transcribed regions are associated with RNA pol II and previously identified transcription start sites (TSSs) (Yamashita et al., 2011).

As expected, these ERBSs are also associated with previously characterized enhancer features, including the pioneer transcription factor FoxA1, histone H3 lysine 4 monomethylation (H3K4me1), the histone acetyltransferases p300 and CBP, p160 steroid receptor coactivator proteins (SRC1, 2, and 3) (Fig. 2.1A). Like the ERBS eRNAs, many of these enhancer features are induced by treatment with E2. These ERBSs are also involved in chromatin looping events that promote physical interactions with their target genes, as defined by chromatin interaction analysis by paired-end tag sequencing (ChIA-PET; (Fullwood et al., 2009)) (Fig. 2.1A). Locus-specific molecular assays, including ChIP-qPCR, RT-qPCR, and 3C-PCR, confirm the localization of RNA pol II and H3K4me1 at these ERBSs (Fig. 2.1, B and C; Figs. 2.3A and 2.3B), as well as the steady-state production of the eRNAs (Fig. 2.1D; Fig. 2.3C; and Fig. 2.5, A through E) and enhancer looping events (Fig. 2.1E; Fig. 2.3D) as the genes are induced by E2. Interestingly, an ERBS not associated with eRNA production lacks many of the enhancer features described above (Fig. 2.2; Fig. 2.7).

Collectively, these genomic and locus-specific analyses illustrate quite clearly the range of features associated with ERα enhancers, including the production of eRNAs.

Global Identification and Characterization of Estrogen-regulated ERα Enhancer Transcripts

To obtain a global view of ERα enhancer transcripts, we developed a computational pipeline that allowed us to identify eRNAs overlapping with intergenic ERBSs (Fig. 2.4A). Starting from a set of all ERBSs (~10,000) defined previously (Welboren et al., 2009), we narrowed the list to those that are intergenic (i.e., >10 kb away from the beginning or end of an annotated RefSeq gene; ~3000). We then separated them in to those that have (~1,500) and those that do not have (~1,500) an overlapping transcript, as defined by GRO-seq. Next, we classified those with overlapping transcripts based on the production of transcripts from both strands of DNA ("Paired"; 715) or from one strand of DNA ("Unpaired"; 882). Finally, we applied a length filter for the transcription unit/primary transcript, defining those <9 kb as "short" and those ≥9 kb as "long". The average length of the primary transcripts in each category, as well as the average length of the overlap for the paired transcripts, are shown schematically in Fig. 2.4B. The position and orientation of the transcripts relative to the ERBSs are shown schematically in Fig. 2.4C. For the purposes of the remaining studies shown herein, we focused on two classes of enhancer transcripts: (1) short unpaired (S-U) and (2) short-short paired (S-S), which we call eRNAs. Many of the transcripts in the remaining classes (i.e., long unpaired, long-short paired, and long-long paired) are likely to

represent long non-coding RNAs and their associated antisense RNAs. They are beyond the scope of the current study and are not considered further herein.

Production of the short unpaired and short-short paired ERα enhancer transcripts is regulated by E2 over a short time course of treatment (0, 10, 40, and 160 min.) (Fig. 2.4D). About 70 percent of the transcripts are E2 upregulated, with maximum effects for most upregulated transcripts occurring at 40 min. and for most downregulated transcripts occurring at 160 min. (Fig. 2.4D). The upregulation is evident in metagenes of the GRO-seq data (Fig. 2.2E) and corresponds to the levels of RNA pol II at the ERBS, as expected (Fig. 2.4, F and G).

The Production of eRNAs from ERBSs Positively Correlates with a Wide Variety of Enhancer Properties

To better understand how the production of eRNAs from ERBSs may relate to enhancer function, we mined a large number of existing genomic data sets from MCF-7 cells (see Materials and Methods, Genomic Datasets). Although all three classes of ERBSs that we examined (i.e., those with S-S paired eRNAs, short unpaired eRNAs, and no eRNAs) have similar mean and median levels of ERα binding by ChIP-seq (Fig. 2.5A), considerable differences were observed among these groups with respect to other enhancer properties. For example, ERBSs producing short bidirectional eRNAs (i.e., S-S paired transcripts) have considerably higher mean and median levels of pioneer factors (e.g., FoxA1, AP2γ; (Carroll et al., 2005; Hurtado et al., 2011; Tan et al., 2011)), ERα coregulators (e.g., CBP and SRC3), enhancer histone modifications (e.g., H3K4me1), as well as the most accessible chromatin

structures (defined by DNase-seq), than ERBS producing no transcripts (Fig. 2.5). Together, these data indicate that the production of eRNAs at ERBS correlates with properties that are generally associated with active enhancers.

Inhibition of eRNA Production by Flavopiridol Does Not Inhibit Enhancer Complex Assembly or Looping to Target Gene Promoters

Our results have shown that the production of eRNAs from ERBSs correlates with many indicators of "active" enhancers (e.g., RNA pol II, coregulators, H3K4me1, looping to target gene promoters), but the precise role of eRNAs in enhancer function are unknown. The act of enhancer transcription may serve to promote the formation of an open chromatin structure required for enhancer function. Alternatively, stably accumulated eRNAs may play a structural role that promotes enhancer loop formation. To address the role of E2-induced eRNAs in ERα enhancer function, we used the small molecule drug flavopiridol (FP), an inhibitor of the cdk9 kinase of the P-TEFb complex (Chao et al., 2000), to block the production and stable, steady-state accumulation of eRNAs originating from ERBSs. MCF-7 cells were pretreated with FP for 1 hour prior to treatment with E2 for the times indicated in the figures. In locus-specific assays using RT-qPCR, FP efficiently blocked the production and accumulation of eRNAs produced from all five of the ERBSs that we examined, as well as the production and accumulation of mRNAs from the E2-regulated target genes controlled by these enhancers (Fig. 2.6, B and D; Fig. 2.7). This experimental system gave us the opportunity to examine the assembly of ERα enhancer complexes in the absence of eRNAs.

Treatment of MCF-7 cells with FP did not affect the E2-dependent binding of ERα or RNA pol II at the ERBSs (Fig. 2.6, A and C), which occurred normally in the presence of the drug. Likewise, treatment with FP did not dramatically affect the recruitment of coregulators to the ERBSs (Fig. 2.6, E and F). Thus, although the production of eRNAs correlates well with markers of active enhancers, the production and stable accumulation of eRNAs are not required for the assembly of ERα enhancer complexes at ERBSs. Furthermore, treatment with FP did not affect E2-dependent looping between ERBSs and E2-regulated target genes (Fig. 2.6, G and H), indicating that the production and stable accumulation of eRNAs are not required enhancer looping, at least under the conditions that we tested herein.

2.4 Discussion

In the studies described herein, we integrated and analyzed a large number of genomic data sets using a novel computational pipeline to provide a comprehensive and global view of ERα enhancers in the MCF-7 human breast cancer cell line. The data sets that we analyzed included (1) GRO-seq from a short time course of E2 treatment, which allowed us to monitor active transcription at ERBSs, (2) ChIP-seq data ± E2, which allowed us to monitor histone modifications (e.g., H3K4 me1 and me3), the binding of pioneer factor (FoxA1) and the binding of coregulators (e.g. CBP and SRC3), and (3) DNase-seq ± E2, which allowed us to monitor the chromatin state at ERBS. From our analyses, we have made a number of observations that shed new light on the activity of ERα at its enhancer sites and

provide new insights about enhancer function in general. These aspects of our studies are discussed below.

ERBS Are Actively Transcribed in an Estrogen-dependent Manner

The results from our GRO-seq analysis indicate that about half of all intergenic ERBSs overlap an actively transcribed region (Fig. 2.4A). By our definition, this excludes transcribed annotated protein coding genes, but may include transcribed lncRNAs and their associated antisense RNAs. By applying a length filter, as described in Fig. 2.4B, we have defined a class of primary (i.e., unprocessed) ERα enhancer transcripts that are similar to the eRNAs described previously (Kim et al., 2010) and have the following properties. First, they originate from one or both strands of DNA, with an average transcription unit (i.e., primary transcript) length of ~3 to 5 kb and, for the bidirectional transcripts, an overlap of ~3 kb (Fig. 2.2A). Second, the majority are upregulated by short treatments with E2 (i.e., 10, 25, or 40 min.) (Fig. 2.4, D and E) with kinetics that, in many cases, precede or match the induction of the target gene (Fig. 2.1D; Fig. 2.3C; Fig. 2.7). Third, steady-state products of the enhancer transcripts are detectable by RT-qPCR using either random hexamer or oligo(dT) primers for RT (Fig. 2.1D; Fig. 2.3C; Fig. 2.7). The signals obtained from the oligo(dT) primers suggest that these transcripts may be polyadenylated, but per perhaps minimally so since they do not give strong signals in poly(A) RNA-seq data sets from MCF-7 cells (data not shown). Finally, many are 5' 7-methylguanosine capped based on TSS-seq (Yamashita et al., 2011) and locus-specific capping assays (data not shown). These results, which are consistent with previous reports of enhancer transcription and eRNAs (De Santa et al., 2010; Djebali et al.,

2012; Hah et al., 2011; Kim et al., 2010; Wang et al., 2011), provide new information by characterizing the active transcription of enhancers and, for specific loci, relating it to the steady-state levels of eRNAs.

About half of the intergenic ERBSs that overlap an actively transcribed region are associated with the production of short eRNA transcripts with the aforementioned properties. Interestingly, half of ERBSs do not overlap an actively transcribed region. Our hypothesis, which we tested in a number of different ways herein, is that the production or absence of eRNAs may functionally distinguish different classes of ERBSs. In this regard, we note that although all ERBSs are enriched for ERE motifs, those associated with eRNAs show a different enrichment of transcription factor motifs than those not associated with eRNAs. Distinctions between ERBSs with and without eRNAs are discussed in more detail below.

eRNA Production Positively Correlates with Features of Active Enhancers

The enrichment of a number of genomic features have been proposed to be marks of active enhancers, including H3K4me1, H3K27ac, p300/CBP, RNA pol II, and an open chromatin architecture (reviewed in (Maston et al., 2012; Natoli and Andrau, 2012)). Our results indicate that the production of eRNAs at ERBSs strongly correlates with the enrichment of these features (Fig. 2.5). In the absence of eRNA production, strong enrichment of these features is not observed, even though ERα binding is evident. In addition, we found that the E2-dependent production of eRNAs at ERBSs also strongly correlates with the binding of the p160 steroid receptor coregulators SRC3 (Fig. 2.5D). SRC proteins interact with ERα in an E2-dependent manner through a hydrophobic cleft on the

ligand binding domain of the receptor (Heery et al., 1997; Torchia et al., 1997). In addition, SRC proteins interact directly with p300 and CBP, thus allowing them to promote the indirect binding of p300 and CBP to DNA-bound ERα (Kim et al., 2001; Torchia et al., 1997). Thus, E2-dependent formation of the DNA-ERα-SRC-p300/CBP complex is likely to represent an initial step in the formation of an active enhancer at ERBSs.

Collectively, our results demonstrate that active transcription at ERBSs, as determined by GRO-seq, tracks with genomic features thought to be marks of active enhancers. From these results, we can begin to understand the order of events that lead to the assembly of an active enhancer complex at ERBSs. The process is initiated by E2-bound ERα, which binds to direct (i.e., ERE-mediated) or indirect (i.e., tethered) sites across the genome and then nucleates the formation of enhancer complexes containing coregulators, some of which function as histone-modifying enzymes, as well as RNA pol II and perhaps looping factors, such as Mediator and cohesin. Histone modification, such as H3K4me1, looping, and eRNA production then follows.

The Assembly of Enhancer Complexes can be Dissociated from eRNA Production

The function of enhancer transcription and the stable, steady-state accumulation of eRNAs are unknown. Some have suggested that the active of transcription helps to create an open chromatin environment that promote enhancer function, while others have suggested that the stable accumulation of eRNAs may play a functional, perhaps even structural, role (Bulger and Groudine, 2011; Maston et al., 2012; Natoli and Andrau, 2012; Orom and Shiekhattar, 2011). These are challenging questions that require continued study. Two

aspects of our results have shed some light on these questions, as well as the order of operations at signal-regulated enhancers. First, using the drug FP, we showed that many features of enhancers, including the assembly of enhancer complexes and the modification of histones, can be dissociated from eRNA production. FP efficiently blocks enhancer transcription and the stable, steady-state accumulation of eRNAs at ERBSs, but had no effect on any of the E2-dependent enhancer features that we examined, including enhancer-promoter looping (Fig. 2.6; Fig. 2.7). Together, these results clearly show that the assembly of enhancer complexes can be dissociated from eRNA production, suggesting that eRNA production occurs after the assembly of active enhancers. These results, however, do not suggest that eRNA production is unnecessary for enhancer function or target gene activation. In fact, FP also inhibits target gene activation (Fig. 2.6, B and D; Fig. 2.7), so the potential role of eRNA production in that aspect of enhancer function could not be assayed in our studies. Further studies will be required to resolve these issues.

2.5 Methods

Cell Culture and Treatments

MCF-7 human breast adenocarcinoma cells were kindly provided by Dr. Benita Katzenellenbogen (University of Illinois, Urbana-Champaign). The cells were maintained in minimal essential medium (MEM) supplemented with Hank's salts (Sigma) and 5% calf serum. The cells were plated for experiments in phenol red-free MEM (Sigma)

supplemented with 5% charcoal-dextran treated calf serum (CDCS) for at least three days prior to hormone or drug treatment. As indicated for the different experiments, the cells were treated with 100 nM E2 for the times specified. For the transcription inhibition experiments, the cells were pretreated with or without 1 μM flavopiridol (Sigma) for 1 hour prior to treatment with E2.

Antibodies

The antibodies used for chromatin immunoprecipitation (ChIP) assays are as follows: ERα (rabbit polyclonal generated in Kraus lab); Pol II (sc-899), SRC2 (sc-343), SRC3 (sc-9119), p300 (sc-585), CBP (sc-369), and non-immune IgG (rabbit polyclonal from Santa Cruz Biotech); H3K4me1 (ab8895), H3K4me3 (ab8580), H3K27ac (ab4729), and H3 (ab1791) (rabbit polyclonal from Abcam).

Chromatin Immunoprecipitation-quantitative PCR (ChIP-qPCR)

MCF-7 cells were grown in estrogen free medium to ~80% confluence and then treated with 100 nM E2 for the indicated time, with or without 1 μM Flavopiridol (Sigma) for 1 h prior to E2 treatment. ChIP analyses were conducted as described previously (Kininis et al., 2007), with a few modifications. Treated MCF-7 cells were crosslinked with 1% formaldehyde for 10 min. at 37°C, followed by quenching with 125 mM glycine for 5 min. at 4°C. The crosslinked cells were washed with PBS, harvested, lysed with lysis buffer [Tris•HCl (pH 7.9), 0.5% SDS, 10 mM EDTA, 1 mM DTT, 1x protease inhibitor cocktail (Roche)] and subjected to sonication for 7 cycles of 20 seconds each at setting high using a

Diagenode Bioruptor to obtain ~500 bp DNA fragments. The lysate was incubated with the antibodies indicated along with a rabbit IgG control after input material was removed, followed by incubation with protein A-agarose beads for 1.5 hours. The immunoprecipitates were collected and washed with wash buffer [20 mM Tris•HCl (pH 7.9), 0.25% NP-40, 0.05% SDS, 2 mM EDTA, 250 mM NaCl, 1x protease inhibitor cocktail (Roche)] at 4°C and eluted by incubating overnight in elution buffer [100 mM $NaHCO_3$, 1 % SDS] at 65°C to reverse the crosslinks, followed by digestion with proteinase K. The ChIP'ed DNA was subjected to phenol:chloroform extraction and analyzed by qPCR using gene-specific primers as below and a 384-well real-time PCR thermocycler with SYBR Green detection. Each experiment was performed a minimum of three times with independent biological samples to ensure reproducibility.

Gene	**Name**	**Sequence (5' to 3')**
• *P2RY2*	ERBS1-Fwd	CCATCAAAGCTGTTGCTTCT
	ERBS1-Rev	CCAGGATAGTGCCAGTGAAC
• *GREB1*	ERBS2-Fwd	TAGGCTTCAAGAGGACCACA
	ERBS2-Rev	AGCAGCAAAACTGCATAGGA
	ERBS3-Rev	TCCCAGTCCATCTATCCTCA

Analysis of eRNAs and mRNAs by Reverse Transcription-quantitative PCR (RT-qPCR)

MCF-7 cells were grown in estrogen free medium to ~80% confluence and then treated with 100 nM E2 for the indicated time, with or without 1 μM Flavopiridol (Sigma)

for 1 h prior to E2 treatment. RT-qPCR detection of eRNAs and mRNAs were performed as described previously (Sun et al.), with some minor modifications. Total RNA was isolated from the treated cells using TRIzol reagent (Invitrogen) and subjected to RT using random hexamers and M-MLV Reverse transcriptase (Promega). The cDNA was then subjected to qPCR analysis using a Roche LightCycler 480 system with SYBR Green detection and gene-specific primers asfollows. Each experiment was performed a minimum of three times with independent biological samples to ensure reproducibility.

Gene/ERBS	**Name**	**Sequence (5' to 3')**
• *P2RY2*/ERBS1	eRNA-Fwd	AGGCAAATCCATTGTCATCC
	eRNA-Rev	AACTGGCTGGATCTTGAAGC
	mRNA-Fwd	CGGTGGACTTAGCTCTGAGG
	mRNA-Rev	GCCTCCAGATGGGTCTATGA
• *GREB1*/ERBS2	eRNA-Fwd	GGGAATAGAGCCCTGAGCTT
	eRNA-Rev	TTGATCTGCTCTTGCCTGAA
	mRNA-Fwd	CCTATTTTGGAATAAAAACTGACC
	mRNA-Rev	GGGGAGAATGACACAAAAGC
• *SBNO2*/ERBS3	eRNA-Fwd	CCTGTATTCTGGGGGCACTA
	eRNA-Rev	CTCACCCCATCCAGTACACC
	mRNA-Fwd	GACTGGGCACCCACAAGGGC
	mRNA-Rev	GGAAGGGCTGGGGGAGGGAG
• *SMAD7*/ERBS4	eRNA-Fwd	GGCATAGCTAGGACCTCACC

	eRNA-Rev	GAGGGAGGAAAGTGGCTTCT
	mRNA-Fwd	AAGAGAAGCATTCTCATTGGAAA
	mRNA-Rev	TCAGGAGTCCTTTCTCTCTCAAA
• *PGR*/ERBS5	eRNA –Fwd	ATGCAGAGCCATTGCAAAAT
	eRNA –Rev	ATCAGCAAGATGCAAACACG
	mRNA-Fwd	TTGCCAAGAAGGTGAAACTG
	mRNA-Rev	CTTTGCATTGTCACCCCATC
• ERBS6	ERBS6-Fwd	TGTGGAAGCTGCATTCTTTG
	ERBS6-Rev	TCAGAACCATGCAGAACCTG

Chromosome Conformation Capture (3C)

Chromosome conformation capture was conducted as previously described (Pan et al., 2008), with the following modifications.

Treatment and Crosslinking of Cells. MCF-7 cells were grown in estrogen free medium to ~80% confluence and then treated with 100 nM E2 for 40 min., with or without 1 μM Flavopiridol (Sigma) for 1 h prior to E2 treatment. They were then fixed with 1% formaldehyde for 10 min. After quenching with 200 mM glycine for 5 min, the cells were lysed by douncing in lysis buffer [10 mM Tris•HCl (pH 7.5), 10 mM NaCl, 0.2% Triton-X, 1x protease inhibitor cocktail (Roche)] and incubating with gentle mixing at 4°C for 30 min. The nuclei were collected by centrifugation.

DNA Digestion. The isolated nuclei were dispersed in 1.2x restriction enzyme buffer supplied by the manufacturer (NEB) with 0.3% SDS, followed by gentle mixing at 37°C for

one hour. Triton-X was then added at 2% final concentration and incubated with gentle mixing at 37°C for one hour. The DNA was digested with gentle mixing at 37°C overnight with 400 U of restriction enzyme (NEB). The enzymes used were *BglII* for *P2RY2* and *BtgI* for *GREB1*.

DNA Ligation. The digestion reaction was terminated by the addition of 1.6% SDS with incubation at 65°C for 20 min. with gentle mixing. The digested nuclei were then transferred to a 50 mL conical tube containing 6.125 mL ligation buffer [1.15x T4 DNA ligase buffer (NEB) and 1% Triton-X100] and incubated at 37°C for one hour with gentle mixing. Ligation was performed by the addition of 2000 U of T4 DNA Ligase (NEB) with incubation at 16°C for 4 hours, followed by incubation at 25°C for 30 min. As a control, similar samples were incubated without ligase. After digestion with Proteinase K, the DNA was de-crosslinked incubating at 65°C overnight and then purified using phenol-chloroform extraction, followed by ethanol precipitation.

Loop Detection. Nested PCR to detect chromatin interactions was performed using Taq DNA polymerase (NEB). Each primer set follows was designed unidirectionally upstream of the restriction enzyme digestion sites such that PCR amplifies any DNA resulting from ligation between the hub and a test site. Digested and ligated bacterial artificial chromosome (BAC) DNA spanning the entire locus for each gene analyzed was used as a PCR control. Each experiment was performed three times with independent biological samples to ensure reproducibility.

GREB1

Name		Outer Primer Sequence (5' to 3')	Inner Primer Sequence (5' to 3')
L	Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	AGACCCCTTACACAGTCA	GACATGTCTTTGATGTTTTC
a	Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	GTTCAAGCAGTCCGAGTA	AGGTGATCTGCCTATCTCT
b	Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	CATGATTTGTTTTATCTTCC	GGAATTGTTCATCTTCTTTC
c	Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	ACCTGAACCTTCTAAGTAGC	CACAGCCAGTTAATTTTTAT
d	Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	AACAGTAGATGCTCTGTGAG	CGAGTAGCTGGGATTACA
e	Fwd	ACTCCATTCTTACTCCAGTT	GAGAATGTTTGACACTGCTA
	Rev	GACTCATTTGAGGTTCGT	GAATCTTCCTTTTCCTCTC

P2RY2

Name		Outer Primer Sequence (5' to 3')	Inner Primer Sequence (5' to 3')
L	Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	AGCAAATGTTTACTCAGAAG	GGAGATGCTTATGTGGTG
a	Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	AGGACAGTTAAGCCTCTG	GGTAGAAAGGGTCAGTCA
b	Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC

	Rev	CAGAAATGTTGTGAGAACTAA	ACATACACAGAGTGCTGTTC
c	Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	CTGGTTTACCAACAATGATA	ATAGCAACCAGAACAGAGA
d	Fwd	GCAGGAGGATTTCAAGTA	AGCAACAAGAGGTAGAGC
	Rev	GACTAAGCTCCAGAGTGTTT	CTCCACCTCCCTTATCTAC

Analysis of GRO-seq Data

GRO-seq data were analyzed using software described previously (Hah et al., 2011) and the approaches described below. Software, scripts, and other information can be obtained by contacting W. Lee Kraus.

Read Alignment. GRO-seq reads were aligned to human reference genome (hg18), including autosomes, X chromosome, and one complete copy of an rDNA repeat (GenBank ID: U13369.1). The SOAP.2.21 software package (Li et al., 2009) was used to align the reads using the following parameters: (1) all n mappings were removed (-r 0); (2) three mismatches were allowed in each mapped read (-v 3); (3) low-quality reads with more than 10 ambiguous bases were removed (-N 10); and (4) for reads failing to align over the entire length of the read, the first 32 bp was used (-l 32).

Transcript Calling. Transcript calling was performed using a two-state hidden Markov model using the GRO-seq data analysis package described in Hah *et al.* (2011). We used a shape setting parameter of 5 and -log transition probability of 200 to predict the transcription units. The predicted transcripts were assigned into six classes as described in

Hah et al. (2011) using annotations from RefSeq, ENSEMBL, and UCSC Known Gene databases.

Defining Classes of ERα Enhancer Transcripts (eRNAs). The repertoire of genomic ERα binding sites (ERBSs) was extracted from ChIP-seq data provided in Welboren *et al.*, (2009) (GEO accession number GSM365926). Those ERBSs >10 kb away from the 5' or 3' ends of annotated genes were defined as "Intergenic ERBSs." They were divided into three classes based on the presence, location, and orientation of GRO-seq-defined transcripts: (1) those overlapping transcripts originating from both strands of DNA, running in opposite directions as a divergent pair; (2) those overlapping a transcript originating from one strand of DNA only ("Unpaired); and (3) those not overlapping a transcript. The transcripts in classes 1 and 2 were further categorized based on the length of the transcript unit/primary transcript as 'short' (length < 9 kb; eRNAs, by our definition) and 'long' (length >9 kb; which likely represent other classes of non-coding RNAs, such as lncRNAs).

Transcript Maps. Individual eRNAs were visualized in the transcript map shown in Fig. 2.2C at genomic positions relative to the associated intergenic ERBS using custom PERL scripts and R. The transcript maps were centered on the ERBSs and the relative positions of each eRNA with respect to the corresponding ERBS was plotted (shorter and longer eRNAs in the pair: blue and red, respectively; unpaired eRNA: red). The transcript maps were ordered based on the length of the shorter eRNA in the pair or on the length of the unpaired eRNA.

Determining Estrogen Regulation of Transcripts and Generating Heat Maps. The effects of E2 treatment on the expression of the eRNAs were analyzed using edgeR

(Robinson et al., 2010), as described previously (Hah et al., 2011). We used the 10 min., 40 min., and 160 min. E2 treatment time points (two biological replicates for each time point) to determine the E2-dependent regulation of eRNAs. The results were plotted as a heat map using Java TreeView (Saldanha, 2004), ordered based on the magnitude of expression at the 40 min. time point.

Analysis of ChIP-seq Data

ChIP-Seq datasets from MCF-7 cells were obtained from the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and the ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) on-line databases. The data sets are listed in Supplemental Table S2. The raw files were aligned to hg18 using BOWTIE (Langmead et al., 2009). Uniquely mappable reads were converted into bigWig files using BEDTools (Quinlan and Hall, 2010) for visualization in the UCSC genome browser.

Genomic Data Analysis and Visualization

We used the following approaches to summarize and visualize genomic data from GRO-seq, ChIP-seq, and ChIA-PET.

Metagenes. Metagenes were used to illustrate the distribution of GRO-seq, ChIP-seq, and ChIA-PET reads around ERα peak maxima (or other genomic features) using the metagene function in our GRO-seq package (Hah et al., 2011).

Boxplots. Boxplot representations were used to minimize the bias caused by outliers in the data, which can overly influence metagene representations. The boxplots allowed for

accurate comparisons across ERBSs with S-S paired eRNAs, S-unpaired eRNAs, and without eRNA. The read distribution in a 2 kb window (± 1 kb) around the ERBSs was calculated and plotted using the boxplot function in R. The inter-quartile regions (IQRs) of the boxplots were used to plot metagenes centered on the ERBSs. All the metagenes and boxplots were scaled to a library size of 15 million reads to normalize against different read densities.

2.6 Genomic Datasets

ChIP-seq

• Pol II (-E2, +E2)	GSM365929, GSM365930
• ERα (-E2, +E2)	GSM365925, GSM365926
• FOXA1 (-E2, +E2)	GSM588929, GSM588930
• CBP (-E2, +E2)	ERR045723, ERR045724
• SRC3 (-E2, +E2)	ERR045716, ERR045715, ERR045719, ERR045720
• H3K4me1 (-E2, +E2)	GSM588569, GSM588568
• H3K4me3 (-E2, +E2)	GSM588571, GSM588570
• ChIA-PET (+E2)	ChIA-PET visualization browser (http://cms1.gis.a-star.edu.sg)*
• DNase1 (-E2, +E2)	GSM822389, GSM822390
• TSS	SRA003625 (GenBank/DDBJ)*

GRO-seq

• MCF-7 (-E2, +E2)	SE43836

2.7 References

Akhtar-Zaidi, B., Cowper-Sal-lari, R., Corradin, O., Saiakhova, A., Bartels, C.F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M.F., *et al.* (2012). Epigenomic enhancer profiling defines a signature of colon cancer. Science *336*, 736-739.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327-339.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., *et al.* (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33-43.

Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F., *et al.* (2006). Genome-wide analysis of estrogen receptor binding sites. Nat Genet *38*, 1289-1297.

Chao, S.H., Fujinaga, K., Marion, J.E., Taube, R., Sausville, E.A., Senderowicz, A.M., Peterlin, B.M., and Price, D.H. (2000). Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. J Biol Chem *275*, 28345-28348.

Cheung, E., and Kraus, W.L. (2010). Genomic analyses of hormone signaling and gene regulation. Annu Rev Physiol *72*, 191-218.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

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

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

Deroo, B.J., and Korach, K.S. (2006). Estrogen receptors and human disease. J Clin Invest *116*, 561-570.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., *et al.* (2012). Landscape of transcription in human cells. Nature *489*, 101-108.

Fernandez, M., and Miranda-Saavedra, D. (2012). Genome-wide enhancer prediction from epigenetic signatures using genetic algorithm-optimized support vector machines. Nucleic Acids Res *40*, e77.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H.*, et al.* (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58-64.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L.*, et al.* (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity *32*, 317-328.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Heery, D.M., Kalkhoven, E., Hoare, S., and Parker, M.G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature *387*, 733-736.

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

Kim, M.Y., Hsiao, S.J., and Kraus, W.L. (2001). A role for coactivators and histone acetylation in estrogen receptor alpha-mediated transcription initiation. EMBO J *20*, 6084-6094.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kininis, M., Chen, B.S., Diehl, A.G., Isaacs, G.D., Zhang, T., Siepel, A.C., Clark, A.G., and Kraus, W.L. (2007). Genomic analyses of transcription factor binding, histone acetylation, and gene expression reveal mechanistically distinct classes of estrogen-regulated promoters. Mol Cell Biol *27*, 5090-5104.

Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R., and Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin *5*, 1.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics *25*, 1966-1967.

Lin, C.Y., Vega, V.B., Thomsen, J.S., Zhang, T., Kong, S.L., Xie, M., Chiu, K.P., Lipovich, L., Barnett, D.H., Stossi, F.*, et al.* (2007). Whole-genome cartography of estrogen receptor alpha binding sites. PLoS Genet *3*, e87.

Maston, G.A., Landt, S.G., Snyder, M., and Green, M.R. (2012). Characterization of enhancer function from genome-wide analyses. Annu Rev Genomics Hum Genet.

Melgar, M.F., Collins, F.S., and Sethupathy, P. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. Genome Biol *12*, R113.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annu Rev Genet.

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

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283-293.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Orom, U.A., and Shiekhattar, R. (2011). Noncoding RNAs and enhancers: complications of a long-distance relationship. Trends Genet *27*, 433-439.

Pan, Y.F., Wansa, K.D., Liu, M.H., Zhao, B., Hong, S.Z., Tan, P.Y., Lim, K.S., Bourque, G., Liu, E.T., and Cheung, E. (2008). Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements. J Biol Chem *283*, 32977-32988.

Pennacchio, L.A., Loots, G.G., Nobrega, M.A., and Ovcharenko, I. (2007). Predicting tissue-specific enhancers in the human genome. Genome Res *17*, 201-211.

Pham, T.H., Benner, C., Lichtinger, M., Schwarzfischer, L., Hu, Y., Andreesen, R., Chen, W., and Rehli, M. (2012). Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. Blood *119*, e161-171.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

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

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Tan, S.K., Lin, Z.H., Chang, C.W., Varang, V., Chng, K.R., Pan, Y.F., Yong, E.L., Sung, W.K., and Cheung, E. (2011). AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. EMBO J *30*, 2569-2581.

Torchia, J., Rose, D.W., Inostroza, J., Kamei, Y., Westin, S., Glass, C.K., and Rosenfeld, M.G. (1997). The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. Nature *387*, 677-684.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A.*, et al.* (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390-394.

Welboren, W.J., van Driel, M.A., Janssen-Megens, E.M., van Heeringen, S.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J *28*, 1418-1428.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221-225.

Won, K.J., Chepelev, I., Ren, B., and Wang, W. (2008). Prediction of regulatory elements in mammalian genomes using chromatin signatures. BMC Bioinformatics *9*, 547.

Yamashita, R., Sathira, N.P., Kanai, A., Tanimoto, K., Arauchi, T., Tanaka, Y., Hashimoto, S., Sugano, S., Nakai, K., and Suzuki, Y. (2011). Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis. Genome Res *21*, 775-789.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res *21*, 1273-1283.

Figure 2.1. The ERα enhancer of the estrogen-responsive P2RY2 gene produces bi-directional transcripts in MCF-7 cells.
(A) Browser tracks of GRO-seq, ChIP-seq (Pol II, ERα, FoxA1, and H3K4me1), ChIA-PET, TSS locations, and gene annotation for *P2RY2* and its distal ERα binding sites (ERBS1). The data are from MCF-7 cells treated with a time course of E2 (GRO-seq) or a single time point of E2 (45 or 60 min.). Previously identified TSSs based on a previously published data set from MCF-7 cells (Yamashita et al. 2011) are located as indicated. The locations of primers used for 3C assays are indicated by orange arrows. The black bars shown for the ChIA-PET data indicate the "head" and "tail" making contact in the gene loops, which are indicated by the dotted black lines. Scale bars show the length of the indicated region.
(B and C) ChIP-qPCR analyses showing recruitment of ERα and Pol II (B) or levels of H3K4me1, me3, and H3 (C) at ERBS1 in response to a time course of E2 treatment. Each bar represents the mean + the SEM for three or more independent biological replicates.
(**D)** RT-qPCR analyses showing the expression of ERBS1 eRNA and *P2RY2* mRNA in response to a time course of E2 treatment. Each bar represents the mean + the SEM for three or more independent biological replicates.
(E) 3C-PCR assay showing E2-induced looping between ERBS1 and the *P2RY2* gene. The lower case letters correspond to the primers denoted by orange arrows shown in panel A. The assays were conducted in the presence (experimental) or absence (control) of DNA ligase, as indicated. Digested and ligated bacterial artificial chromosome (BAC) DNA spanning the entire *P2YR2* locus was used as a PCR control. The size of the PCR fragments in bp is shown. One representative experiment from three conducted is shown.

[Figure 2.1 is on the next page]

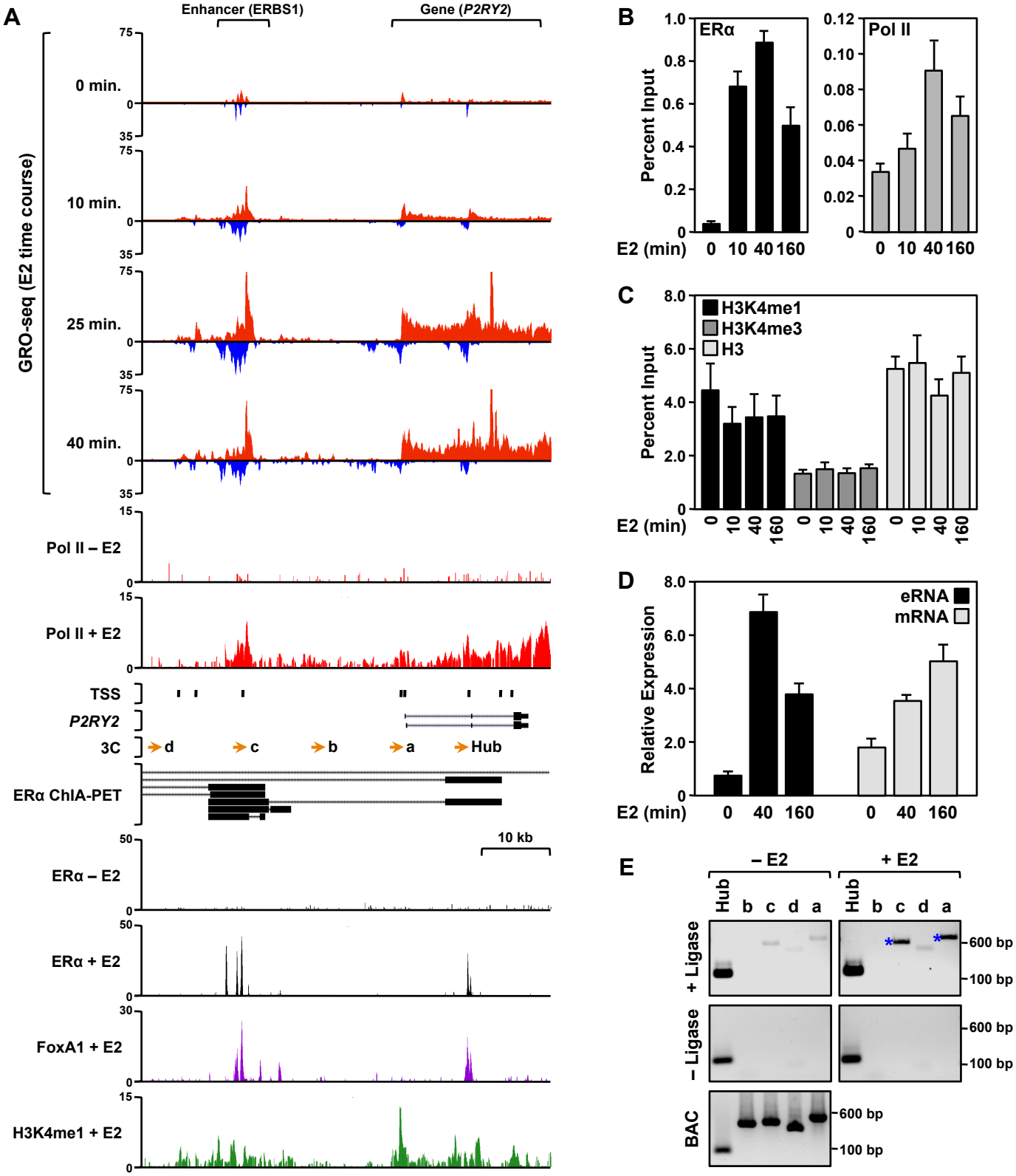

A
Enhancer (ERBS1)
Gene (P2RY2)
GRO-seq (E2 time course)
0 min.
10 min.
25 min.
40 min.
Pol II – E2
Pol II + E2
TSS
P2RY2
3C
d
c
b
a
Hub
ERα ChIA-PET
10 kb
ERα – E2
ERα + E2
FoxA1 + E2
H3K4me1 + E2
B
ERα
Pol II
Percent Input
E2 (min)
0 10 40 160
C
H3K4me1
H3K4me3
H3
Percent Input
E2 (min)
D
eRNA
mRNA
Relative Expression
E2 (min)
0 40 160 0 40 160
E
– E2
+ E2
Hub b c d a Hub b c d a
+ Ligase
– Ligase
BAC
600 bp
100 bp

Figure 2.2. Schematics of genomic loci examined in detail in this study.
Genome browser tracks showing gene annotations and ERα ChIP-seq data. Red arrows indicated the ERBSs that were examined by ChIP-qPCR. Orange arrows indicated the location of primer sets that were designed for the 3C analyses. The black arrows above the gene annotations indicate the direction of transcription for the annotated genes. Scale bars show the length of the indicated region. **(A)** ERBS1 and the *P2RY2* gene. **(B)** ERBS2 and the *GREB1* gene. **(C)** ERBS3 and the *SBNO2* gene. **(D)** ERBS4 and the *SMAD7* gene. **(E)** ERBS5 and the *PGR* gene. **(F)** ERBS6, which was used as a control enhancer that does not produce eRNAs.

[Figure 2.2 is on the next page]

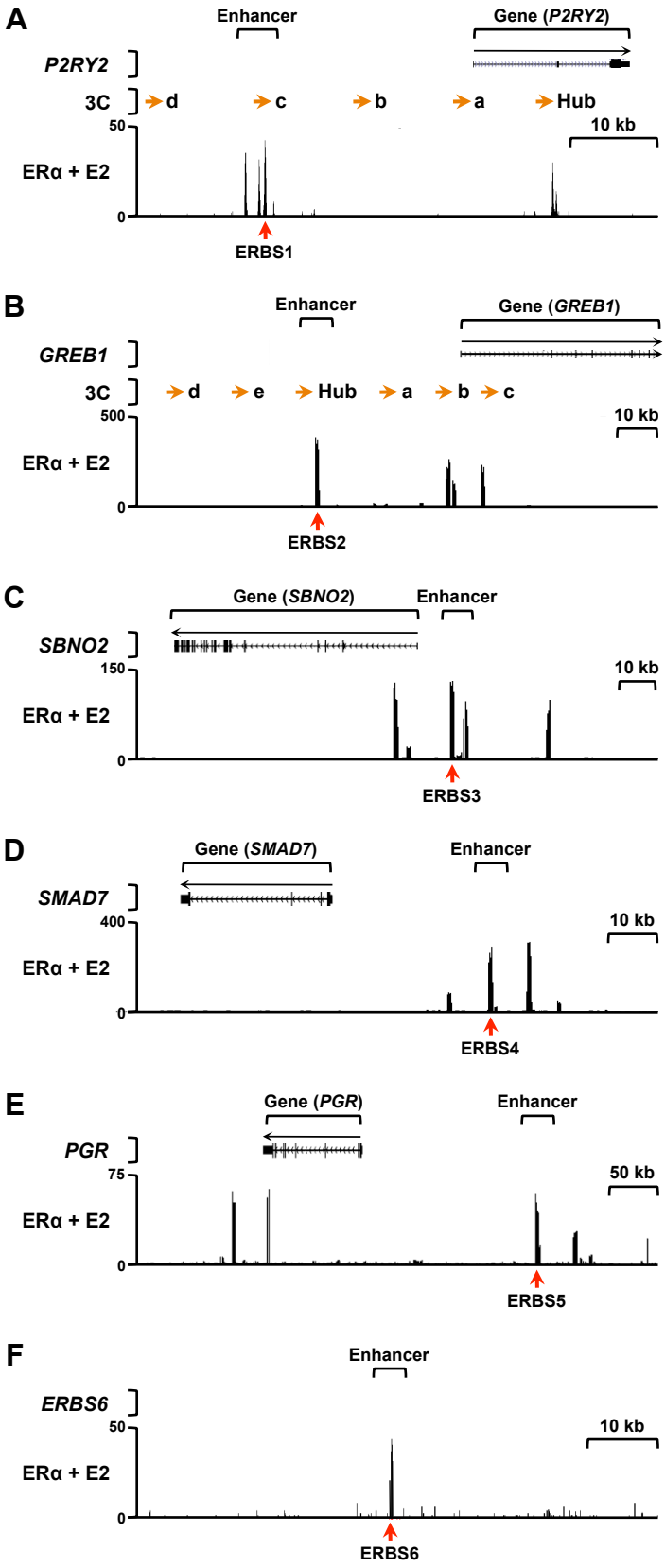

A
Enhancer
Gene (P2RY2)
P2RY2
3C
d
c
b
a
Hub
10 kb
50
ERα + E2
0
ERBS1
B
Enhancer
Gene (GREB1)
GREB1
3C
d
e
Hub
a
b
c
500
10 kb
ERα + E2
0
ERBS2
C
Gene (SBNO2)
Enhancer
SBNO2
150
10 kb
ERα + E2
0
ERBS3
D
Gene (SMAD7)
Enhancer
SMAD7
400
10 kb
ERα + E2
0
ERBS4
E
Gene (PGR)
Enhancer
PGR
75
50 kb
ERα + E2
0
ERBS5
F
Enhancer
ERBS6
50
10 kb
ERα + E2
0
ERBS6

Figure 2.3. Dynamics of ERα enhancer activity for the E2-responsive GREB1 gene in MCF-7 cells.

Locus-specific molecular assays for E2-responsive enhancers. Each bar represents the mean + the SEM for three or more independent biological replicates.

(A) ChIP-qPCR analyses for ERα and Pol II at a distal enhancer of the GREB1 gene (ERBS2) in response to a time course of E2 treatment.

(B) ChIP-qPCR analyses for H3K4me1, H3K4me3, and bulk H3 at ERBS2 in response to a time course of E2 treatment.

(C) RT-qPCR analyses for ERBS2 eRNA and *GREB1* mRNA in response to a time course of E2 treatment.

(D) 3C-PCR assays for E2-induced looping between ERBS2 and the *GREB1* gene. The lower case letters correspond to the primers denoted by orange arrows shown in Supplemental Figure 2.2. The assays were conducted in the presence (experimental) or absence (control) of DNA ligase, as indicated. Digested and ligated bacterial artificial chromosome (BAC) DNA spanning the entire *GREB1* locus was used as a PCR control. The size of the PCR fragments in bp is shown. One representative experiment from three conducted is shown.

[Figure 2.3 is on the next page]

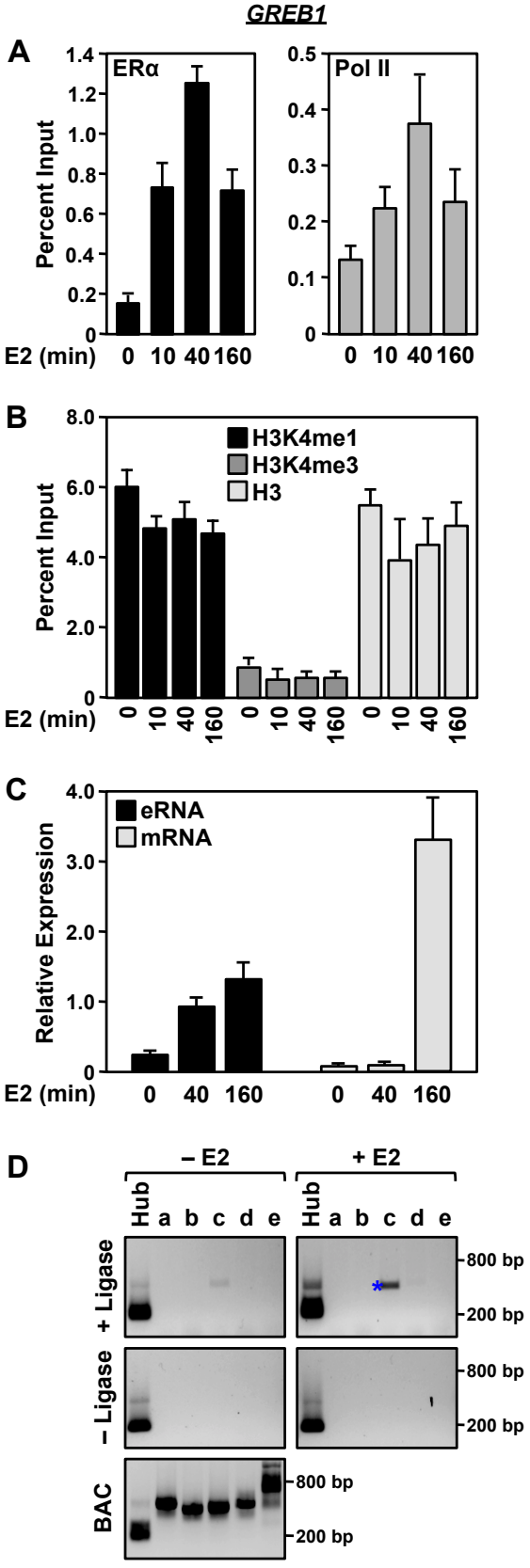

GREB1
A
ERα
Pol II
Percent Input
E2 (min) 0 10 40 160
B
H3K4me1
H3K4me3
H3
Percent Input
E2 (min)
C
eRNA
mRNA
Relative Expression
E2 (min) 0 40 160 0 40 160
D
– E2
+ E2
Hub a b c d e
+ Ligase
– Ligase
BAC
800 bp
200 bp

Figure 2.4. Genome-wide identification of ERα enhancer transcripts in MCF-7 cells using GRO-seq.

(A) Flowchart of ERBS classification in MCF-7 cells based on genomic location, eRNA production, and length of the transcribed region based on ChIP-seq and GRO-seq.

(B) Schematics of average transcribed regions overlapping ERBSs in MCF-7 cells in five classes: (a) short unpaired, (b) long unpaired, (c) short-short paired, (d) short-long unpaired, and (e) long-long, paired. "Short" and "Long" indicates a transcribed region <9 kb or ≥9 kb, respectively. Red and blue boxes indicate transcription from opposite strands.

(C) Graphical representation of the positions and orientations of eRNAs (indicated by red and blue lines) relative to ERα binding sites (indicated by yellow oval and line) for unpaired and paired eRNAs in MCF-7 cells. The position relative to the ERBS is indicated in kb. a through e correspond to the categories shown in panel B. Red and blue lines indicate transcription from opposite strands.

(D) Heat map showing the expression of E2-regulated short unpaired (S-U) and short/short paired (S-S) eRNAs over a time course of E2 treatment in MCF-7 cells based on GRO-seq data. The data were median centered and scaled to the 0 min. time point. Yellow and blue indicate upregulated and downregulated transcripts, respectively. Only unique transcripts are shown (i.e., those transcripts that overlap more than one ERBS are represented once).

(E) Metagene analyses of GRO-seq reads surrounding ERBS associated with short/short paired transcripts, short unpaired transcripts, or no transcripts in MCF-7 cells ± E2 treatment.

(F) Metagene analyses of Pol II ChIP-seq reads surrounding ERBS associated with short/short paired transcripts, short unpaired transcripts, or no transcripts in MCF-7 cells ± E2 treatment.

(G) Box plot representation of GRO-seq and Pol II ChIP-seq reads associated with short/short paired transcripts (S-S), short unpaired transcripts (S-U), or no transcripts in MCF-7 cells ± E2 treatment.

[Figure 2.4 is on the next page]

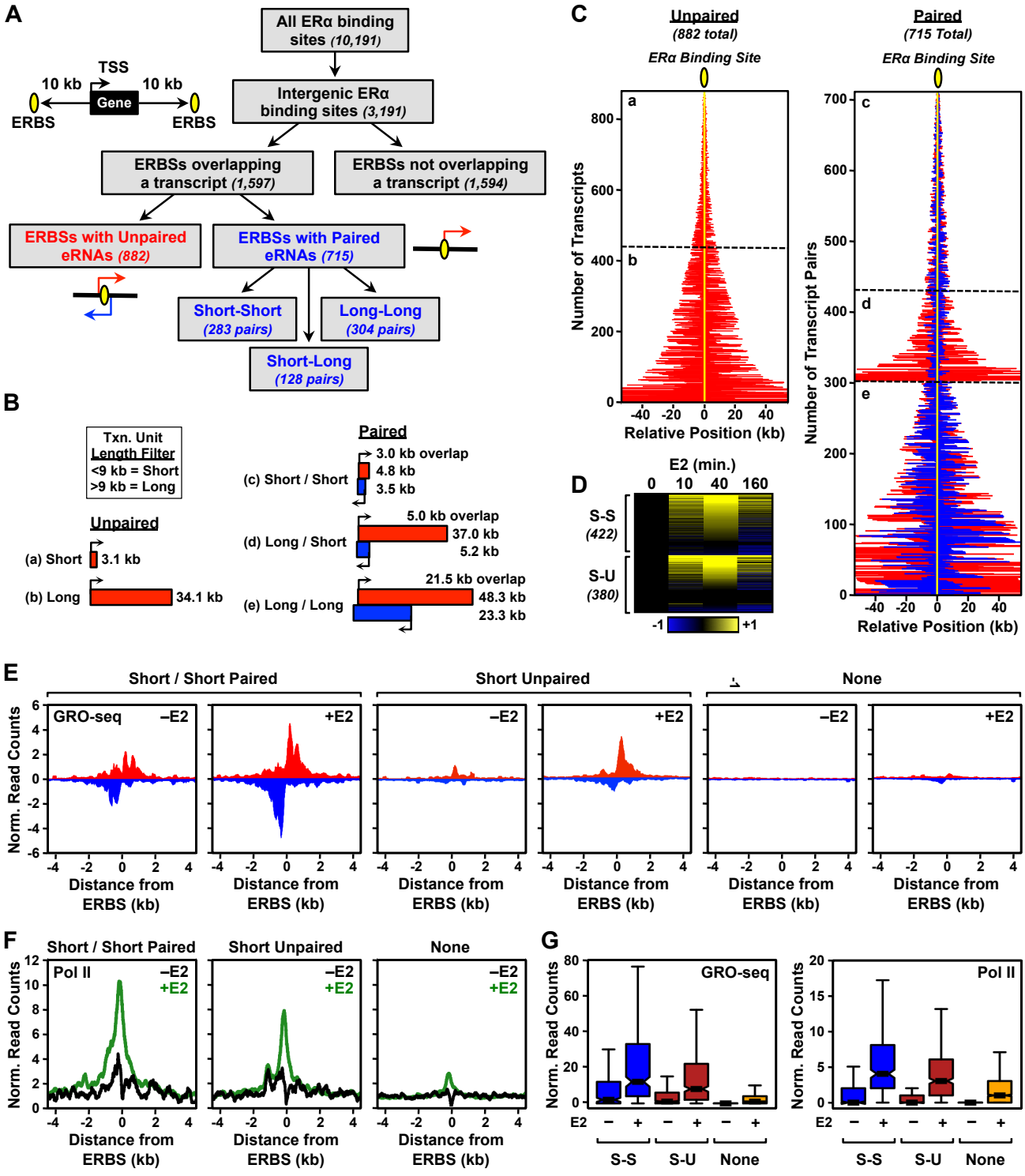

A
All ERα binding sites (10,191)
Intergenic ERα binding sites (3,191)
ERBSs overlapping a transcript (1,597)
ERBSs not overlapping a transcript (1,594)
ERBSs with Unpaired eRNAs (882)
ERBSs with Paired eRNAs (715)
Short-Short (283 pairs)
Long-Long (304 pairs)
Short-Long (128 pairs)
10 kb
TSS
10 kb
Gene
ERBS
ERBS
B
Txn. Unit Length Filter
<9 kb = Short
>9 kb = Long
Unpaired
(a) Short 3.1 kb
(b) Long 34.1 kb
Paired
(c) Short / Short
3.0 kb overlap
4.8 kb
3.5 kb
(d) Long / Short
5.0 kb overlap
37.0 kb
5.2 kb
(e) Long / Long
21.5 kb overlap
48.3 kb
23.3 kb
C
Unpaired
(882 total)
ERα Binding Site
Number of Transcripts
Relative Position (kb)
Paired
(715 Total)
ERα Binding Site
Number of Transcript Pairs
Relative Position (kb)
D
E2 (min.)
0 10 40 160
S-S (422)
S-U (380)
-1 +1
E
Short / Short Paired
Short Unpaired
None
GRO-seq
–E2
+E2
Norm. Read Counts
Distance from ERBS (kb)
F
Short / Short Paired
Short Unpaired
None
Pol II
–E2
+E2
Norm. Read Counts
Distance from ERBS (kb)
G
GRO-seq
Pol II
Norm. Read Counts
E2
S-S
S-U
None

Figure 2.5. The production of eRNAs from ERBSs positively correlates with the recruitment of coactivators, the levels of histone modifications, and the chromatin state in MCF-7 cells.

Browser tracks, metagenes, and boxblots showing a positive correlation between eRNA production at ERBS with known markers of enhancer function. *(Left two panels)* Browser track representations of coactivator or histone modification ChIP-seq data, or DNase-seq data, as indicated on the y-axis for ERBS1 and ERBS6. *(Middle three panels)* Metagene analyses of ChIP-seq or DNase-seq read counts for sets of ERBSs with short-short paired, short unpaired, or no transcripts in the presence *(green line)* or absence (*black line*) of E2 treatment. *(Right panel)* Box plot representations of ChIP-seq or DNase-seq data for sets of ERBSs with short-short paired *(blue boxes)*, short unpaired *(maroon boxes)*, or no transcripts *(yellow boxes)* in the presence (+) or absence (-) of E2 treatment.

(A) ERα ChIP-seq.
(B) FOXA1 ChIP-seq.
(C) CBP ChIP-seq.
(**D)** SRC3 ChIP-seq.
(**E)** H3K4me1 ChIP-seq.
(**F)** H3K4me3 ChIP-seq.
(**G)** DNase-seq.

[Figure 2.5 is on the next page]

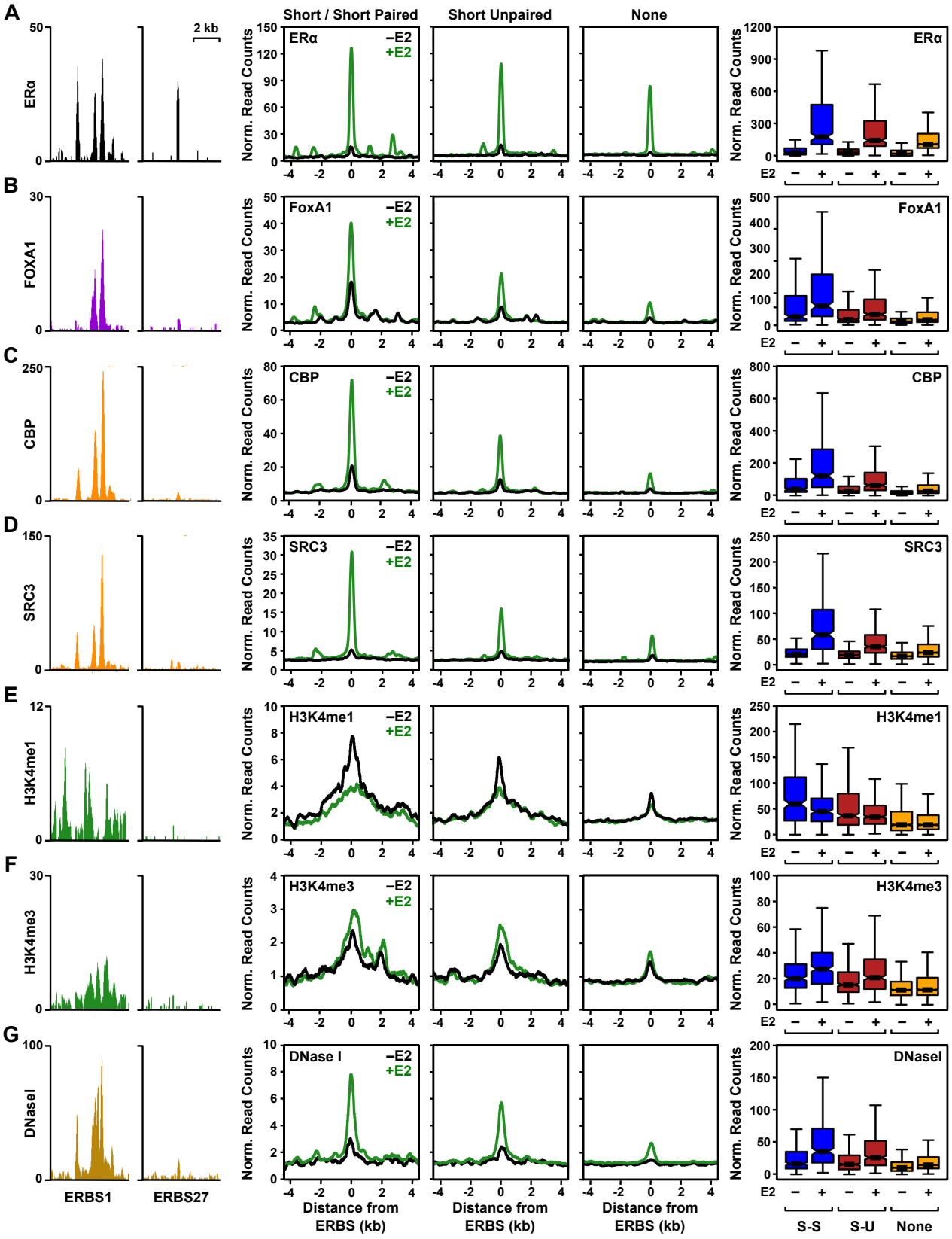

A
B
C
D
E
F
G
ERα
FOXA1
CBP
SRC3
H3K4me1
H3K4me3
DNaseI
2 kb
ERBS1
ERBS27
Short / Short Paired
Short Unpaired
None
Norm. Read Counts
–E2
+E2
Distance from ERBS (kb)
E2
S-S
S-U
None

Figure 2.6. Inhibition of eRNA production by flavopiridol does not inhibit ERα, Pol II, or coregulator binding, alter H3K4me1 or H3K27ac levels, or prevent enhancer looping at ERBSs in MCF-7 cells.

Locus-specific assays for E2-responsive enhancers showing the effects of a 1 hour pre-treatment with flavopiridol (FP) on various molecular outcomes in MCF-7 cells . Each bar represents the mean + the SEM for three or more independent biological replicates.

(A and B) Treatment with flavopiridol inhibits the E2-dependent production and steady-state accumulation of eRNAs and target gene mRNAs. RT-qPCR analyses for selected eRNAs and mRNAs in response to E2 treatment. (A) ERBS1 eRNA/*P2RY2* mRNA and (B) *ERBS2* eRNA/GREB1 mRNA.

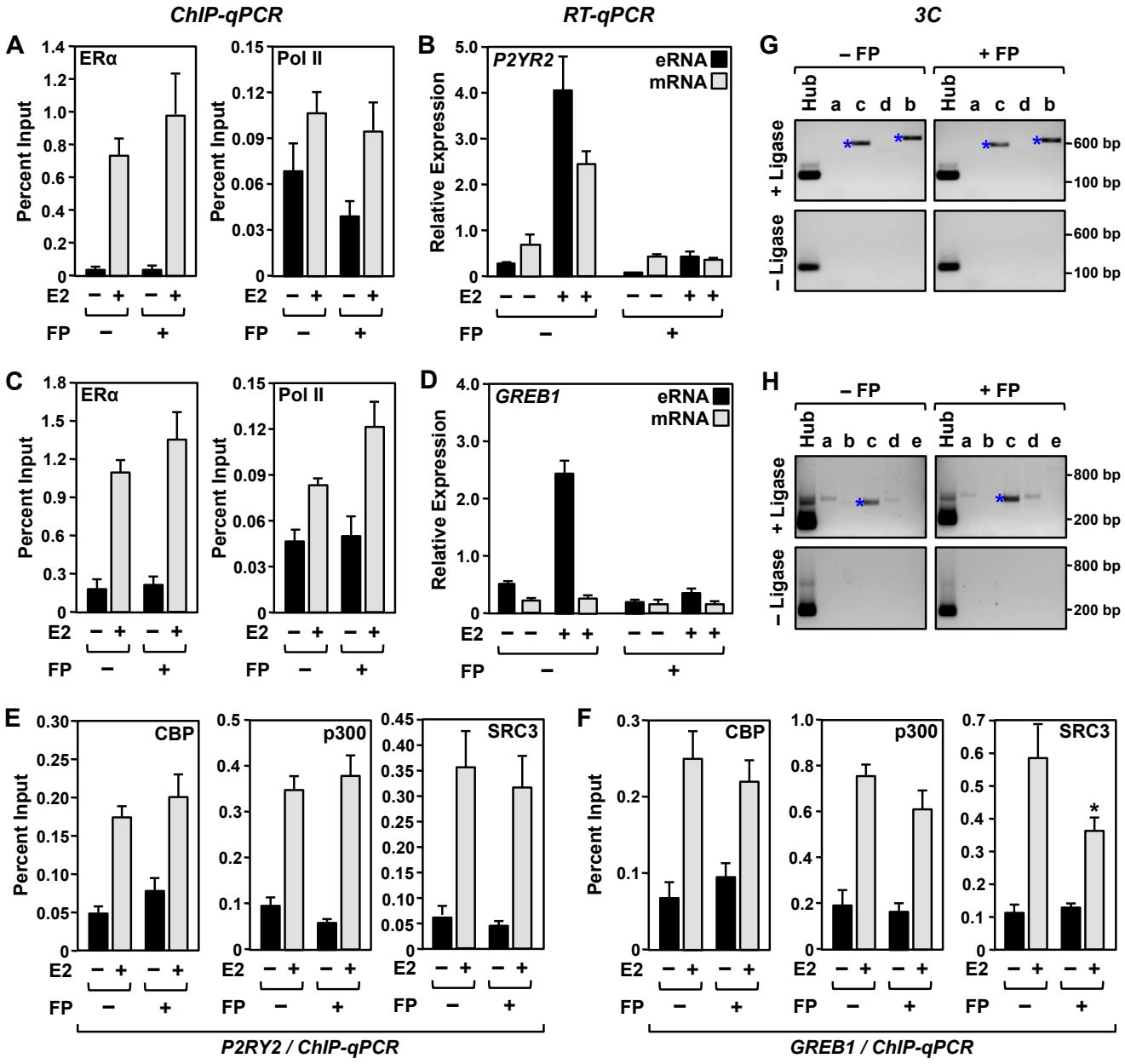
(C and D) ChIP-qPCR analyses for ERα *(left)* and Pol II *(right)* for (C) ERBS1 and (D) ERBS2 in the absence or presence of E2 and flavopiridol, as indicated.

(E and F) ChIP-qPCR analyses for CBP *(left)*, p300 *(middle)*, and Pol II *(right)* for (E) ERBS1 and (F) ERBS2 in the absence or presence of E2 and flavopiridol, as indicated.

(G and H) ChIP-qPCR analyses for H3K4me1 *(left)*, H3K27ac *(middle)*, and H3 *(right)* for (G) ERBS1 and (H) ERBS2 in the absence or presence of E2 and flavopiridol, as indicated.

(I and J) 3C-PCR analyses showing that looping between distal ERBSs and target genes in the presence of E2 is not blocked by flavopiridol (FP). (I) ERBS1/*P2RY2* and (J) *ERBS2*/GREB1. The lower case letters correspond to the primers denoted by orange arrows shown in Fig. 2.1A. The assays were conducted in the presence (experimental) or absence (control) of DNA ligase, as indicated. The size of the PCR fragments in bp is shown. One representative experiment from three conducted is shown.

[Figure 2.6 is on the next page]


ChIP-qPCR
RT-qPCR
3C
A
ERα
Pol II
Percent Input
E2
FP
B
P2YR2
eRNA
mRNA
Relative Expression
G
– FP
+ FP
Hub
a c d b
+ Ligase
– Ligase
600 bp
100 bp
C
GREB1
H
a b c d e
800 bp
200 bp
D
E
CBP
p300
SRC3
P2RY2 / ChIP-qPCR
F
GREB1 / ChIP-qPCR

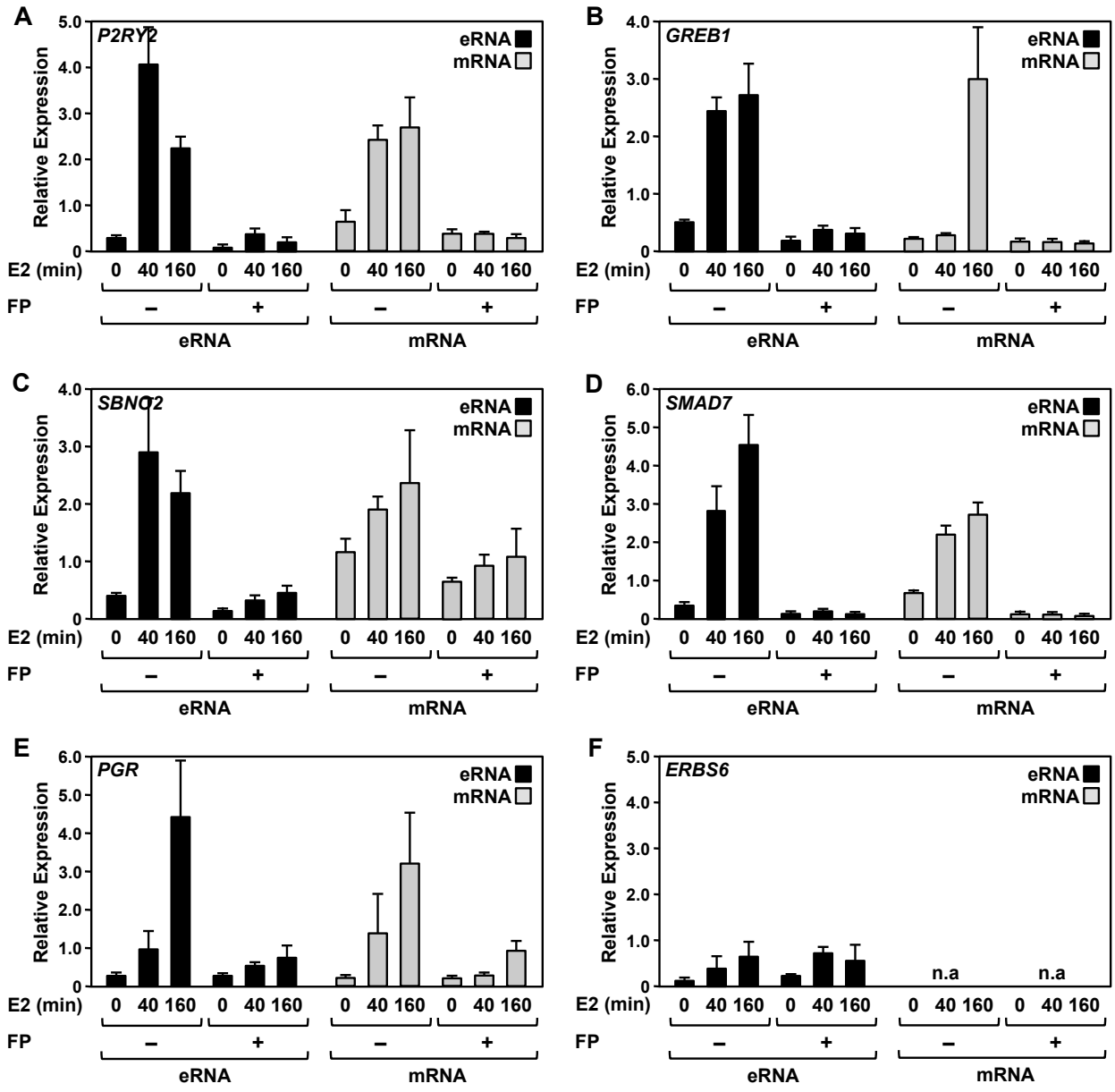


Figure 2.7. Expression of E2-responsive eRNAs and mRNAs in response to a time course of E2 treatment in MCF-7 cells.

RT-qPCR assays for eRNAs and mRNAs from enhancer/gene pairs in response to a time course of E2 treatment in MCF-7 cells with (+) or without (-) a 1 hour pretreatment with flavopiridol (FP). Each bar represents the mean + the SEM for three or more independent biological replicates. **(A)** ERBS1 and the *P2RY2* gene. **(B)** ERBS2 and the *GREB1* gene. **(C)** ERBS3 and the *SBNO2* gene. **(D)** ERBS4 and the *SMAD7* gene. **(E)** ERBS5 and the *PGR* gene. **(F)** ERBS6, which was used as a control enhancer that does not have eRNA transcripts.

CHAPTER THREE

Dynamic Assembly and Activation of Estrogen Receptor α Through Coregulator Switching

The study described in this chapter is from the prior publication in Murakami *et al.*, Genes and Development, 2017. The study was conducted in collaboration with Anusha Nagari, M.S. I initiated the study, designed and performed the experiments, and prepared the manuscript with W.L.K. A.N. performed analysis on the genomic datasets.

3.1 Summary

Although many features of active transcriptional enhancers have been defined by genomic assays, we lack a clear understanding of the order of events leading to enhancer formation and activation, as well as the dynamics of coregulator interactions within the enhancer complex. Herein, we used selective loss- or gain-of-function mutants of estrogen receptor alpha (ERα) to define two distinct phases of ligand-dependent enhancer formation. In the first phase (0 to 20 min.), p300 is recruited to ERα by Mediator, as well as p300's acetylhistone-binding bromodomain, to promote initial enhancer formation, which is not competent for sustained activation. In the second phase (20 to 45 min.), p300 is recruited to ERα by steroid receptor coregulators (SRCs) for enhancer maturation and maintenance. Successful transition between these two phases ('coregulator switching') is required for proper enhancer function. Failure to recruit p300 during either phase leads to abortive enhancer formation and a lack of target gene expression. Our results reveal an ordered and cooperative assembly of ERα enhancers requiring functional interplay among p300, Mediator, and SRCs, which has implications for hormone-dependent gene regulation in breast cancers. More broadly, our results demonstrate the unexpectedly dynamic nature of coregulator interactions within enhancer complexes, which are likely to be a defining feature of all enhancers.

3.2 Introduction

Transcription factors (TFs) control cell type-specific gene transcription by binding to their cognate DNA motifs in chromatin and nucleating enhancer formation by recruiting a set of transcriptional coregulators (Reiter et al., 2017). In many cases, enhancer formation and activation occur as an endpoint of cellular signaling pathways, allowing integration of extracellular cues with intracellular responses (Heinz et al., 2015). Despite wide variation in the types of TFs that may be expressed in a given cell, the enhancers that they form share several common features (Heinz et al., 2015; Shlyueva et al., 2014). For example, enhancers are enriched with a common set of histone modifications or 'marks,' including histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac) (Heintzman et al., 2009). Enhancers are also enriched with a common set of coregulators, such as the bromodomain-containing lysine acetyltransferases p300 and CBP (referred to collectively as p300/CBP), the RNA polymerase II (Pol II)-binding Mediator complex, and the ATP-dependent chromatin-remodeling complexes Swi/Snf (Heintzman et al., 2009; Reiter et al., 2017; Visel et al., 2009). While some of these features may be found at enhancers prior to full activation, others are further enriched or occur only in response to cellular signaling events (Heinz et al., 2015).

Recent studies have focused on additional enhancer features that generally correlate with enhancer activity, such as enhancer transcription and looping to target gene promoters. Genomic assays have revealed that many enhancers bound by RNA polymerase II (Pol II) are actively transcribed, producing enhancer RNAs ('eRNAs') (De Santa et al., 2010; Hah et al.,

2011; Kim et al., 2010) and loop to nearby target genes (Kulaeva et al., 2012). While looping may provide a logical framework for enhancer-promoter communication, the function of enhancer transcription is less clear. Three models have been proposed: (1) the act of enhancer transcription may help to open the chromatin and allow enhancer formation, (2) the eRNAs may act in *cis* at the enhancer from which they were transcribed to regulate enhancer function, or (3) they may act in trans as regulatory RNAs to control promoter function (Kim and Shiekhattar, 2015). Regardless of what the function is, enhancer transcription represents a robust mark of active enhancers that can be used to track enhancer activity (Hah et al., 2013; Kim and Shiekhattar, 2015; Wang et al., 2011). Although the features noted above (i.e., TF binding, enrichment of histone marks, coregulator recruitment, enhancer transcription, and looping) are readily identifiable, they provide only cryptic clues to enhancer biology. We lack a clear and detailed understanding of (1) the order of events leading to enhancer formation and activation, (2) the specific roles of each coregulator and feature enriched at an enhancer, and (3) the functional interrelationships among them.

Estrogen receptor alpha (ERα), a member of the nuclear receptor superfamily, is a ligand-regulated TF that is activated by the binding to estrogenic ligands, including 17β-estradiol (E2) (Burns and Korach, 2012; Nilsson et al., 2001). Within minutes of ligand-dependent activation, ERα dimerizes, binds to genomic DNA in chromatin, and promotes the formation of enhancers at the ERα binding sites to rapidly and transiently induce the transcription of target genes (Hah et al., 2011; Nilsson et al., 2001). Like the enhancers formed by other TFs, ERα enhancers are associated with the shared enhancer features noted above, including coregulators that associate with ERα in response to estrogen (Dasgupta et

al., 2014; Hah et al., 2013). Mediator and the steroid receptor coregulators (SRCs) are two of the best studied ERα coregulators, both of which interact directly with ERα through α-helix 12 in the major transactivation domain (i.e., activation function 2, or AF-2) of the ligand binding domain (LBD) (Dasgupta et al., 2014; Kang et al., 2002; Malik and Roeder, 2010; Shiau et al., 1998).

The Mediator complex, which binds to ERα through the Med1 subunit, interacts with RNA polymerase II (Pol II) and may interact with regulatory RNAs to promote enhancer-promoter looping (Chen and Roeder, 2011; Lai et al., 2013; Malik and Roeder, 2010). In contrast, the SRCs serve as scaffold proteins to recruit additional coregulators, many of which possess protein-modifying activities, to ERα binding sites, including p300/CBP (Dasgupta et al., 2014; Yi et al., 2015). The interactions of p300/CBP with ERα are mediated through the SRC proteins (Dancy and Cole, 2015; Kim et al., 2001; Yi et al., 2015). p300/CBP acetylate histones, TFs, and other coregulators, and have also been used as a mark of active enhancers (Dancy and Cole, 2015; Dasgupta et al., 2014; Visel et al., 2009). Amplifications, mutations, and aberrant functions of Mediator, SRCs, and p300/CBP have been implicated in some types of cancers, including hormone-dependent cancers (Anzick et al., 1997; Barbieri et al., 2012; Dasgupta et al., 2014). Indeed, p300 is one of the most commonly mutated genes across cancer types, including breast cancers (Chang et al., 2016; Kandoth et al., 2013).

Understanding the dynamics of enhancer formation and function is a critical component of overall understanding of enhancer biology. In the studies described herein, we have used selective loss- or gain-of-function ERα mutants, combined with genomic and

molecular biology assays, to dissect the order of assembly and activation of ERα enhancers, as well as the downstream biology.

3.3 Results

ERα L540Q Selectively Recruits Mediator, but Not SRCs, to ERα Binding Sites

To gain a better understanding of the molecular mechanisms that lead to the assembly and activation of enhancers at TF binding sites across the genome, we used ERα as a model signal-regulated TF. We applied a set of complementary sequencing-based genomic assays and qPCR-based locus-specific molecular assays in time course experiments designed to explore the kinetics of ERα enhancer assembly. We hypothesized that a selective loss-of-function ERα mutant defective in one activity, but not others, might provide unique insights into the order of enhancer assembly. ERα L540Q (ERαLQ) is one such mutant, which retains ligand, DNA, and Mediator binding, but is impaired in SRC binding, as determined by biochemical and cell-based assays (Acevedo et al., 2004; Ince et al., 1993; Schodin et al., 1995). As a result, ERαLQ exhibits reduced E2-dependent transcription (Acevedo et al., 2004; Ince et al., 1993; Schodin et al., 1995).

For our studies with this mutant ERα, we used previously described ERα-negative MDA-MB-231 breast cancer cells expressing either ERα wild-type (ERαWt) or ERαLQ (231/ERαWt and 231/ERαLQ cells, respectively) (Acevedo et al., 2004). To examine the selective loss of function of ERαLQ at ER binding sites genome-wide, we performed ChIP-

seq for ERα, Med1, SRCs (using a 'pan' SRC antibody), and p300 in both 231/ERαWt and 231/ERαLQ cells after 45 min of E2 treatment (note: Med1, SRCs, and p300 are expressed to similar levels in both cell types; Appendix S1). As expected, we observed E2-dependent binding of both ERαWt and ERαLQ to genomic loci, with considerable overlap between the sites of ERα binding between the two cell lines (Fig. 3.1, A and B; Appendix S2A, S2B, S3A, and S3B). In addition, we observed E2-dependent recruitment of Mediator (i.e., Med1) by both ERαWt and ERαLQ when filtered for common sites with similar levels of ERα binding (Fig. 3.1, A and B; Appendix S3, A and B). ERαWt also exhibited robust E2-dependent recruitment of SRC and p300, whereas ERαLQ exhibited significantly reduced recruitment of SRC and p300 (Fig. 3.1, A and B; Appendix. S3, A and B). Similar results were observed for individual SRC proteins (i.e., SRC1, 2, and 3) and CBP in ChIP-qPCR assays (Appendix S4 and S5). Thus, ERαLQ is a selective- loss-of-function mutant genome-wide.

Impaired SRC Recruitment Results in Impaired E2-dependent Transcription

Although previously characterized as a transcriptionally dead dominant-negative mutant in cell-based reporter gene assays (Ince et al., 1993; Schodin et al., 1995), ERαLQ has not been examined genome-wide in sequencing-based transcription assays. To determine how the impaired recruitment of SRC-p300/CBP observed with ERαLQ affects E2-dependent gene transcription, we used global run-on sequencing (GRO-seq), a genomic assay that detects sites of active transcription across the genome (Core et al., 2008). We treated 231/ERαWt and 231/ERαLQ cells with E2 for 0, 20, and 45 min and then subjected

them to GRO-seq. We identified 1,240 genes with significant E2-dependent regulation (up or down) in 231/ERαWt cells after 45 min. of treatment (Fig. 3.1C). In total, 747 genes were uniquely regulated by E2 in 231/ERαWt cells versus 231/ERαLQ cells, of which about half were up-regulated (Fig. 3.1, C and D).

Surprisingly, we also identified 788 genes with significant E2-dependent regulation in 231/ERαLQ cells (Fig. 3.1C; Appendix S3C). The majority of these genes were a subset of the genes regulated in 231/ERαWt cells, although we observed some uniquely regulated genes as well (Fig. 3.1, C and D; Appendix S3, C and D). We considered two sets of ERαLQ-regulated genes: (1) those regulated by both ERαWt and ERαLQ and (2) those regulated by ERαLQ only (Appendix S3D). In ChIP-seq assays, we observed impaired recruitment of SRC at ERα binding sites located near the genes regulated by both ERαWt and ERαLQ, but not at ERα binding sites located near the genes regulated by ERαLQ only (Appendix S6). These results suggest that the ERα enhancers regulating the expression of these genes may have an alternate mechanism for recruiting SRC. They also suggest that SRC is not sufficient for the activation of these genes because ERαWt recruits SRC to these binding sites, but does not promote E2-dependent gene transcription.

In spite of the transcriptional activity of ERαLQ with some genes, the extent of up-regulation was significantly less than the extent of up-regulation with ERαWt, especially for the 45 min. time point (Fig. 3.1D). Thus, the selective loss SRC-p300 binding by ERαLQ results in impaired E2-dependent transcription for a major subset of ERα target genes. For the purposes of this study, we focused on the genes up-regulated by E2 treatment only in the 231/ERαWt cells to compare enhancer formation by ERαWt versus ERαLQ.

ERα L540Q Supports a Subset of Active Enhancer Features In Spite of Impaired SRC Recruitment

The selective recruitment of Mediator versus SRCs with ERαLQ presents an interesting opportunity to dissect the molecular mechanisms of enhancer complex formation, perhaps by producing incompletely formed enhancers. The initial indications of this were manifest in reduced enhancer transcription with ERαLQ, as determined by GRO-seq, at a number of ERα enhancers (Fig. 3.1E; Appendix S3, A and B). To explore this in more detail, we surveyed the status of features thought to be indicative of active enhancers, such as Pol II recruitment, enhancer-promoter looping, and H3K27 acetylation. In ChIP-qPCR assays with five different ERα enhancers, we found that Pol II recruitment by ERαLQ was significantly reduced compared to ERαWt in most cases (Fig. 3.2A; Appendix S7), suggesting that the initiation of enhancer transcription is impaired with ERαLQ. This was also reflected in impaired Pol II loading at target gene promoters with ERαLQ (Appendix S8). Thus, we examined additional enhancer features likely to occur before the initiation of enhancer transcription.

Enhancer-promoter looping is a mechanism by which distal enhancers communicate with their target gene promoters (Dekker et al., 2002; Ong and Corces, 2011). We used chromatin conformation capture (3C)-qPCR assays with five different enhancer-promoter pairs to determine if ERαLQ is capable of promoting enhancer-promoter looping. Interestingly, ERαLQ promoted E2-dependent looping similar to, or even better than, ERαWt (Fig. 3.2B; Appendix S9). Likewise, the enrichment of H3K27ac, a histone modification mediated by p300 and used as a mark of active enhancers (Jin et al., 2011; Tang et al., 2013),

was similar for ERαWt and ERαLQ as determined by ChIP-seq (Fig. 3.2, C and D; Appendix S3, A and B). H3K27ac enrichment was maintained with ERαLQ despite reduced p300 recruitment (Fig. 3.1, A and B; Appendix S3, A and B). H3K27ac enrichment was, however, reduced in the presence of the p300 acetyltransferase inhibitor C646 (Bowers et al., 2010) (Fig. 3.2E; Appendix S10), suggesting p300 activity at an earlier point during ERα enhancer formation. Collectively, these initial studies indicate that ERαLQ is a selective loss-of-function mutant that still supports some aspects of enhancer formation and function (Fig. 3.2F).

Impaired SRC-p300 Recruitment Leads to Abortive E2-dependent Transcription at Enhancers and Promoters

We showed previously that only a subset of ERα binding sites are transcribed, indicating that ERα binding, per se, is not sufficient to establish an active enhancer or promote target gene transcription (Hah et al., 2013). To link impaired enhancer formation with ERαLQ to the transcriptional outcomes of target genes, we mined our ERα ChIP-seq and GRO-seq datasets from 231/ERαWt and 231/ERαLQ cells. We started with the 367 genes up-regulated by E2 with ERαWt, but not ERαLQ (i.e., ERαWt only), and determined the nearest common ERα binding site in the regulatory region, assuming that it would control the expression of the neighboring gene (Fig. 3.3A). We then determined the levels of transcription at the ERα binding sites and the corresponding target genes in 231/ERαWt and 231/ERαLQ cells (Fig. 3.3A). Interestingly, the E2-dependent transcription of the enhancer and the nearest gene were similar for ERαWt and ERαLQ at 20 min. of treatment (Fig. 3.3B).

However, continued up-regulation of enhancer and gene transcription at 45 min. of treatment was only observed with ERαWt (Fig. 3.3B). These results indicate a two-component response, with impaired ERαLQ activity only evident during the second component.

To explore the molecular underpinnings of the abortive transcriptional response at enhancers and promoters with ERαLQ in more detail, we monitored SRC and p300 recruitment by ChIP-qPCR for five different enhancers in 231/ERαWt and 231/ERαLQ cells during the same time course of E2 treatment. As expected, the kinetics of ERαLQ binding to chromatin was comparable to ERαWt (Fig. 3.3C, *left*; Appendix S11A and S11, B-E, *left*). Also as expected, SRC recruitment by ERαLQ was impaired at both time points of E2 treatment (20 and 45 min.) (Fig. 3.3C, *middle*; Appendix S11, B-E, *middle*). In contrast, p300 was recruited to similar levels by both ERαWt and ERαLQ after 20 min. of treatment. However, p300 recruitment by ERαLQ was significantly impaired at 45 min. of treatment compared to ERαWt (Fig. 3.3C, *right*; Appendix S11, B-E, *right*), coinciding with the two-component transcriptional response described above (Fig. 3.3B). These temporal effects with p300 recruitment were observed genome-wide in 231/ERα cells (Appendix S12), as well as in ERα knockdown MCF-7 cells re-expressing either ERαWt or ERαLQ (Appendix S13). Collectively, these results suggest that the first phase of ERα enhancer formation ('enhancer priming'), which is supported by ERαLQ, requires the recruitment of p300 through an SRC-independent mechanism. In contrast, the second phase ('enhancer maintenance'), which is not supported by ERαLQ, requires the recruitment of p300 through an SRC-dependent mechanism. Interestingly, these ERα enhancer events do not require the pioneer factor FoxA1, since the temporal effects with p300 were observed in both 231/ERα cells, which do

not express FoxA1, and MCF-7 cells, which do express FoxA1 (Appendix S14).

SRC-independent Enhancer Priming Requires Mediator and the Bromodomain of p300

The experiments in Fig. 3.3 suggest that p300 is recruited to liganded ERα in an SRC-independent manner during the first phase of enhancer formation (Fig. 3.4A). How might this occur? Black *et al.* showed that p300 is recruited by Mediator to GAL4-VP16-activated promoters to activate transcription (Black et al., 2006). Since Mediator and p300 are both recruited to enhancers, as well as promoters (Heintzman et al., 2009; Stumpf et al., 2006; Visel et al., 2009), we considered the possibility that p300 is recruited by Mediator to ERα enhancers during the first phase of activation. We found that E2-dependent recruitment of p300 was dramatically reduced upon knockdown of the ERα-interacting Mediator subunit, Med1, in 231/ERαWt cells (Fig. 3.4, B and C; Appendix S15, A-E). These results indicate that Mediator plays a role in p300 recruitment in the initial phase of enhancer formation. In addition, E2-dependent transcription of the enhancers and the corresponding target genes was significantly reduced by Med1 knockdown, but perhaps to a lesser extent than might be expected from given the dramatic reduction in p300 recruitment (Fig. 3.4D; Appendix S15, F-I).

p300 contains a bromodomain and a PHD domain, which allow p300 to bind to chromatin through acetylated and methylated histones, respectively (Dancy and Cole, 2015). Thus, we asked if these domains might also facilitate p300 enrichment at ERα enhancers, as well as p300-dependent enhancer priming and target gene transcription. We explored these questions using MCF-7 breast cancer cells, which express endogenous ERα. Treatment of

the cells with the p300/CBP bromodomain inhibitor, SGC-CBP30 (Hay et al., 2014), did not affect ERα binding to native ERα binding sites, as expected, as determined on multiple ERα enhancers (Fig. 3.5A, *left*; Appendix S16, A-D, *left*). In contrast, SGC-CBP30 reduced p300 recruitment to ERα binding sites throughout the E2 treatment time course (Fig. 3.5A, *middle*; Appendix S16, A-D, *middle*), indicating that the p300 bromodomain is also required for p300 recruitment during enhancer priming. Interestingly, we also observed reduced recruitment of Mediator upon treatment with SGC-CBP30 (Fig. 3.5A, *right*; Appendix S16, A-D, *right*), suggesting a cooperative recruitment of p300 and Mediator to the ERα binding sites.

To further investigate the role of the p300 bromodomain in the recruitment of p300 during E2-dependent enhancer priming, we expressed a fragment of p300 that contains the bromodomain (BRP) (Delvecchio et al., 2013) fused to the Fc region of rabbit IgG (BRP-Fc) in MCF-7 cells (Fig. 3.5B). We performed ChIP using the Fc region for immunoprecipitation and examined the E2-dependent recruitment of BRP-Fc to native ERα binding sites in MCF-7 cells. BRP-Fc was enriched at ERα enhancers at 20 min. of E2 treatment, but was depleted by 45 min. (Fig. 3.5C; Appendix S16E), consistent with the time course of p300 recruitment and SRC-dependence determined in the experiments described above. In follow-up experiments in MCF-7 cells, we found that inhibition of the p300 bromodomain (with SGC-CBP30) and the p300 acetyltransferase activity (with C646) both significantly reduced E2-dependent gene expression (Fig. 3.5D; Appendix S16, F-I), demonstrating the critical role of bromodomain-dependent/SRC-independent recruitment of p300, as well as p300 acetyltransferase activity, in ERα-mediated gene expression. Taken together, these studies have revealed a cooperative recruitment of p300 and Mediator at ERα

enhancers for enhancer priming, prior to the formation of a mature and fully active ERα enhancer complex.

Forced Recruitment of p300 to Inactive ERα Binding Sites Increases Mediator Recruitment and Induces E2-dependent Gene Expression

The experiments in Figs. 3.1 through 3.5 identified critical roles for p300 recruitment in two phases of enhancer formation (priming and maintenance) by at least three distinct mechanisms (Mediator-, bromodomain-, and SRC-dependent recruitment). To fully test and explore this hypothesis, we engineered an experimental system to examine the function of p300 at ERα enhancers independent of recruitment by other ERα coregulators. First, we deleted a part of α-helix 12 (Glu 542 through His 547) in the AF-2 domain of ERα to generate a ligand binding- and DNA binding-competent, but transcriptionally impaired ERα (ERαΔH12) (Brzozowski et al., 1997; Shiau et al., 1998). Next, we fused the p300/CBP binding domain (PID) of SRC2 (Li and Chen, 1998; Sheppard et al., 2001; Voegel et al., 1998) to ERαΔH12 to generate ERαΔH12-PID (Fig. 3.6A). Note that SRC2(PID) functions as a potent activation domain by recruiting p300/CBP (Acevedo and Kraus, 2003; Kim et al., 2001). We expressed ERαWt, ERαΔH12, or ERαΔH12-PID in MBA-MB-231 cells using a Dox-inducible system to generate 231/ERαWt, 231/ERαΔH12, and 231/ERαΔH12-PID cells, respectively (Fig. 3.6A).

As expected, all three ERαs were similarly recruited to chromatin after E2 treatment (Fig. 3.6B, *left*; Appendix S17, A and B, *left*). p300 was robustly recruited in an E2-dependent manner by chromatin-bound ERαWt and ERαΔH12-PID, but exhibited impaired recruitment

by ERαΔH12, at three different ERα enhancers (Fig. 3.6B, *right*; Appendix S17, A and B, *left middle*). Med1 recruitment and H3K27ac enrichment occurred effectively with ERαΔH12-PID, but were impaired with ERαΔH12 (Fig. 3.6C; Appendix S17, A and B, *right panels*). Likewise, enhancer-promoter looping and target gene activation occurred effectively with ERαΔH12-PID, but were impaired with ERαΔH12 (Fig. 3.6, D and E; Appendix S17, C-F). Together, these results with forced recruitment of p300 to an otherwise inactive ERα binding site demonstrate that p300 recruitment is sufficient for enhancer formation (e.g., Med1 recruitment, H3K27ac, looping) and target gene activation.

The Mediator- and SRC-dependent Functions of p300 Link ERα Enhancer Activity to Clinical Outcomes in ER-positive Breast Cancer Patients

Our results have identified ordered and cooperative interactions among p300, Mediator, and SRCs leading to the formation, activation, and maintenance of ERα enhancers. To explore the biological outcomes of these events, we used the p300 bromodomain inhibitor SGC-CBP30, the p300 acetyltransferase inhibitor C646, and siRNAs targeting Med1 in MCF-7 proliferation assays. Each treatment alone effectively inhibited the growth of MCF-7 cells over a 6 day time course of E2 treatment (Fig. 3.7A; Appendix S18, A and B). Moreover, combination treatments with Med1 siRNAs and either p300 inhibitor were even more effective than the treatments alone (Fig. 3.7A; Appendix S18, A and B). These studies link the functions of Mediator and p300 in ERα-dependent transcription to downstream biological outcomes.

Our results using ERαL540Q indicate an important role for SRCs in the maintenance of ERα enhancer activity. SRC3 is frequently amplified in ER-positive breast cancers (Anzick et al., 1997; Gojis et al., 2010). An increased dosage of SRCs are likely to support E2-mediated mitogenic growth of breast cancers by promoting ERα enhancer maintenance. Using datasets from The Cancer Genome Atlas (TCGA), we found that 15% and 5% of ER-positive breast tumors contain amplifications of *SRC2* and *SRC3*, respectively (Fig. 3.7B). These gene amplifications correlate with the overexpression of the cognate mRNAs (Fig. 3.7C). Next, we examined the clinical outcomes of patients with ER-positive breast cancers, with or without *SRC* amplification. In this analysis, the gene expression profiles for the patient samples with or without *SRC* amplification were compared to generate a list of genes more highly expressed in the group with *SRC* amplification relative to the group without *SRC* amplification, which we called the "SRC signature" (Appendix S18C). The expression of each SRC signature gene was then used to segregate ER-positive luminal breast cancers into groups with "high" or "low" expression of the signature genes using curated microarray data with associated clinical outcome data that are independent of the TCGA data. The collective clinical outcomes stratified according to each SRC signature gene expression were represented in meta-Kaplan-Meier plots (Fig. 3.7D and S18D) (Mihaly et al., 2013; Pongor et al., 2015). Importantly, elevated expression of both the SRC2 and SRC3 signature genes correlates with unfavorable outcomes of patients with luminal breast cancers. Collectively, these cell growth and clinical data analyses link the Mediator- and SRC-dependent functions of p300 at ERα enhancers to clinical outcomes in ER-positive breast cancer patients.

3.4 Discussion

In the current study, we identified a critical functional interplay among p300, Mediator, and SRCs that recruits p300 and determines its activity throughout the time course of ERα enhancer formation and activation. Importantly, we identified two distinct mechanistic phases of ERα enhancer assembly and function, which are defined by the mode of p300 recruitment. Successful transition between these two phases ('coregulator switching') is required for proper enhancer function. Collectively, our studies have revealed the detailed molecular and temporal mechanisms of signal-regulated enhancer assembly and activity that are likely to be applicable to a wide variety of signal-regulated enhancers. Furthermore, our results link p300, Mediator, and SRCs function at ERα enhancers to hormone-dependent gene regulation in ER-positive breast cancers.

An Ordered and Cooperative Assembly and Function of ERα Enhancers

Previous results have suggested a critical role for SRCs, which bind directly to ERα in a ligand-dependent manner, in recruiting p300 to ERα binding sites. Our results using a selective loss-of-function ERα mutant (i.e., L540Q), however, suggest a more intricate and varied set of mechanisms for recruiting p300 to ERα binding sites in two distinct phases. In the first phase (0 to 20 min.), p300 is recruited through a Mediator-dependent mechanism, perhaps through direct interactions, as well as the bromodomain of p300, which binds to acetylated histones like other bromodomains (Zeng and Zhou, 2002). In the second phase, p300 is recruited though an SRC-dependent mechanism, which is also likely to occur through

direct interactions. The former is required for initial enhancer formation and enhancer priming, while the latter is required for enhancer maturation and the maintenance of enhancer activity (Fig. 3.7F). We refer to the dynamic changes in the mode of p300 recruitment as 'coregulator switching." Failure to recruit p300 during either phase leads to abortive enhancer formation and a lack of target gene expression. These results suggest a critical role of SRCs in the later stages of ERα enhancer function, but not in the initiation of enhancer formation.

A number of independent lines of investigation support the conclusions stated above. For example, siRNA-mediated knockdown of Med1, the ERα-interacting component of the Mediator complex, reduces the E2-dependent recruitment of p300 to ERα binding sites (Fig. 3.4C). Furthermore, a fragment of p300 containing the acetyllysine-binding bromodomain, but lacking the SRC binding domain, is recruited to ERα binding sites within the first 20 minutes following E2 treatment (Fig. 3.5C). Conversely, the competitive p300 bromodomain inhibitor, SGC-CBP-30, blocks p300 recruitment to ERα binding sites (Fig. 3.5A).

Additional experiments using loss- or gain-of-function ERα mutants also support the conclusions stated above. For example, the selective loss-of-function ERα mutant ERα L540Q, which recruits Mediator, but not SRCs, supports enhancer formation in the first phase, but is impaired in enhancer maturation and maintenance in the second phase (Figs. 3.1B and 3.3). Interestingly, the abortive enhancer formation with ERα L540Q results in a dramatic reduction of enhancer transcription (Fig. 3.3B), but maintains a subset of active enhancer features, including enrichment of H3K27ac (Fig. 3.2D) and enhancer-promoter chromatin looping (Fig. 3.2B). In contrast, forced recruitment of p300 to an inactive ERα

binding site using the selective gain-of-function ERα mutant ERαΔH12-PID promotes enhancer activation, enhancer-promoter looping, and target gene transcription (Fig. 3.6). Interestingly, ERαΔH12-PID also promotes the recruitment of Mediator (Fig. 3.6C), while SGC-CBP-30 inhibits the recruitment of Mediator (Fig. 3.5A), illustrating a functional interplay between these two coregulators at enhancers, which has been demonstrated previously at promoters (Black et al., 2006).

Recruitment of p300 is Necessary and Sufficient for Enhancer Formation at ERα Binding Sites

Since p300 and CBP were identified in the 1980s, they have been shown to function as key coregulators for many TFs through dual roles as scaffolding proteins for the recruitment of other coregulators, and as enzymes for the modification of histones and other transcription-related proteins (Dancy and Cole, 2015; Wang et al., 2013). Genome-wide analyses have shown that p300 and CBP are associated with active enhancers and, as such, are good markers of active enhancers (Heintzman et al., 2009; Visel et al., 2009). Our results with the p300 catalytic inhibitor, C646, and the p300 bromodomain inhibitor, SGC-CBP-30, indicate that p300 is required for enhancer formation at ERα binding sites (Fig. 3.5D). Furthermore, our results with the selective gain-of-function ERα mutant, ERαΔH12-PID, which recruits p300 to a transcriptionally impaired ERα, suggest that p300 recruitment is sufficient for enhancer formation at ERα binding sites (Fig. 3.6). Although we cannot rule out cooperative effects with other coregulators that may be partially or fully recruited to ERαΔH12, our results are consistent with previous studies using the yeast Gal4 DNA binding

domain to recruit the CBP acetyltransferase domain to activate a reporter gene (Bannister and Kouzarides, 1996). Collectively, our results indicate an essential role for p300 in ERα enhancer formation and function. In this regard, it makes sense that nature has devised and exploited multiple modes for p300 recruitment to TFs, such as ERα (i.e., via Mediator, SRCs, and the p300 bromodomain).

Successful transition between the two phases of ERα enhancer formation, reflected in the different modes of p300 recruitment, is required for proper enhancer function. Failure to transition leads to abortive enhancer formation and reduced E2-dependent gene activation. Experimentally, this can be demonstrated by inhibiting each of the modes of p300 recruitment, using Med1 knockdown or a p300 bromodomain inhibitor in the first phase (Figs. 3.4, 3.5A, and 3.5D), or the SRC-binding mutant ERα L540Q in the second phase (Fig. 3.3). These results suggest that the enhancer complex is dynamic, with different protein-protein interaction surfaces used to accomplish distinct outcomes at specific times. This observation is consistent with a number studies showing rapid and dynamic binding and exchange of coregulators at nuclear receptor binding sites (Stenoien et al., 2001; Voss and Hager, 2014).

As noted above, p300 and CBP have an intrinsic acetyltransferase activity that allows acetylation of lysine residues in substrate proteins, which include histones, TFs, and coregulators (Dancy and Cole, 2015). In fact, p300 acetylates ERα, which can enhance the DNA binding activity of ERα (Kim et al., 2006). Studies with the p300 catalytic inhibitor, C646, indicate that p300 acetyltransferase activity is required for both phases of ERα assembly and function (Figs. 3.2E and 3.5D). Although the relevant substrates within the

enhancer complex are, at present, unknown, it is possible that they may differ from one phase of enhancer activation to the next. In our assays, p300 recruitment and activity track with enhancer transcription, and impaired recruitment of p300 by ERα L540Q correlates with reduced enhancer transcription (Fig. 3.3, B and C). In this regard, Bose et al. have recently shown that non-coding RNAs, including eRNAs, can stimulate CBP (and presumably p300) histone acetyltransferase activity (Bose et al., 2017). It is interesting to speculate that p300-dependent enhancer transcription in the first phase may have a feed-forward effect to stimulate p300 acetyltransferase activity in the second phase through eRNA production.

p300, Mediator, and SRCs: Links Between ERα Enhancer Activity, Breast Cancer Cell Growth, and Clinical Outcomes in Breast Cancers

E2 signaling, which promotes p300, Mediator, and SRC recruitment, ERα enhancer formation, and target gene activation, has potent mitogenic effects on the proliferation of ER-positive breast cancer cells. In this regard, p300, Mediator, and SRCs have all been linked to human diseases, including cancers (Anzick et al., 1997; Gojis et al., 2010; Spaeth et al., 2011; Wang et al., 2013). In cell proliferation assays, we observed a dramatic effect of p300 inhibitors on the growth of ER-positive breast cancer cells, which was further enhanced by the knockdown of Med1 (Fig. 3.7A). These cell-based results mirror the molecular interplay between p300 and Mediator that we observed during ERα enhancer formation, providing a molecular link between ERα enhancer activity, E2-dependent gene expression, and breast cancer cell growth.

The genes encoding SRC2 and SRC3 are up-regulated in a subset of ER-positive luminal breast cancers (Fig. 3.7C; (Anzick et al., 1997; Gojis et al., 2010)). In addition, SRC2- and SRC3-dependent gene signatures track with clinical outcomes in breast cancer patients (i.e., lower expression, better outcomes) (Fig. 3.7D). These clinical results mirror the molecular interplay between p300 and SRC that we observed during ERα enhancer formation, providing yet another molecular link between ERα enhancer activity, E2-dependent gene expression, and breast cancer cell growth. Interestingly, amplification of SRCs in breast cancer cells may act to cause the aberrant recruitment of p300 to ERα binding sites, which may drive spurious enhancer formation or activation. Given the role of SRCs in ERα enhancer maintenance described above, amplification of SRCs may promote sustained enhancer activation resulting in prolonged mitogenic responses. Such effect would justify the on-going efforts to develop drugs that inhibit the interactions between ERα and SRCs (Rodriguez et al., 2004; Song et al., 2016), which can be a potent way of inhibiting SRC-, p300-, and Mediator-dependent transcription by ERα (Acevedo and Kraus, 2003; Kim et al., 2001).

3.5 Materials and Methods

Antibodies

The antibodies used were as follows: ERα (custom rabbit polyclonal antiserum generated in the Kraus Laboratory against the first 113 amino acids of human ERα (Kraus

and Kadonaga, 1998)); ERα (Enzo Biochem, ADI-SRA-1000-F); pan-SRC (rabbit polyclonal antiserum generated in the Kraus Laboratory against amino acids 624-1130 of mouse SRC2 (Acevedo et al., 2004)); p300 (Bethyl Laboratories, A300-358A); p300 (Santa Cruz Biotechnology, sc-584x); Med1 (Bethyl Laboratories, A300-793A); Med1 (Santa Cruz Biotechnology, sc-5334x); Pol II (Santa Cruz Biotechnology, sc-899x); H3K27ac (Abcam, ab4729); SRC1 (Santa Cruz Biotechnology, sc-8995); SRC2 (Santa Cruz Biotechnology, sc-8996); SRC3 (Santa Cruz Biotechnology, sc-9119); CBP (Santa Cruz Biotechnology, sc-369); β-actin (Cell Signaling, 3700S); snRNP70 (Abcam, ab83306); β-tubulin (Abcam, ab6046); and PARP-1 (custom rabbit polyclonal antiserum now available from Active Motif; cat. no. 39559).

Cell Culture and Treatments

MDA-MB-231 cells were obtained from the American Type Culture Collection (ATCC). Parental MDA-MB-231 cells, or MDA-MB-231 expressing ERα wild-type, L540Q, ΔH12, or ΔH12-PID, generated as described below, were maintained in phenol red-free Dulbecco's modified Eagle's medium/nutrient mixture F-12 Ham (DMEM/F-12; Sigma, D2906) supplemented with 10% charcoal-dextran treated calf serum (Sigma, C8056), 6 ng/mL human recombinant insulin (Sigma, I5500), 3.75 ng/mL hydrocortisone (Sigma, H0888), 16 μg/mL glutathione (Sigma, G6013), 100 units/mL penicillin/streptomycin (Gibco, 15140122), and 25 μg/mL gentamicin (Gibco, 15710064). The same conditions were used for experiments.

MCF-7 cells, kindly provided by Dr. Benita Katzenellenbogen (University of Illinois at Urbana-Champaign, IL), were maintained in minimal essential medium (MEM; Sigma, M1018) supplemented with 5% calf serum (Sigma, C8056), 100 units/mL penicillin-streptomycin (Gibco, 15140122), and 25 μg/mL gentamicin (Gibco, 15710064). Three days prior to experiments, the medium was switched to minimal essential medium Eagle (Sigma, M3024) supplemented with 5% charcoal-dextran treated calf serum (Sigma, C8056).

293T cells were obtained from the ATCC. There were maintained in Dulbecco's modified Eagle medium/high glucose (Sigma, D7777) supplemented with 10% FBS (Sigma, F2442), 100 units/mL penicillin-streptomycin (Gibco, 15140122), and 25 μg/mL gentamicin (Gibco, 15710064).

For experiments, the cells were treated with vehicle (DMSO) or 100 nM 17β-estradiol (E2) (Sigma, E8875) for the specified amount of time. Where indicated, the cells were treated with vehicle (DMSO), 25 μM of the p300/CBP acetyltransferase inhibitor C646 (Sigma, SML0002) (Bowers et al., 2010), or 50 μM the p300/CBP bromodomain inhibitor SGC-CBP30 (Sigma, SML1133) (Hay et al., 2014) for 30 min. prior to E2 treatment unless noted otherwise.

Preparation of Transgenic Cells with Ectopic Protein Expression

MDA-MB-231 cells with constitutive expression of ERα wild-type or L540Q were prepared as described previously (Acevedo et al., 2004). MDA-MB-231 cells with doxycycline-inducible expression of ERα wild-type, L540Q, ΔH12, or ΔH12-SRC2(PID) were generated by lentivirus-mediated transduction using the pINDUCER20 vector, kindly

provided by Dr. Thomas Westbrook (Baylor College of Medicine, Houston, TX) (Meerbrey et al., 2011). Before use, pINDUCER20 was modified to replace the Gateway® cloning sites with restriction enzyme sites suitable for conventional or Gibson cloning. cDNAs encoding the wild-type or variant ERαs noted above were constructed and cloned into the modified pINDUCER20 vector. A cDNA fragment encoding amino acids 1010-1130 of human SRC2 containing the p300-interacting domain (PID) (Acevedo and Kraus, 2003; Kim et al., 2001) was cloned into the ERαΔH12 cDNA-containing vector to generate a C-terminal fusion. A cDNA fragment encoding the p300 Bromodomain-Ring-PHD domain (p300 BRP) (Delvecchio et al., 2013) was ligated to a cDNA fragment encoding the rabbit immunoglobulin G (IgG) Fc region and cloned into the modified pINDUCER20 vector. The protein expressed from this construct functions as an antibody-like p300 chromatin binding domain.

The pINDUCER20-based vectors described above were transfected into 293T cells using GeneJuice (Millipore, 70967) along with pCMV-VSV-G, pCMV-GAG-pol-Rev, and pAdVantage (Promega, E1711) for recombinant lentivirus production. The resulting viruses were filtered through a 0.44 μm filter and used to infect target cells. Infected cells were selected using 600 μg/mL G481 (Life Technologies, 11811031) to generate a pool of resistant cells stably harboring the transgene. After selection, ectopic expression of the transgenes was induced by treating the cells with 25 to 500 ng/mL doxycycline (Dox; titrated for each cell line to achieve similar expression levels) for 24 hours prior to experiments.

MCF-7 cells with simultaneous Dox-dependent shRNA-mediated knockdown of endogenous ERα and ectopic expression of ERα was prepared by sequential transduction of

MCF-7 cells as follows. pTRIPZ-shRNA vectors were purchased from Dharmacon: shNegativeControl (RHS4743) and shERα targeting the *ESR1* 3'UTR (RHS4740-EG2099, V3THS_405935). The pTRIPZ-shRNA vectors were transfected into 293T cells using GeneJuice (Millipore, 70967) along with psPAX2 and pMD2.G for recombinant lentivirus production. The resulting viruses were filtered through a 0.44 μm filter, concentrated using Lenti-X™ concentrators (Clontech, 631232) according the manufacturer's protocol, and used to infect target cells. Infected cells were selected using 1 μg/mL Puromycin (Sigma, P8833) to generate a pool of resistant cells stably harboring the transgene. After selection, shRNA expression was induced by treating the cells with 200 to 500 ng/mL Dox for 72 hours prior to experiments. The knockdown efficiency was determined by Western blotting as described below. After confirming efficient knockdown, the MCF-7-pTRIPZ-shERα cells were transduced with the pINDUCER20-based vectors described above. After selection, the cells were treated with 10-500 ng/mL Dox for 72 hours to determine the Dox concentration that promotes efficient knockdown of endogenous ERα and re-expression of the ectopic ERα (Wt and L540Q mutant) from the transgenes at a level similar to endogenous ERα in MCF-7 cells.

siRNA-mediated Knockdown

siRNA-mediated knockdown was performed using Lipofectamine RNAiMax reagent (Invitrogen, 13778150) following the manufacturer's instructions using the following siRNAs: siControl (MISSION universal negative control #1; Sigma, SIC001); siMed1 (Sigma, SASI_Hs01_00089551); and siERα (Sigma, SASI_Hs01_00078598). The siRNAs

were diluted to 100 nM in antibiotic-free, serum-free MEM and mixed by vortex in a tube. Lipofectamine RNAiMAX reagent was added to the diluted siRNAs and mixed by pipetting. The transfection complexes were transferred onto each well or plate by pipetting. The cells were seeded over the siRNA/lipofectamine transfection complex (typically 5 x 10^4 cells per well in a 6-well plate) to a final siRNA concentration of 10 nM and incubated at 37°C, 5% CO_2 for 48 to 72 hours before experiments.

Western Blotting

The cells were collected by scraping in PBS containing complete protease inhibitor cocktail (Roche, 11697498001) and pelleted by gentle centrifugation. To prepare whole cell lysates, the cell pellets were resuspended in Lysis Buffer [20 mM HEPES pH 7.5, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% glycerol, 0.5% NP-40, 1 mM DTT, and 1x complete protease inhibitor cocktail (Roche, 11697498001)] and mixed by occasional gentle vortexing for 15 min at 4°C. The lysates were clarified by centrifugation at full speed in a microcentrifuge at 4°C. The supernatants were collected as whole cell lysates, snap frozen, and stored at -80°C until use.

Where indicated, the cells were separated into cytoplasmic and nuclear fractions. Briefly, the cells were lysed in Isotonic Buffer [10 mM Tris•HCl pH7.5, 300 mM sucrose, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 0.5% NP-40, 1 mM DTT, and 1x complete protease inhibitor cocktail (Roche, 11697498001)] by repeated pipetting. The nuclei were pelleted by gentle centrifugation and the supernatant was collected as cytoplasmic lysate, snap frozen, and stored at -80°C until used. The nuclear fraction was prepared by extracting the nuclei in

Lysis Buffer with occasional gentle vortexing for 15 min at 4°C. The nuclear extracts were clarified by centrifugation at full speed in a microcentrifuge at 4°C, snap frozen, and stored at -80°C until used.

For assays, the cell lysates were thawed, and the protein concentration was measured by Bradford assay (BioRad, 5000006). The lysates were mixed with water and 5x SDS loading buffer, run on SDS-PAGE gels, and subjected to Western blotting.

Analysis of mRNA and eRNA Expression by RT-qPCR

RNA expression analysis by RT-qPCR was performed essentially as previously described (Hah et al., 2013). Total RNA was isolated using TRI Reagent (Sigma, T9424) and treated with RQ-1 DNase (Promega, M6101). 1 μg of total RNA was reverse transcribed using random hexamers (Roche. 11034731001) and M-MLV reverse transcriptase (Promega, M1705). mRNA and eRNA expression was analyzed by quantitative PCR (qPCR) with SYBR Green (Lonza, 50512) using a LightCycler 480 (Roche) and the following primers:

<u>MDA-MB-231 cells</u>

- *OTUB2* eRNA Forward 5’-GGAATTCCCAAAGAGCAAA-3’
- *OTUB2* eRNA Reverse 5’-TCTCGCCTGTGATGACTCAG-3’
- *OTUB2* mRNA Forward 5’-TCAGCAAAAGGTTCACCGC-3’
- *OTUB2* mRNA Reverse 5’-GTAGGAATAGCCCAAGGCCC-3’
- *TGFA* eRNA Forward 5’-TTTCTGTTCCTGGCTTGGCA-3’
- *TGFA* eRNA Reverse 5’-AGCCAGGTGACCTAGTGGTA-3’

- *TGFA* mRNA Forward 5’-GACTGGTCCCCCTTTCATGG-3’
- *TGFA* mRNA Reverse 5’-TCGTGAGCCCTCGGTAAGTA-3’
- *CR595588* eRNA Forward 5’-GCTCCAGGCAGTGTAGGAAG-3’
- *CR595588* eRNA Reverse 5’-AGACTCTGTTGGCCCTGTTG-3’
- *CR595588* mRNA Forward 5’-AGTCGGTGGGGTGTGAGTTA-3’
- *CR595588* mRNA Reverse 5’-TTGGGAAGCGTGGGTTATGT-3’
- *HK1* eRNA Forward 5’-AATTTCAGGGGAAGCCTGGG-3’
- *HK1* eRNA Reverse 5’-GACTCTCTGGCAGTCACACC-3’
- *HK1* mRNA Forward 5’-ACGTGTCCTTCCTCCTGTCT-3’
- *HK1* mRNA Reverse 5’-GATCCCGGACTCTTAGCTGC-3’
- *C2orf18* eRNA Forward 5’-TCCACATGGTTGTCTCTGCC-3’
- *C2orf18* eRNA Reverse 5’-TGCCTGAAGCTTGACCTCTG-3’
- *C2orf18* mRNA Forward 5’-ACCTGCCTGCCTAGAGAACT-3’
- *C2orf18* mRNA Reverse 5’-CAACGCCCAGGATACCAGAA-3’

<u>MCF-7 cells</u>

- *GREB1* eRNA Forward 5’-TGCTGGCTGCTTAAAAACCT-3’
- *GREB1* eRNA Reverse 5’-TGAAAACCCACACTTCCAAA-3’
- *GREB1* mRNA Forward 5’-CCTATTTTGGAATAAAAACTGACC-3’
- *GREB1* mRNA Reverse 5’-GGGGAGAATGACACAAAAGC-3’
- *P2RY2* eRNA Forward 5’-AAGGGCTAATGTTTGGCACA-3’
- *P2RY2* eRNA Reverse 5’-AGGGAGGTCCAGGAGGTCTA-3’
- *P2RY2* mRNA Forward 5’-CGGTGGACTTAGCTCTGAGG-3’

- *P2RY2* mRNA Reverse 5’-GCCTCCAGATGGGTCTATGA-3’
- *PGR* eRNA Forward 5’-CATTGAGTCATGGCCTTTGAT-3’
- *PGR* eRNA Reverse 5’-CCTTTCAGATGGGAGCTAGG-3’
- *PGR* mRNA Forward 5’-TTGCCAAGAAGGTGAAACTG-3’
- *PGR* mRNA Reverse 5’-CTTTGCATTGTCACCCCATC-3’
- *SBNO2* eRNA Forward 5’-CCTGTATTCTGGGGGCACTA-3’
- *SBNO2* eRNA Reverse 5’-CTCACCCCATCCAGTACACC-3’
- *SBNO2* mRNA Forward 5’-GACTGGGCACCCACAAGGGC-3’
- *SBNO2* mRNA Reverse 5’-GGAAGGGCTGGGGGAGGGAG-3’
- *SMAD7* eRNA Forward 5’-GGCATAGCTAGGACCTCACC-3’
- *SMAD7* eRNA Reverse 5’-GAGGGAGGAAAGTGGCTTCT-3’
- *SMAD7* mRNA Forward 5’-AAGAGAAGCATTCTCATTGGAAA-3’
- *SMAD7* mRNA Reverse 5’- TCAGGAGTCCTTTCTCTCTCAAA-3’

Chromatin Immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) was performed as previously described with some modifications (Hah et al., 2013). Cells were grown to ~80% confluence and were crosslinked with 1% formaldehyde in PBS at 37°C for 10 min. The crosslinking reaction was quenched by adding glycine to a final concentration of 125 mM. The cells were collected by scraping in 1x PBS containing 1x complete protease inhibitor cocktail (Roche, 11697498001). The cells were pelleted by brief centrifugation in a microcentrifuge and lysed by pipetting in Farnham Lysis Buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40, 1 mM

DTT, 1x complete protease inhibitor cocktail). The nuclei were collected by brief centrifugation in a microcentrifuge and resuspended in SDS Lysis Buffer (Tris•HCl pH 7.9, 1% SDS, 10 mM EDTA, 50 mM, 1 mM DTT, 1x complete protease inhibitor cocktail) by pipetting and incubation on ice for 10 min. The chromatin was then sheared to ~200-500 bp DNA fragments by sonication using a Bioruptor Plus sonicator (Diagenode) for 25-30 cycles of 30 seconds on and 30 seconds off. Protein concentrations in the solubilized chromatin were determined using the BCA Protein Assay Kit (Pierce, 23225), and a normalized amount of soluble chromatin was precleared with Protein A or Protein G Dynabeads (Invitrogen, 10001D and 10003D, respectively) or Protein A or Protein G agarose beads (Millipore, 16-125 and Invitrogen, 15920010, respectively), and incubated overnight with 8 μL of polyclonal antiserum or 2.5 to 5.0 μg of commercial antibody.

The immune complexes from the ChIP were precipitated by the addition of Protein A or Protein G Dynabeads or agarose beads (depending on the antibody used) and washed once with each of the following wash buffer in sequence: (1) Low Salt Wash Buffer (20 mM Tris•HCl pH 7.9, 2 mM EDTA, 125 mM NaCl, 0.05% SDS, 1% Triton X-100, 1x complete protease inhibitor cocktail); (2) High Salt Wash Buffer (20 mM Tris•HCl pH 7.9, 2 mM EDTA, 500 mM NaCl, 0.05% SDS, 1% Triton X-100, 1x complete protease inhibitor cocktail); (3) LiCl Wash Buffer (10 mM Tris•HCl pH 7.9, 1 mM EDTA, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1x complete protease inhibitor cocktail); and (4) 1x Tris-EDTA (TE) containing 1x complete protease inhibitor cocktail. The precipitated immune complexes were transferred to a new tube in 1x TE/complete protease inhibitor cocktail before elution of the genomic DNA fragments in Elution Buffer [40 mM Tris•HCl pH 7.9, 10

mM EDTA, 100 mM NaCl, 100 mM $NaHCO_3$, 1% SDS, and 50 μg Proteinase K (Life Technologies, 2542)] for 2 hours at 55°C. H3K27ac ChIP assays were performed in the presence of the deacetylase inhibitors nicotinamide (5 mM) and sodium butyrate (10 mM) until the elution of the ChIPed DNA. The crosslinks were reversed by incubating overnight at 65 °C, and the genomic DNA was purified using phenol:chloroform:isoamyl acid extraction (Sigma, P2069) followed by ethanol precipitation. The ChIPed DNA was (1) analyzed by qPCR with SYBR Green (Lonza, 50512) using a LightCycler 480 (Roche) and the primers listed below or (2) subjected to ChIP-seq library preparation as described below. Non-specific background signals in all ChIP assays were determined using IgG (for purified antibodies) or no antibody control (for antiserum). The data were expressed as the percent of input or relative enrichment (fold change).

ChIP using the p300 BRP-Fc fusion protein was performed using the standard ChIP-seq protocol described above except that the antibody addition was omitted. Instead, chromatin-associated p300 BRP-Fc was directly precipitated with Protein A Dynabeads using the rabbit Fc tag.

Primers for ChIPing non-histone proteins

- *OTUB2* enhancer Forward 5’-GGAATTCCCAACAGAGCAAA-3’
- *OTUB2* enhancer Reverse 5’-TCTCGCCTGTGATGACTCAG-3’
- *TGFA* enhancer Forward 5’-GGAGAAAGGAGGTGGAACGG-3’
- *TGFA* enhancer Reverse 5’-GACTCAAAGTGACAGGGGCA-3’
- *CR595588* enhancer Forward 5’-GTCACTTGTTCTCCTGCGTG-3’

- *CR595588* enhancer Reverse 5’-GGGAAGCAGTGCTCATCCAG-3’
- *HK1* enhancer Forward 5’-CCCTCCTGAATGACAGATGG-3’
- *HK1* enhancer Reverse 5’-CTGCCTGACTCACACTGGAA-3’
- *C2orf18* enhancer Forward 5’-CATGTGACCCCAAAGAGGAG-3’
- *C2orf18* enhancer Reverse 5’-CATCCAGGCTTAACCAGAGG-3’
- *GREB1* enhancer Forward 5’-TGCTGGCTGCTTAAAAACCT-3’
- *GREB1* enhancer Reverse 5’-TGAAAACCCACACTTCCAAA-3’
- *P2RY2* enhancer Forward 5’-AAGGGCTAATGTTTGGCACA-3’
- *P2RY2* enhancer Reverse 5’-AGGGAGGTCCAGGAGGTCTA-3’
- *PGR* enhancer Forward 5’-ATGCAGAGCCATTGCAAAAT-3’
- *PGR* enhancer Reverse 5’-ATGCAGAGCCATTGCAAAAT-3’
- *SBNO2* enhancer Forward 5’-CCTGTATTCTGGGGGCACTA-3’
- *SBNO2* enhancer Reverse 5’-CTCACCCCATCCAGTACACC-3’
- *SMAD7* enhancer Forward 5’-GGCATAGCTAGGACCTCACC-3’
- *SMAD7* enhancer Reverse 5’-GAGGGAGGAAAGTGGCTTCT-3’

<u>Primers for ChIPing H3K27ac</u>

- *OTUB2* enhancer Forward 5’-CCGAGCCTCTCCTCATTTCC-3’
- *OTUB2* enhancer Reverse 5’-CCATCAATGGTGGCAGGAGA-3’
- *TGFA* enhancer Forward 5’-TTTCTGTTCCTGGCTTGGCA-3’
- *TGFA* enhancer Reverse 5’-AGCCAGGTGACCTAGTGGTA-3’
- *CR595588* enhancer Forward 5’-ACAGGGCCAACAGAGTCTTG-3’

- *CR595588* enhancer Reverse 5'-CATGCTGCACACAGATCACG-3'
- *HK1* enhancer Forward 5'-TGCTGACAATCCAGCAAGGAA-3'
- *HK1* enhancer Reverse 5'-GATTTACTCGGAGAGTGCCCC-3'
- *C2orf18* enhancer Forward 5'-AACACAGGACAAGGGAGCAG-3'
- *C2orf18* enhancer Reverse 5'-GGGGTCAGGCAGACACATAC-3'

ChIP-seq Library Preparation

ChIP-seq libraries were prepared as previously described with some slight modifications (Luo et al., 2014). Briefly, input DNA or ChIPed DNA was subjected to additional purification using Agencourt AMPure XP beads (Beckman Coulter, A63881). Two and a half to 10 ng of purified genomic DNA was subjected to end repair using an end-repair mix (Enzymatics, Y9140-LC-L) and 0.1 mM dNTPs. A single dA base was added to the end repaired DNA using Klenow 3′→5′ Exo-minus (Enzymatics, P7010-HC-L), 1x Blue Buffer (Enzymatics, B0110), and 0.2 mM dATP to facilitate adaptor ligation. TruSeq DNA Sample Prep Kit adaptors (custom synthesized by IDT and HPLC purified) were partially annealed to form double-stranded adaptors on one terminus, and annealed to add DNA through double-stranded DNA-DNA ligation using T4 DNA ligase (Enzymatics, L6030-HC-L) in DNA Rapid Ligase Buffer (Enzymatics, B1010). The adaptor-ligated DNA was amplified by PCR for 6 to 11 cycles and purified by electrophoresis on a 1% agarose gel containing SYBR Gold (Invitrogen, S11494). The DNA was excised from the agarose gel and eluted using a QIAquick Gel Extraction Kit (Qiagen, 28706). The quality of the library was assessed using a D1000 ScreenTape (Agilent, 5067-5582) on a 2200 TapeStation

(Agilent) and quantified using a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32851). The libraries with unique adaptor barcodes were multiplexed and sequenced on an Illumina HiSeq 2000 (single-end, 50 base reads).

Analysis of ChIP-seq Data

Quality Control and Alignment. Quality of the ChIP-seq datasets was assessed using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The ChIP-seq reads were aligned to the human reference genome (hg19) using BOWTIE (version 0.12.7) with the default parameters (Langmead et al., 2009). Uniquely mapped reads were visualized on the UCSC genome browser as bigWig files prepared using BEDTools (Quinlan and Hall, 2010) and custom R scripts (available on request).

Peak Calling. ERα peak calling from a total of four replicates per condition in two sequencing runs was performed using MACS software (Zhang et al., 2008) using the default p-value and input condition as a control. Peak calling for wild-type (Wt) and L540Q (LQ) ERα was performed separately. The peak calling was performed separately for each sequencing run, and the peak calls from the two runs were compared to yield a set of common peaks. The final universe of ERα peaks from Wt and LQ were then compared using the mergePeaks function in HOMER (Heinz et al., 2010) to identify ERα peaks shared between the Wt and LQ conditions. The read counts ± 500 bp around the summit of the common ERα peaks were used to calculate the fold change (FC) where,

$$FC = Wt/LQ$$

In the subsequent analyses, a set of common ERα peaks with approximately equivalent binding strength in Wt and LQ was selected using the cut off $0.75 < FC < 1.25$. This set of ERα peaks was used to determine differences between Wt and LQ in the recruitment of coregulators or the enrichment of H3K27ac. For boxplot representations of ChIP-seq data, read counts ± 500 bp surrounding the summit of the common ERα peaks were counted and visualized.

Comparing Called Peaks. Peaks for wild-type (Wt) and L540Q (LQ) ERα were rank ordered separately by the number of read counts in a fixed 1 kb window (± 500 bp) surrounding each ERα peaks. The genomic coordinates of the top 6500 peaks for each ERα (i.e., ERαWt or ERαLQ) were compared. The overlap of the ERαWt and ERαLQ peaks were then visualized in a Venn diagram.

Chromosome Conformation Capture (3C) Assays

Chromosome conformation capture (3C) assays were performed as previously described with some modifications (Dekker et al., 2002; Wang et al., 2009). The cells in one 10 cm diameter plate at ~80% confluence (~4 x 10^6 cells) were crosslinked in 1% formaldehyde in PBS at 37°C for 10 min. Crosslinking was quenched by the addition of glycine to a final concentration of 125 mM. The cells were scraped and collected in PBS containing 1x complete protease inhibitor cocktail (Roche, 11697498001), and then pelleted by centrifugation in a microfuge at 4°C. The cells were resuspended in 3C Lysis Buffer (10 mM Tris•HCl pH7.5, 10 mM NaCl, 0.2% Triton X-100, 1 mM DTT, 1x complete protease inhibitor cocktail) and lysed by pipetting and gentle mixing at 4°C for 30 min. to release the

nuclei, which were collected by centrifugation in a microcentrifuge at 500 x g for 5 min. The pelleted nuclei were resuspended in 3C Digestion Buffer [1.2x NEB Restriction Enzyme Buffer 3 (NEB, B7003S), 0.3% SDS] and incubated at 37°C for 1 hour with gentle mixing. Triton X-100 was then added to the nuclear suspension to a final concentration of 2%. After continued incubation at 37°C for 1 hour with gentle mixing, 400 units *BglII* restriction enzyme (NEB, R0144S) were added and the chromatin was digested overnight at 37°C with gentle mixing. SDS was added to the *BglII*-digested chromatin to a final concentration of 1.6% and the mixture was incubated in 65°C for 20 min. to quench the digestion. The digested chromatin was then transferred to a 15 mL conical tube, diluted with 6.5 mL of 3C Ligation Buffer (50 mM Tris•HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 1% Triton X-100, 10 mM DTT), and incubated at 37°C for 1 hour with gentle mixing.

The digested chromatin was then ligated with 2000 units of T4 DNA Ligase (NEB, M0202L) at 16°C for 4 hours, followed by 30 min. at room temperature. The ligated chromatin was deproteinized by treatment with 300 μg Proteinase K (Sigma, P2308) at 37°C for 1 hour, reverse crosslinked at 65°C overnight, purified by phenol:chloroform:isoamyl acid (Sigma, P2069) extraction, and precipitated with ethanol. The 3C DNA was collected by centrifugation, resuspended in 1x TE, and further purified using a NucleoSpin Gel and PCR Clean-Up Kit (Takara, 740609.250) with buffer NTB (Takara, 740595.150) following the manufacturer's instructions. Bacterial artificial chromosomes (BAC) spanning the loci to be tested were subjected to digestion with *BglII*, ligation, phenol:chloroform:isoamyl acid purification, and ethanol precipitation to generate a standard curve to normalize the amplification efficiency for each primer set.

The purified 3C DNA was quantified using a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32851) and then normalized to (1) 2.5 ng/μL for qPCR assays with SYBR Green (Lonza, 50512) or (2) 25 ng/μL for qPCR assays with TaqMan probes (Thermo Fisher Scientific) using a QuantiTech Probe PCR Kit (Qiagen, 204343) with the primers and probes listed below. qPCR assays with SYBR Green were performed on the 3C DNA to quantify: (1) the input of the locus to be tested using the loading primer mix, (2) the input of the *GAPDH* locus using the *GAPDH* loading primer mix, and (3) the interaction frequencies of the *GAPDH* locus for normalization using the *GAPDH* interaction primer mix. A serial dilution of the 3C DNA was used to generate a standard curve for each primer set. qPCR assays with TaqMan probes were performed to quantify the interaction frequencies in the locus to be tested using the hub and each test primer. A standard curve for the TaqMan qPCR assay was generated using a serial dilution of the digested/ligated BAC control DNA. The interaction frequencies determined by the TaqMan qPCR assay were normalized to the input of the test locus, the input of the *GAPDH* locus, and the interaction frequencies of the *GAPDH* locus.

BAC templates:

Locus	*BAC Clone ID*
• *OTUB2*	RP11-666E24
• *TGFA*	RP11-771E16, RP11-36L20
• *CR595588*	CTD-2363K16
• *HK1*	RP11-652D17

• *C2orf18*	CTD-2246L24

TaqMan probes:

Locus	*Sequence (5' to 3')*
• *OTUB2*	TTCCTCTCCGGGCCTGACCT
• *TGFA*	ATCTAGGAAACCTCCGTGGGGCTAGTCT
• *CR595588*	CTCGTGAGGCTTATTCACTACCATGAGAACAGG
• *HK1*	TCTAGGATCACAGCTTGGATCTGTGAGTC
• *C2orf18*	ATCTTCAGTGTCCAGGAAGAAGGTACGG

Primers:

Locus	*Name*	*Sequence (5' to 3')*
• *GAPDH*	Interaction Forward (Wang et al., 2009)	CCTTCTCCCCATTCCGTCTT
	Interaction Reverse (Wang et al., 2009)	TGTGCGGTGTGGGATTGTC
	Loading Forward (Wang et al., 2009)	ACAGTCCATGCCATCACTGCC
	Loading Reverse (Wang et al., 2009)	GCCTGCTTCACCACCTTCTTG
• *OTUB2*	Loading Forward	CAGTAATGTTCTCAGACTTC
	Loading Reverse	CTAGAGCTCTGACTCCAC
	Hub	AGTCAGAGCTCTAGGGA
	Test 1	AGTCAGGCAGGGAGAT
	Test 2	GGGATTCACACCCAGAT
	Test 3	AAAGGAGGTGTCGTCTAG

	Test 4	ATGGGTTTGGAGCAGAT
	Test5	CGAAGAAAGAAGCCCTTTATAA
	Test 6	CTGGGGGCTTCTTAGAT
• *TGFA*	Loading Forward	CAATGCTCAGGTTCCAAGTA
	Loading Reverse	CTGTTAGGAGTCTCGGTTAATG
	Hub	AGACTCCTAACAGCCAGTT
	Test 1	CAGAGAGAAGGTGCTGTG
	Test 2	TGGATTCAAATCCAGGATCC
	Test3	TGAATCAGTAGTCGGAATATAGA
• *CR595588*	Loading Forward	ACAAAAGAGGAATGATGGCT
	Loading Reverse	CCAGGAAAGATGAGAAGCAT
	Hub	TGAATCATGGGGACAGTTTC
	Test 1	GCGACCTCTTTCTACCA
	Test 2	CCGCCTGGGATAAAAGTT
	Test 3	CATCTTTGGCCTTTCCAGA
	Test 4	GACCTTCTGCTCTTAAGAAAAC
• *HK1*	Loading Forward	GCATCTAAGCTCCTCCTTTT
	Loading Reverse	CTCTACCCTAGCTCTTGACT
	Hub	TCTATACAACTGGGACCAC
	Test1	ATAGCTTCTCTTGAAAGATTTAGA
	Test2	GGTAGTAGACACTTCTAAACAAC
	Test 3	CCTTGCATGAGCCACAC

	Test 4	CTCACTGTCAAGTTATCAAGAA
• *C2orf18*	Loading Forward	TCTTCTGTGTCCTTTCTGTG
	Loading Reverse	ACTCTACTACACTGTCCTCC
	Hub	TGTAAAATGGACTTGGTGAT
	Test 1	ATTCTGTGCCTGCAAAGAA
	Test 2	GAAACTACGGAGTGTGTTTG
	Test 3	ATAGGATGGAAACCACCAGT
	Test 4	TCTTTGATCAGGGTCAGGT
	Test 5	CCCTGCTCACATCTCCTAA

Preparation of Global Run-on Sequencing (GRO-seq) Libraries

GRO-seq libraries were prepared as previously described with some modifications (Luo et al., 2014). Approximately 2 x 10^6 cells were seeded on a 15 cm^2 plate and grown to ~80% confluence. The cells were washed twice with ice-cold PBS and collected by scraping in Swelling Buffer [10 mM Tris•HCl pH 8.0, 2 mM MgOAc, 3 mM $CaCl_2$, 0.5 mM DTT, 1x complete protease inhibitor cocktail, and SUPERase•In (Ambion)] on ice. The cells were pelleted by centrifugation and resuspended in Lysis Buffer [10 mM Tris•HCl pH 8.0, 2 mM MgOAc, 3 mM $CaCl_2$, 10 mM NaCl, 0.5% NP-40, 300 mM sucrose, 0.5 mM DTT, 1x complete protease inhibitor cocktail, and SUPERase•In (Ambion)]. To isolate the nuclei, the swollen cells were lysed by pipetting up and down 60 times through a narrow bore tip. The nuclei were pelleted by brief centrifugation, equilibrated, and dispersed in Freezing Buffer to

5 x 10^6 nuclei/100 µL [50 mM Tris•HCl pH 8.3, 5 mM $MgCl_2$, 40% glycerol, 0.1 mM EDTA, and SUPERase•In (Ambion)], aliquoted, and stored at -80°C.

Run-on was performed as previously described with some modifications (Luo et al., 2014). The nuclei were incubated in the presence of 5'-bromo-UTP and α-^{32}P radiolabeled-CTP for 5 min., with subsequent quenching by the addition of RQ-1 DNase (Promega, M6101). RNA was isolated from the run-on reaction mixture and hydrolyzed by incubating with 0.2 N NaOH on ice for 15 min. The fragmented RNA was treated with T4 PNK (NEB, M0201S) in the absence of ATP to dephosphorylate the 3'-terminus. Nascent transcripts were isolated from the total RNA with two rounds of affinity purification using anti-BrdU antibody-conjugated agarose beads (Santa Cruz Biotechnology, sc-32323). PolyA tails were added to the nascent transcripts using 1 mM ATP and 0.5 U/µL *E. coli* Poly(A) Polymerase (NEB, M0276S) at 37°C for 8 min. cDNA was generated using the oNTI223HIseq primer (custom synthesized by IDT and HPLC purified) (Ingolia et al., 2009) and purified on an 8% polyacrylamide TBE-urea gel. The purified cDNA was then circularized using CircLigase (Epicentre, CL4111K), relinerized using APE1 (NEB, M0282S), and amplified using TruSeq small RNA-seq PCR primers (custom synthesized by IDT and HPLC purified) to generate the GRO-seq library. The library quality was assessed using a D1000 ScreenTape (Agilent, 5067-5582) on a 2200 TapeStation (Agilent) and quantified using a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32851). The libraries with unique adaptor barcodes were multiplexed and sequenced on an Illumina HiSeq 2000 (single-end, 50 bp reads) for a total of ~47 million raw reads per biological replicate. The experiment has two biological and two

technical replicates. The technical replicates were merged into a single data file before further analysis after confirming strong positive correlations between them.

Analysis of GRO-seq Data and Integration with ChIP-seq Data

GRO-seq data analyses were performed as previously described (Hah et al., 2013; Nagari et al., 2017) using the groHMM software package available from Bioconductor (http://bioconductor.org/packages/release/bioc/html/groHMM.html) (Chae et al., 2015; Danko et al., 2014).

Quality Control. Quality of the GRO-seq datasets was assessed using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The GRO-seq reads were subjected to trimming using Cutadapt (Martin, 2011) to remove the polyA tails and adapter sequences to maximize the mappability.

Alignment and Gene Annotation. The trimmed GRO-seq reads were aligned to the human reference genome (hg19) and one complete copy of an rDNA repeat (GenBank ID: U13369.1) using the BWA aligner (Li and Durbin, 2010). The mapped reads were visualized as UCSC genome browser tracks after conversion to bigWig files using the groHMM package (Chae et al., 2015; Danko et al., 2014). A complete set of annotated coding genes was assembled from the RefSeq, UCSC, and Gencode databases. Overlapping annotations were removed from the assembled set of genes to avoid multiple counting. The set of annotated long non-coding RNA (lncRNA) genes was downloaded from the LNCipedia 2.0 database. The lists of coding and lncRNA gene annotations were combined and used in the subsequent analysis.

Calling Differential Gene Expression. Differential gene expression between experimental conditions was determined using edgeR (Robinson et al., 2010). Expression levels in a given condition were determined using the number of read counts between +1 to +13 kb from the 5' end of the gene annotations (Hah et al., 2011). Significantly regulated genes were determined using a false discovery rate (FDR) cutoff of 5%. The read counts +1 to +13 kb from the 5' end of each gene were visualized in boxplots as Reads Per Kilobase of gene per Million mapped reads (RPKM).

Calculation of Pausing Indexes. The read counts in windows from the TSS to +1 kb and from +1 kb to the 3' end to the end of the gene were collected as "TSS" and "gene body," respectively, and represented in boxplots as Reads Per Kilobase of gene per Million mapped reads (RPKM). Pausing indexes were calculated using the pausing index function in the groHMM software package (Chae et al., 2015).

Box Plots. For ChIP-seq, box plot representations were used to quantitatively represent the read distribution in a fixed 1 kb window (± 500 bp) surrounding each ERα binding site using the box plot function in R. Wilcoxon rank sum tests were performed to determine the statistical significance of all comparisons.

Metaplots. Metaplots of GRO-seq data were generated using the metagene function in the groHMM software package (Chae et al., 2015)

Analysis of ERα Binding Sites Nearest to Regulated Genes. The ERα peaks from the ChIP-seq analyses that were (1) present in both the Wt and LQ conditions and (2) located nearest to the genes up-regulated in the Wt 45 min. E2 condition were obtained using custom Perl scripts (available on request). The GRO-seq RPKM values of the gene (i.e., gene

transcription) and the ERα peaks (i.e., enhancer transcription) were represented as box plots and line plots (average RPKM).

Cell Proliferation Assays

On day -2 (minus 2), MCF-7 cells were seeded at 4,000 cells per well in 24-well plates in minimal essential medium Eagle (Sigma, M3024) supplemented with 5% charcoal-dextran treated calf serum (Sigma, C8056) and allowed to rest at 37°C, 5% CO_2 for 16 hours. On day -1 (minus 1), the specified siRNAs were transfected using Lipofectamine RNAiMax (Thermo Fisher Scientific, 13778150) according to the manufacturer's protocol. On day 0, the cells were treated with control vehicle (DMSO), 4 μM C646, or 10 μM SGC-CBP-30 ± 100 nM E2. The cells were collected every two days, fixed in 10% formaldehyde for 10 min., and stored at 4°C. After the final time point was collected, all samples were stained with 0.1% crystal violet in 75 mM phosphoric acid for 30 min. After washing with a large volume of water, the crystal violet was extracted from the cells using 10% acetic acid and measured as absorbance at 562 nm.

Analysis of Somatic Mutations, Copy Number Alterations, and Gene Expression Using TCGA Data Sets

Relevant gene sets and expression data were accessed from The Cancer Genome Atlas (TCGA) (http://www.cbioportal.org) (Cerami et al., 2012; Gao et al., 2013; Network, 2012). Bar graphs showing the somatic mutations and copy number alterations in the *NCOA1* (SRC1), *NCOA2* (SRC2), and *NCOA3* (SRC3) genes were generated using the

OncoPrint tool on cBioPortal (http://www.cbioportal.org/index.do) and the METABRIC breast cancer cohort of TCGA datasets (Network, 2012; Pereira et al., 2016). The ER-positive dataset was selected for ER status in the clinical attributes on the study summary page, the genes of interest were specified in the query window, and OncoPrint was generated in response to the query (Cerami et al., 2012; Gao et al., 2013; Network, 2012).

For mRNA expression analysis in relation to copy number alterations, the clinical attributes, copy number alterations, and mRNA expression data from the METABRIC breast cancer cohort of TCGA (Network, 2012; Pereira et al., 2016) were downloaded from cBioPortal (Cerami et al., 2012; Gao et al., 2013) as separate files. The following analysis was performed using custom R scripts (available on request). The datasets on separate files were merged based on patient IDs, and ER status in the clinical attributes was used to sort and extract the data for ER-positive samples. The datasets of the ER-positive breast cancer samples were binned based on the copy number of the *NCOA2* (SRC2) or *NCOA3* (SRC3) genes, and the z-score for the expression of the corresponding mRNA were visualized in boxplots. Due to the low number of samples with *NCOA1* (SRC1) copy number alterations, the mRNA expression analysis for *NCOA1* based on the copy number alteration was omitted.

Kaplan-Meier Analyses

Kaplan-Meier analyses were performed using the Genotype 2 Outcome (G-2-O) online tool (http://www.g-2-o.com/) (Pongor et al., 2015). First, a list of genes more highly expressed in Luminal A or B breast cancer patient samples with amplification of *SRC2* or *SRC3*, compared to samples without amplification of *SRC2* or *SRC3*, was prepared using

TCGA datasets, namely *SRC2* or *SRC3* signature gene sets. Next, the expression of the genes in the *SRC2* or *SRC3* signature gene sets were determined using curated publicly available microarray gene expression datasets linked to clinical outcomes (Pongor et al., 2015). The patients were stratified based on the top 50% (high) or the bottom 50% (low) of expression for each of the *SRC2* or *SRC3* signature genes, and the survival rates were calculated for each group of patients for each signature gene. The overall survival rates were plotted as a metagene (average) of all the signature genes to give higher statistical power in the analysis (Pongor et al., 2015).

Genomic Data Sets

The ChIP-seq and GRO-seq datasets generated from MDA-MB-231-ERα Wt and MDA-MB-231-ERα L540Q cells for the current study were deposited in the NCBI's Gene Expression Omnibus (GEO) database under the super series accession number GSE95123. The ChIP-seq and GRO-seq datasets can be found under subseries GSE95121 and GSE95122, respectively.

3.6 References

Acevedo, M.L., and Kraus, W.L. (2003). Mediator and p300/CBP-steroid receptor coactivator complexes have distinct roles, but function synergistically, during estrogen receptor alpha-dependent transcription with chromatin templates. Mol Cell Biol *23*, 335-348.

Acevedo, M.L., Lee, K.C., Stender, J.D., Katzenellenbogen, B.S., and Kraus, W.L. (2004). Selective recognition of distinct classes of coactivators by a ligand-inducible activation domain. Mol Cell *13*, 725-738.

Anzick, S.L., Kononen, J., Walker, R.L., Azorsa, D.O., Tanner, M.M., Guan, X.Y., Sauter, G., Kallioniemi, O.P., Trent, J.M., and Meltzer, P.S. (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science *277*, 965-968.

Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature *384*, 641-643.

Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theurillat, J.P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N.*, et al.* (2012). Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet *44*, 685-689.

Black, J.C., Choi, J.E., Lombardo, S.R., and Carey, M. (2006). A mechanism for coordinating chromatin modification and preinitiation complex assembly. Mol Cell *23*, 809-818.

Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA binding to CBP stimulates histone acetylation and transcription. Cell *168*, 135-149 e122.

Bowers, E.M., Yan, G., Mukherjee, C., Orry, A., Wang, L., Holbert, M.A., Crump, N.T., Hazzalin, C.A., Liszczak, G., Yuan, H.*, et al.* (2010). Virtual ligand screening of the p300/CBP histone acetyltransferase: identification of a selective small molecule inhibitor. Chem Biol *17*, 471-482.

Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.A., and Carlquist, M. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. Nature *389*, 753-758.

Burns, K.A., and Korach, K.S. (2012). Estrogen receptors and human disease: an update. Arch Toxicol *86*, 1491-1504.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E.*, et al.* (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov *2*, 401-404.

Chae, M., Danko, C.G., and Kraus, W.L. (2015). groHMM: a computational tool for identifying unannotated and cell type-specific transcription units from global run-on sequencing data. BMC Bioinformatics *16*, 222.

Chang, M.T., Asthana, S., Gao, S.P., Lee, B.H., Chapman, J.S., Kandoth, C., Gao, J., Socci, N.D., Solit, D.B., Olshen, A.B., *et al.* (2016). Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol *34*, 155-163.

Chen, W., and Roeder, R.G. (2011). Mediator-dependent nuclear receptor function. Semin Cell Dev Biol *22*, 749-758.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Dancy, B.M., and Cole, P.A. (2015). Protein lysine acetylation by p300/CBP. Chem Rev *115*, 2419-2452.

Danko, C.G., Chae, M., Martins, A., and Kraus, W.L. (2014). groHMM: GRO-seq Analysis Pipeline. In Bioconductor (http://bioconductor.org/packages/release/bioc/html/groHMM.html: Bioconductor).

Dasgupta, S., Lonard, D.M., and O'Malley, B.W. (2014). Nuclear receptor coactivators: master regulators of human health and disease. Annu Rev Med *65*, 279-292.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E., and Panne, D. (2013). Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. Nat Struct Mol Biol *20*, 1040-1046.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., *et al.* (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal *6*, pl1.

Gojis, O., Rudraraju, B., Gudi, M., Hogben, K., Sousha, S., Coombes, R.C., Cleator, S., and Palmieri, C. (2010). The role of SRC-3 in human breast cancer. Nat Rev Clin Oncol *7*, 83-89.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res *23*, 1210-1223.

Hay, D.A., Fedorov, O., Martin, S., Singleton, D.C., Tallant, C., Wells, C., Picaud, S., Philpott, M., Monteiro, O.P., Rogers, C.M., *et al.* (2014). Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. J Am Chem Soc *136*, 9308-9319.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol *16*, 144-154.

Ince, B.A., Zhuang, Y., Wrenn, C.K., Shapiro, D.J., and Katzenellenbogen, B.S. (1993). Powerful dominant negative mutants of the human estrogen receptor. J Biol Chem *268*, 14026-14032.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Jin, Q., Yu, L.R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.E., Wang, C., Brindle, P.K., Dent, S.Y., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J *30*, 249-262.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A.*, et al.* (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333-339.

Kang, Y.K., Guermah, M., Yuan, C.X., and Roeder, R.G. (2002). The TRAP/Mediator coactivator complex interacts directly with estrogen receptors alpha and beta through the TRAP220 subunit and directly enhances estrogen receptor function in vitro. Proc Natl Acad Sci U S A *99*, 2642-2647.

Kim, M.Y., Hsiao, S.J., and Kraus, W.L. (2001). A role for coactivators and histone acetylation in estrogen receptor alpha-mediated transcription initiation. EMBO J *20*, 6084-6094.

Kim, M.Y., Woo, E.M., Chong, Y.T., Homenko, D.R., and Kraus, W.L. (2006). Acetylation of estrogen receptor alpha by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor. Molecular endocrinology *20*, 1479-1493.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, T.K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. Cell *162*, 948-959.

Kraus, W.L., and Kadonaga, J.T. (1998). p300 and estrogen receptor cooperatively activate transcription via differential enhancement of initiation and reinitiation. Genes Dev *12*, 331-342.

Kulaeva, O.I., Nizovtseva, E.V., Polikanov, Y.S., Ulianov, S.V., and Studitsky, V.M. (2012). Distant activation of transcription: mechanisms of enhancer action. Mol Cell Biol *32*, 4892-4897.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497-501.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Li, H., and Chen, J.D. (1998). The receptor-associated coactivator 3 activates transcription through CREB-binding protein recruitment and autoregulation. J Biol Chem *273*, 5948-5954.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589-595.

Luo, X., Chae, M., Krishnakumar, R., Danko, C.G., and Kraus, W.L. (2014). Dynamic reorganization of the AC16 cardiomyocyte transcriptome in response to TNFalpha signaling revealed by integrated genomic analyses. BMC Genomics *15*, 155.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet *11*, 761-772.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal *17*.

Meerbrey, K.L., Hu, G., Kessler, J.D., Roarty, K., Li, M.Z., Fang, J.E., Herschkowitz, J.I., Burrows, A.E., Ciccia, A., Sun, T.*, et al.* (2011). The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. Proc Natl Acad Sci U S A *108*, 3665-3670.

Mihaly, Z., Kormos, M., Lanczky, A., Dank, M., Budczies, J., Szasz, M.A., and Gyorffy, B. (2013). A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. Breast Cancer Res Treat *140*, 219-232.

Nagari, A., Murakami, S., Malladi, V.S., and Kraus, W.L. (2017). Computational Approaches for Mining GRO-Seq Data to Identify and Characterize Active Enhancers. Methods Mol Biol *1468*, 121-138.

Network, C.G.A. (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61-70.

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

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283-293.

Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J.*, et al.* (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun *7*, 11479.

Pongor, L., Kormos, M., Hatzis, C., Pusztai, L., Szabo, A., and Gyorffy, B. (2015). A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. Genome Med *7*, 104.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Curr Opin Genet Dev *43*, 73-81.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Rodriguez, A.L., Tamrazi, A., Collins, M.L., and Katzenellenbogen, J.A. (2004). Design, synthesis, and in vitro biological evaluation of small molecule inhibitors of estrogen receptor alpha coactivator binding. J Med Chem *47*, 600-611.

Schodin, D.J., Zhuang, Y., Shapiro, D.J., and Katzenellenbogen, B.S. (1995). Analysis of mechanisms that determine dominant negative estrogen receptor effectiveness. J Biol Chem *270*, 31163-31171.

Sheppard, H.M., Harries, J.C., Hussain, S., Bevan, C., and Heery, D.M. (2001). Analysis of the steroid receptor coactivator 1 (SRC1)-CREB binding protein interaction interface and its importance for the function of SRC1. Mol Cell Biol *21*, 39-50.

Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., and Greene, G.L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell *95*, 927-937.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Song, X., Chen, J., Zhao, M., Zhang, C., Yu, Y., Lonard, D.M., Chow, D.C., Palzkill, T., Xu, J., O'Malley, B.W.*, et al.* (2016). Development of potent small-molecule inhibitors to drug the undruggable steroid receptor coactivator-3. Proc Natl Acad Sci U S A *113*, 4970-4975.

Spaeth, J.M., Kim, N.H., and Boyer, T.G. (2011). Mediator and human disease. Semin Cell Dev Biol *22*, 776-787.

Stenoien, D.L., Nye, A.C., Mancini, M.G., Patel, K., Dutertre, M., O'Malley, B.W., Smith, C.L., Belmont, A.S., and Mancini, M.A. (2001). Ligand-mediated assembly and real-time cellular dynamics of estrogen receptor alpha-coactivator complexes in living cells. Mol Cell Biol *21*, 4404-4412.

Stumpf, M., Waskow, C., Krotschel, M., van Essen, D., Rodriguez, P., Zhang, X., Guyot, B., Roeder, R.G., and Borggrefe, T. (2006). The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit Med1/TRAP220. Proc Natl Acad Sci U S A *103*, 18504-18509.

Tang, Z., Chen, W.Y., Shimada, M., Nguyen, U.T., Kim, J., Sun, X.J., Sengoku, T., McGinty, R.K., Fernandez, J.P., Muir, T.W.*, et al.* (2013). SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53. Cell *154*, 297-310.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Voegel, J.J., Heine, M.J., Tini, M., Vivat, V., Chambon, P., and Gronemeyer, H. (1998). The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. EMBO J *17*, 507-519.

Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet *15*, 69-81.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A.*, et al.* (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390-394.

Wang, F., Marshall, C.B., and Ikura, M. (2013). Transcriptional/epigenetic regulator CBP/p300 in tumorigenesis: structural and functional versatility in target recognition. Cellular and molecular life sciences : CMLS *70*, 3989-4008.

Wang, Q., Li, W., Zhang, Y., Yuan, X., Xu, K., Yu, J., Chen, Z., Beroukhim, R., Wang, H., Lupien, M.*, et al.* (2009). Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. Cell *138*, 245-256.

Yi, P., Wang, Z., Feng, Q., Pintilie, G.D., Foulds, C.E., Lanz, R.B., Ludtke, S.J., Schmid, M.F., Chiu, W., and O'Malley, B.W. (2015). Structure of a biologically active estrogen receptor-coactivator complex on DNA. Mol Cell *57*, 1047-1058.

Zeng, L., and Zhou, M.M. (2002). Bromodomain: an acetyl-lysine binding domain. FEBS letters *513*, 124-128.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Figure 3.1. Reduced SRC recruitment at ERα L540Q binding sites impairs enhancer formation and target gene transcription.

(A) SRC and p300 recruitment is impaired at ERα L540Q binding sites. ChIP-seq browser tracks for ERα, Med1, SRC (pan), and p300 at the *OTUB2* locus in MBA-MB-231 cells expressing ERα wild-type or ERα L540Q (231/ERαWt and 231/ERαLQ cells, respectively) ± E2 for 45 min.

(B) Box plots of ERα, Med1, SRC (pan), and p300 ChIP-seq read counts ± 45 min. of E2 treatment at ERα binding sites shared between 231/ERαWt and 231/ERαLQ cells. Box plots marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 2.2 \times 10^{-16}$; Wilcoxon rank-sum test).

(C) ERα L540Q exhibits impaired transcriptional activity. Venn diagram showing the number of genes significantly regulated by 45 min. of E2 treatment in 231/ERαWt and 231/ERαLQ cells as measured by GRO-seq. FDR = 0.05.

(D) Box plots showing the read counts for 367 genes up-regulated by wild-type ERα, but not by the L540Q mutant, upon E2 treatment as shown in panel (C). Box plots marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 1.54 \times 10^{-05}$; Wilcoxon rank-sum test).

(E) Browser tracks of the *OTUB2* locus in 231/ERαWt (*left*) and 231/ERαLQ (*right*) cells showing ERα ChIP-seq after 45 min. of E2 treatment and GRO-seq with a time course E2 treatment.

[Figure 3.1 is on the next page]

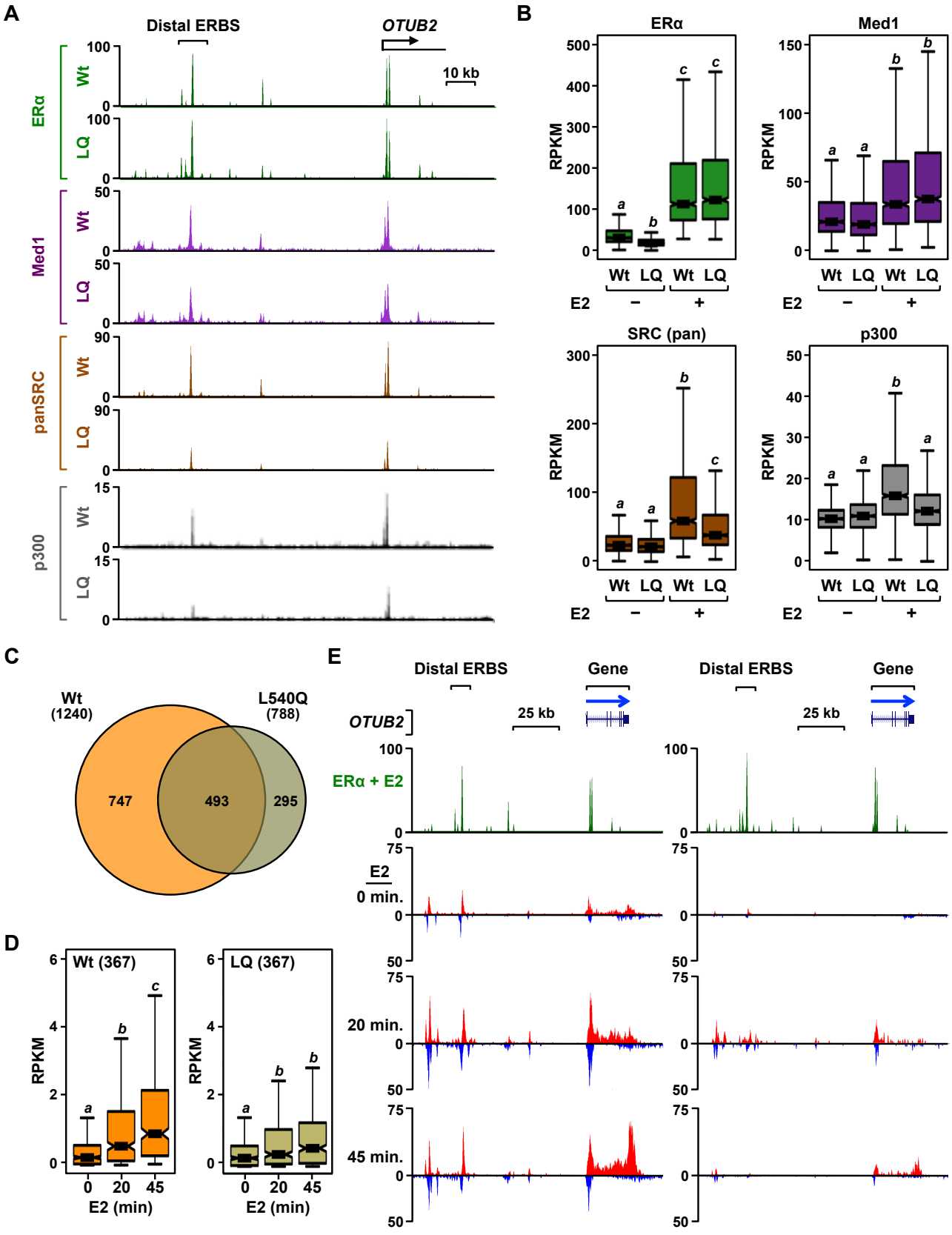

A
Distal ERBS
OTUB2
10 kb
ERα
Med1
panSRC
p300
Wt
LQ
B
ERα
Med1
SRC (pan)
p300
RPKM
Wt LQ Wt LQ
E2 – +
C
Wt (1240)
L540Q (788)
747
493
295
D
Wt (367)
LQ (367)
RPKM
E2 (min)
E
Distal ERBS
Gene
25 kb
OTUB2
ERα + E2
E2
0 min.
20 min.
45 min.

Figure 3.2. Selective impairment of enhancer features with ERα L540Q.
(A) Pol II recruitment is impaired at ERα L540Q binding sites. ChIP-qPCR assays for Pol II in 231/ERαWt and 231/ERαLQ cells treated ± E2 for 45 min. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Bars marked with different letters (*a*, *b*, *c*, *d*) are significantly different from each other ($p < 0.05$; two-way ANOVA).
(B) E2-dependent enhancer-promoter chromatin loop formation is maintained with ERα L540Q, in spite of impaired SRC recruitment. *(Top)* Browser tracks for ERα ChIP-seq and GRO-seq in 231/ERαWt cells after 45 min. of E2 treatment, shown with the location of the 3C primers. *(Bottom)* 3C-qPCR assays showing chromatin looping from a distal ERα binding site to the *OTUB2* promoter in 231/ERαWt and 231/ERαLQ cells ± 45 min. of E2 treatment. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to Wt at each genomic location in each condition ($p < 0.05$; two-way ANOVA).
(C) E2-stimulated H3K27ac levels are maintained with ERα L540Q, in spite of impaired SRC recruitment. ChIP-seq browser tracks for ERα and H3K27ac at the *OTUB2* locus in 231/ERαWt and 231/ERαLQ cells treated with E2 for 45 min.
(D) Box plots of H3K27ac ChIP-seq read counts at ERα binding sites shared between 231/ERαWt and 231/ERαLQ cells ± 45 min. of E2 treatment. Box plots marked with different letters (*a*, *b*) are significantly different from each other ($p < 2.2 \times 10^{-16}$; Wilcoxon rank-sum test).
(E) p300 HAT activity is required for E2-stimulated increases of H3K27ac levels. ChIP-qPCR assays for H3K27ac in 231/ERαWt cells treated ± E2 for 45 min. in the presence of the p300/CBP HAT inhibitor C646. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Bars marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 0.05$; two-way ANOVA).
(F) A table summarizing the features of ERα Wt and ERα L540Q binding sites upon 45 min. of E2 treatment.

[Figure 3.2 is on the next page]

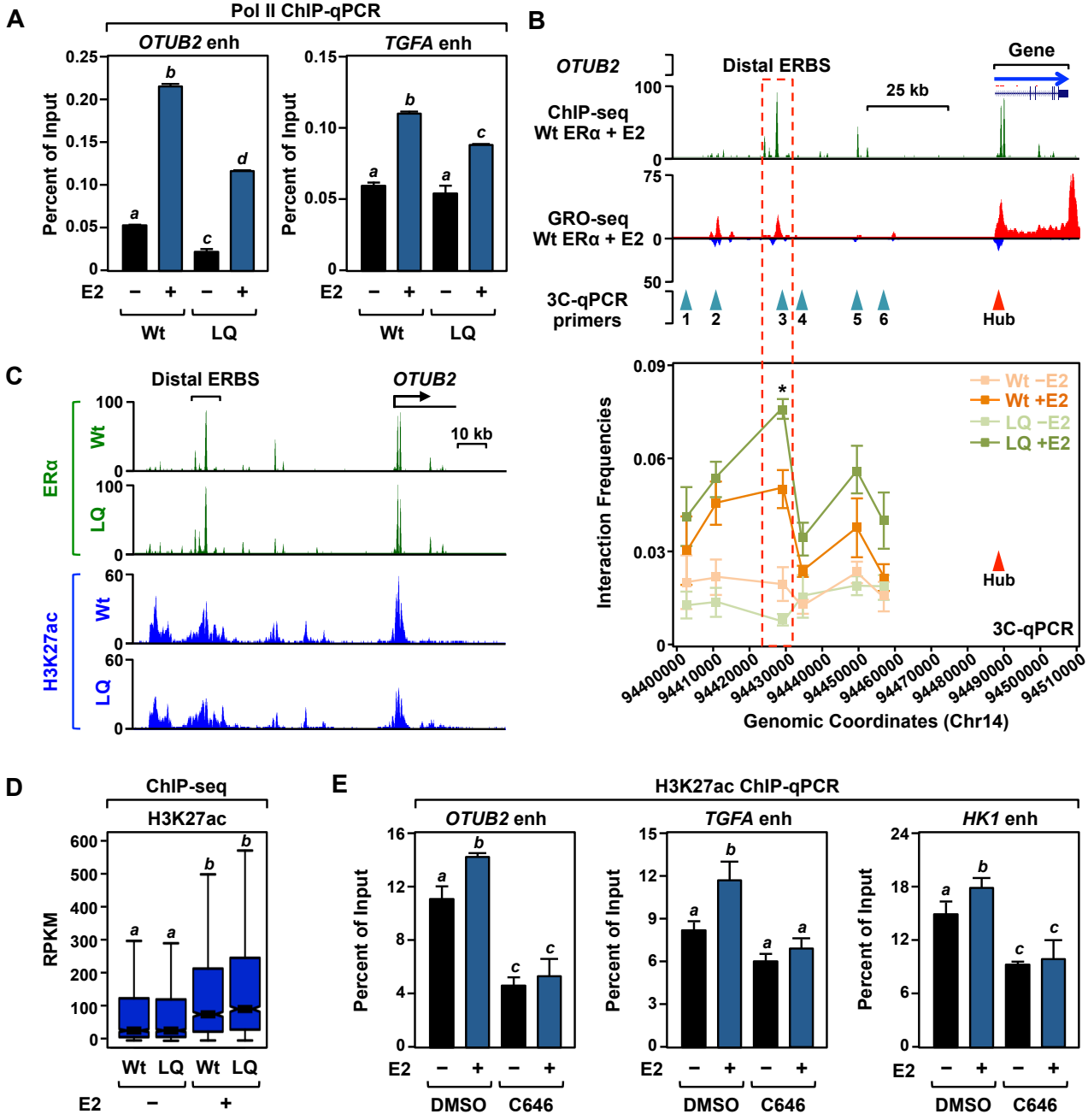

A
Pol II ChIP-qPCR
OTUB2 enh
TGFA enh
Percent of Input
E2
Wt
LQ
B
OTUB2
Distal ERBS
Gene
25 kb
ChIP-seq
Wt ERα + E2
GRO-seq
Wt ERα + E2
3C-qPCR
primers
Hub
Wt −E2
Wt +E2
LQ −E2
LQ +E2
Interaction Frequencies
3C-qPCR
Genomic Coordinates (Chr14)
C
Distal ERBS
OTUB2
10 kb
ERα
H3K27ac
Wt
LQ
D
ChIP-seq
H3K27ac
RPKM
E
H3K27ac ChIP-qPCR
OTUB2 enh
TGFA enh
HK1 enh
DMSO
C646
F
Ligand Binding
Chromatin Binding
Mediator Recruitment
SRC Recruitment
p300 Recruitment
H3K27ac Enrichment
Chromatin Looping
Pol II Recruitment
Transcription
Wt + + + + + + + + +
LQ + + + − − + + − −

Figure 3.3. Impaired SRC recruitment with ERα L540Q causes abortive enhancer formation and target gene transcription.
(A) Schematics of the selection pipeline for the genes and enhancers analyzed in panel (B). 367 genes up-regulated by E2 only in 231/ERαWt cells *(right)* were associated with the nearest ERαWt binding site *(left)*. GRO-seq data for the promoters of the genes and the nearest ERα binding sites were analyzed as described in panel (B).
(B) Box plots *(left and middle)* and line graphs *(right)* of GRO-seq normalized read counts for 367 genes up-regulated by ERα Wt, but not by ERα L540Q, upon E2 treatment (*top*) and the nearest ERα enhancers (*bottom*). In the box plots *(left and middle)*, bars marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 6 \times 10^{-09}$; Wilcoxon rank-sum test). In the line graphs *(right)*, each point represents the average read counts. Asterisks indicate significant differences compared to Wt at each time point ($p < 0.05$; Student's t-Test).
(C) p300 is recruited in an SRC-independent manner during the initial phase of enhancer formation ("enhancer priming"). ChIP-qPCR assays for ERα, SRC (pan), and p300 in doxycycline (Dox)-inducible 231/ERαWt and 231/ERαLQ cells treated with a time course of E2. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to Wt at each time point (*, $p < 0.05$; **, $p < 0.0005$; two-way ANOVA).

[Figure 3.3 is on the next page]

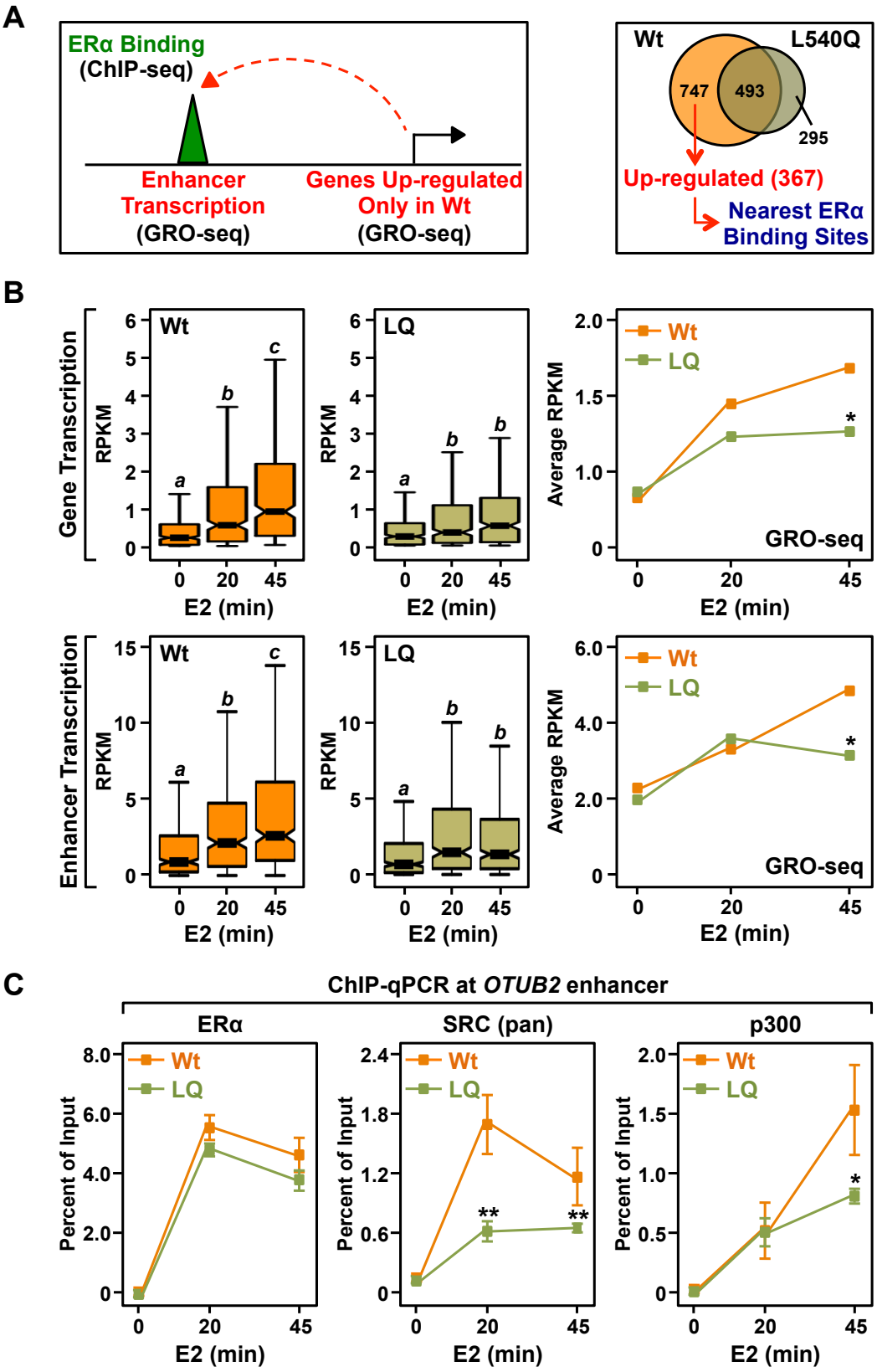

A
ERα Binding
(ChIP-seq)
Enhancer
Transcription
(GRO-seq)
Genes Up-regulated
Only in Wt
(GRO-seq)
Wt
L540Q
747
493
295
Up-regulated (367)
Nearest ERα
Binding Sites
B
Gene Transcription
RPKM
Wt
LQ
E2 (min)
Average RPKM
GRO-seq
Enhancer Transcription
C
ChIP-qPCR at OTUB2 enhancer
ERα
SRC (pan)
p300
Percent of Input

Figure 3.4. SRC-independent ERα enhancer priming requires Mediator.
(A) Schematic showing the distinct SRC-independent and SRC-dependent phases of p300 recruitment during enhancer formation.
(B) Western blot showing siRNA-mediated knockdown of Med1 in Dox-inducible 231/ERαWt cells.
(C) SRC-independent recruitment of p300 requires Mediator. ChIP-qPCR assays for Med1 and p300 in Med1-depleted Dox-inducible 231/ERαWt cells treated with a time course of E2. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the siRNA control at each time point (*, $p < 0.005$; **, $p < 0.0001$; two-way ANOVA).
(D) Mediator is required for E2-induced gene expression. RT-qPCR assays in Med1-depleted 231/ERαWt cells treated with a time course of E2. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the siRNA control at each time point (*, $p < 0.05$; **, $p < 0.005$; two-way ANOVA).

[Figure 3.4 is on the next page]

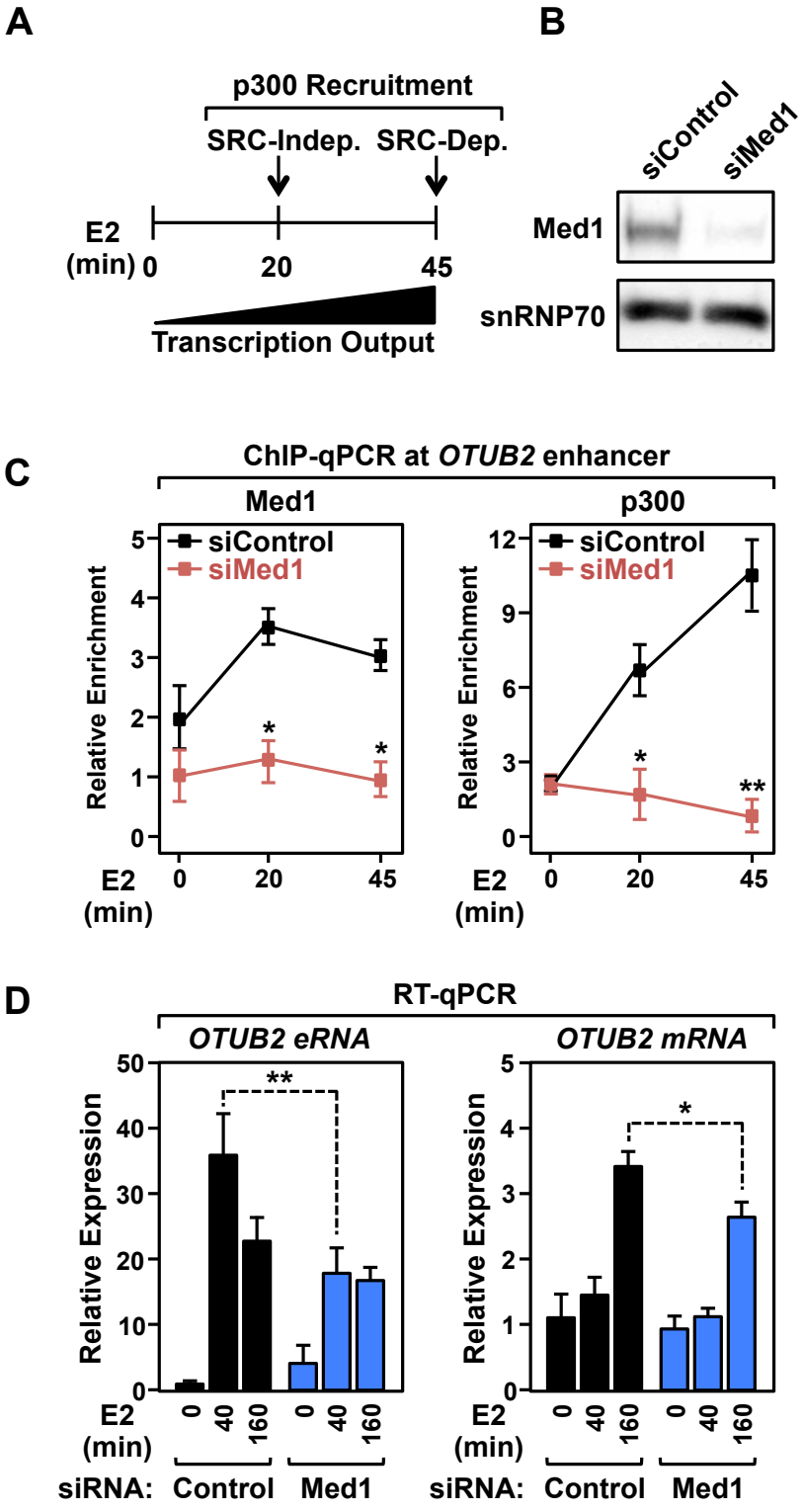

A
p300 Recruitment
SRC-Indep.
SRC-Dep.
E2
(min)
0
20
45
Transcription Output
B
siControl
siMed1
Med1
snRNP70
C
ChIP-qPCR at OTUB2 enhancer
Med1
p300
siControl
siMed1
Relative Enrichment
E2
(min)
D
RT-qPCR
OTUB2 eRNA
OTUB2 mRNA
Relative Expression
siRNA:
Control
Med1

Figure 3.5. The p300 bromodomain is required for ERα enhancer priming.
(A) The p300 bromodomain is required for p300 recruitment during enhancer priming. ChIP-qPCR assays for ERα, p300, and Med1 in MCF-7 cells treated with a time course of E2 in the presence of the p300/CBP bromodomain inhibitor SGC-CBP-30 (CBP30). Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the DMSO control at each time point (*, $p < 0.05$; **, $p < 0.005$; two-way ANOVA).
(B) Ectopic expression of the p300 Bromo-RING-PHD (BRP) cassette in MCF-7 cells. (*Top*) Schematics of p300 and the p300 BRP cassette: KIX, CREB interaction domain; Bromo, bromodomain; RING, ring domain; PHD, plant homeodomain; HAT, histone acetyltransferase catalytic domain; SID, SRC interaction domain. (*Bottom*) Western blot showing doxycycline (Dox)-dependent expression of the rabbit IgG Fc-fused p300-BRP cassette in MCF-7 cells.
(C) The p300 bromodomain is recruited to ERα binding sites during enhancer priming. ChIP-qPCR assays for the IgG Fc-fused p300-BRP cassette in MCF-7 cells treated with a time course of E2. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to 0 min. of E2 treatment ($p < 0.005$; one-way ANOVA).
(D) The p300 bromodomain and p300 acetyltransferase activity are required for E2-induced gene expression. RT-qPCR assays in MCF-7 cells treated with a time course of E2 in the presence of the p300 HAT inhibitor C646 or the bromodomain inhibitor SGC-CBP-30 (CBP30). Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the DMSO control at each time point (*, $p < 0.05$; **, $p < 0.005$; two-way ANOVA).

[Figure 3.5 is on the next page]

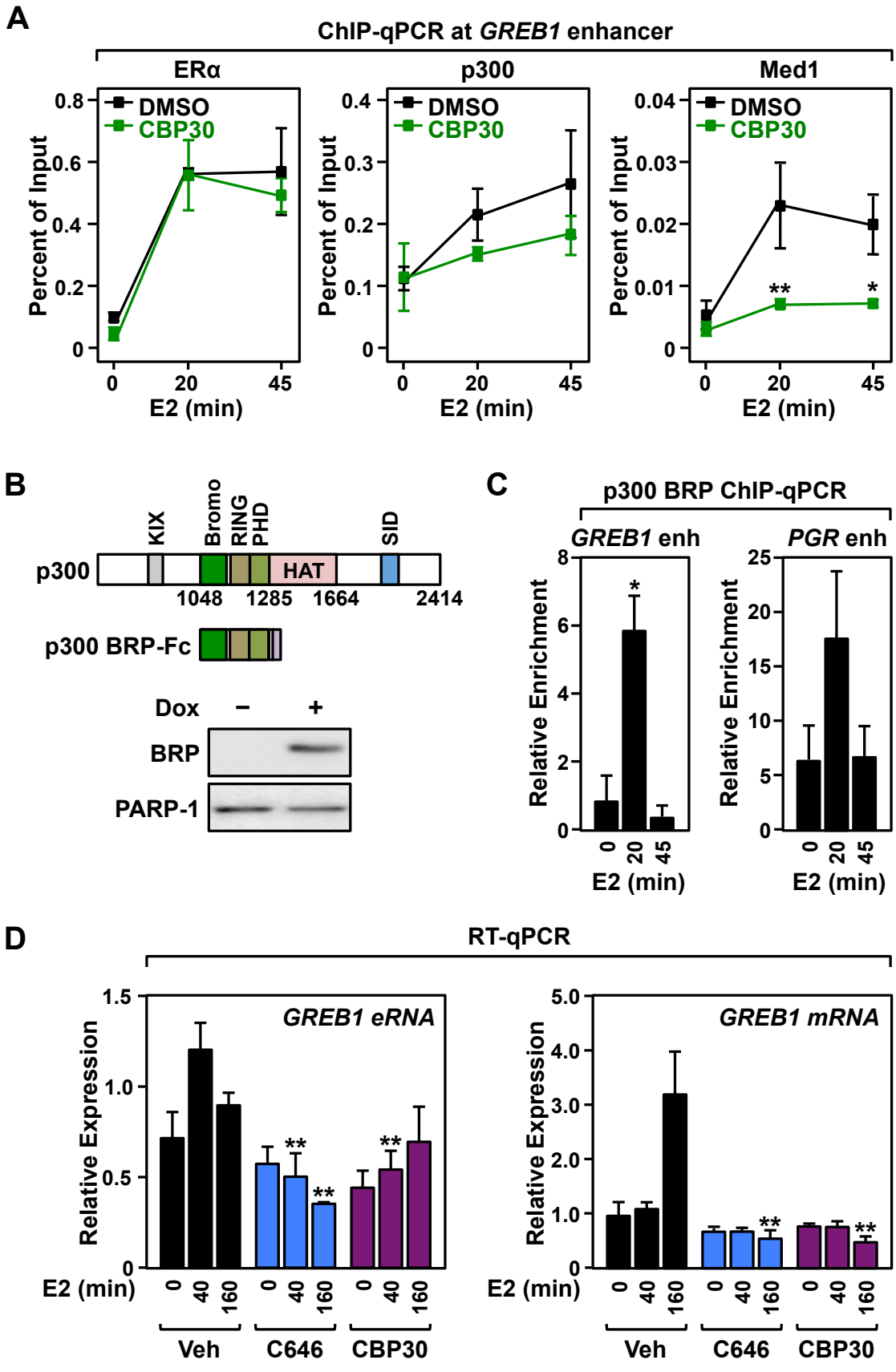

A
ChIP-qPCR at GREB1 enhancer
ERα
p300
Med1
DMSO
CBP30
Percent of Input
E2 (min)
B
KIX
Bromo
RING
PHD
SID
HAT
p300
1048
1285
1664
2414
p300 BRP-Fc
Dox
BRP
PARP-1
C
p300 BRP ChIP-qPCR
GREB1 enh
PGR enh
Relative Enrichment
D
RT-qPCR
GREB1 eRNA
GREB1 mRNA
Relative Expression
Veh
C646
CBP30

Figure 3.6. Forced recruitment of p300 to ERα binding sites promotes Mediator recruitment, enhancer formation, and target gene expression.
(A) Ectopic expression of ERα wild-type (Wt), ΔH12, and ΔH12-PID in MDA-MB-231 cells (231/ERαWt, 231/ERαΔH12, and 231/ERαΔH12-PID, respectively). *(Top)* Schematics of ERα Wt, ΔH12, and ΔH12-PID: AF-1, activation domain 1; DBD, DNA binding domain; LBD, ligand binding domain; AF-2, activation domain 2; PID, p300 interaction domain of SRC2. (*Bottom*) Western blot showing doxycycline (Dox)-dependent expression of ERα Wt, ΔH12, and ΔH12-PID in MDA-MB-231 cells.
(B) The SRC2 PID is sufficient to recruit p300 to chromatin. ChIP-qPCR assays for ERα and p300 in MDA-MB-231 cell lines expressing the ERαs described in panel (A) ± E2 treatment for 45 min. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the vehicle control in each cell line (*, $p < 0.005$; two-way ANOVA).
(C) Forced recruitment of p300 to an inactive ERα binding site restores Mediator recruitment and H3K27ac enrichment. ChIP-qPCR assays for Med1 and H3K27ac in the MDA-MB-231 cell lines expressing the ERαs described in panel (A) ± E2 treatment for 45 min. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the vehicle control in each cell line (*, $p < 0.05$; **, $p < 0.0001$; two-way ANOVA).
(D) Forced recruitment of p300 to an inactive ERα binding site restores enhancer-promoter chromatin looping. *(Top)* Browser tracks for ERα ChIP-seq and GRO-seq in 231/ERαWt cells after 45 min. of E2 treatment, shown with the location of the 3C primers. *(Bottom)* 3C-qPCR assays showing chromatin looping from a distal ERα binding site to the *OTUB2* promoter to in 231/ERαΔH12 or 231/ERαΔH12-PID cells. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to 231/ERαΔH12 cells at each genomic location in each condition ($p < 0.0001$; two-way ANOVA).
(E) Forced recruitment of p300 to an inactive ERα binding site restores E2-responsive eRNA production and gene expression. RT-qPCR assays in MDA-MB-231 cells expressing the ERαs described in panel (A) over a time course of E2 treatment. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the vehicle control in each cell line (*, $p < 0.05$; **, $p < 0.005$; two-way ANOVA).

[Figure 3.6 is on the next page]

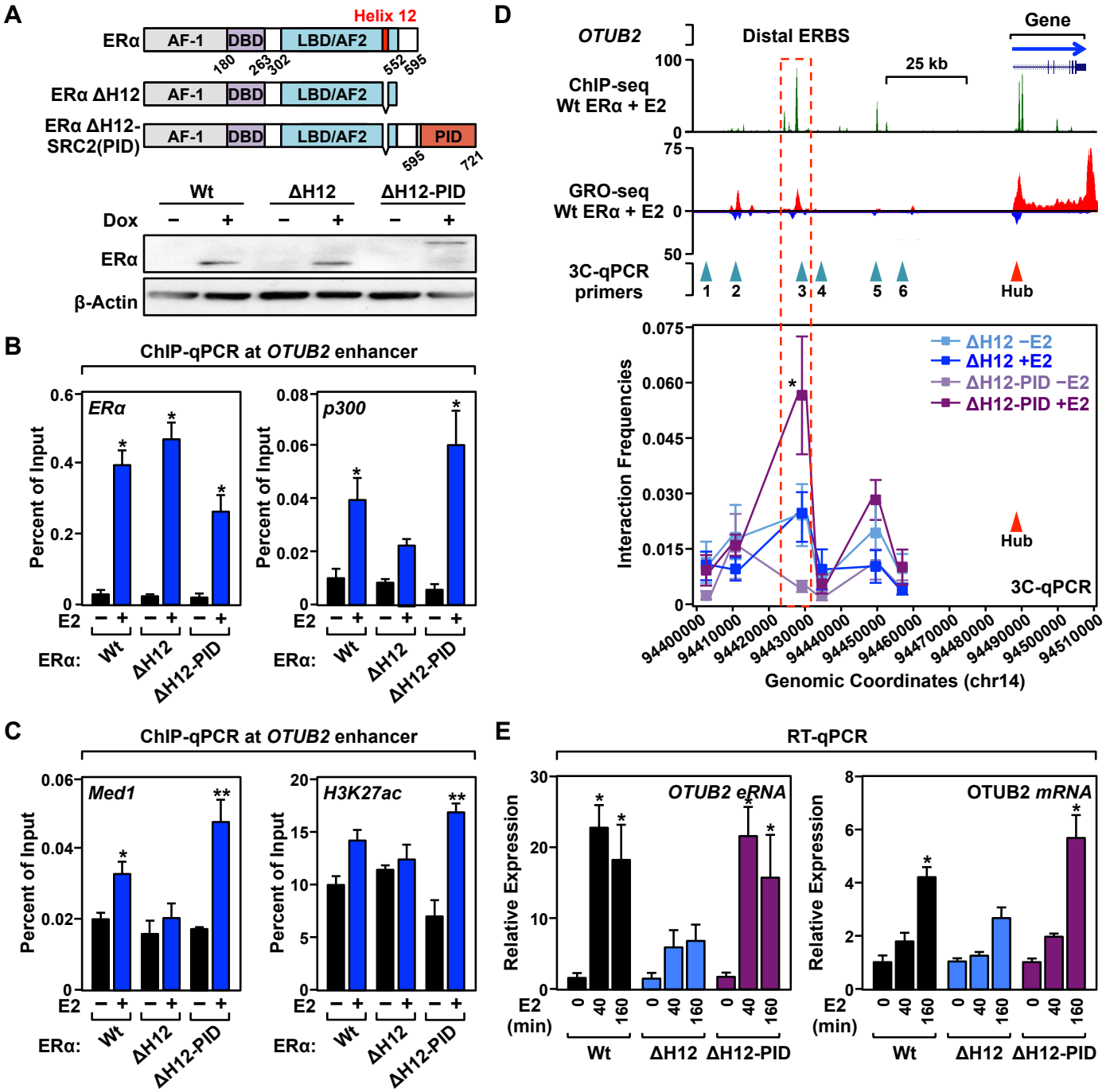
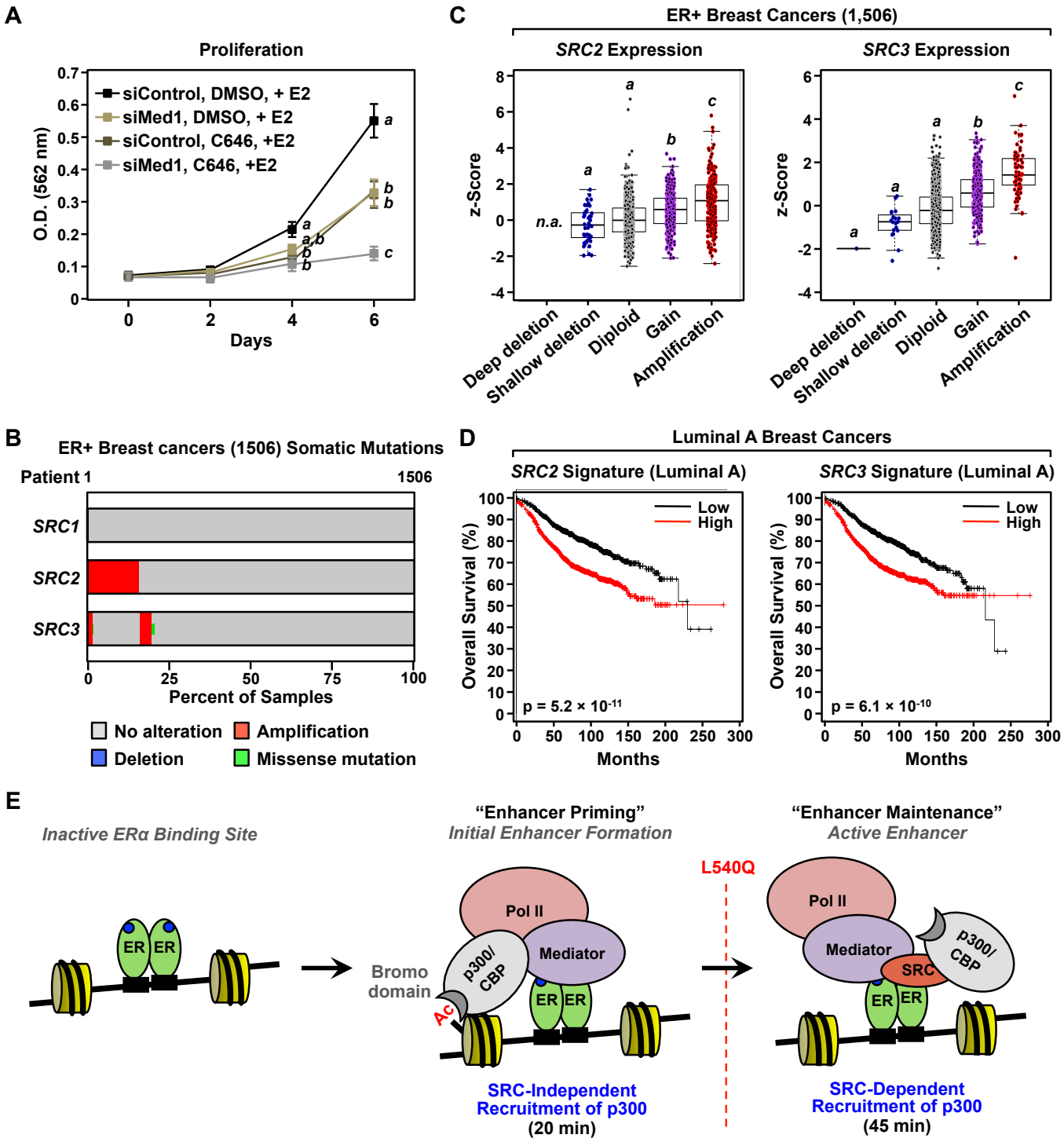

A
Helix 12
ERα
AF-1
DBD
LBD/AF2
180
263
302
552
595
ERα ΔH12
ERα ΔH12-SRC2(PID)
PID
721
Wt
ΔH12
ΔH12-PID
Dox
ERα
β-Actin
B
ChIP-qPCR at OTUB2 enhancer
ERα
p300
Percent of Input
E2
ERα:
C
Med1
H3K27ac
D
OTUB2
Distal ERBS
Gene
25 kb
ChIP-seq Wt ERα + E2
GRO-seq Wt ERα + E2
3C-qPCR primers
Hub
ΔH12 −E2
ΔH12 +E2
ΔH12-PID −E2
ΔH12-PID +E2
Interaction Frequencies
3C-qPCR
Genomic Coordinates (chr14)
E
RT-qPCR
OTUB2 eRNA
OTUB2 mRNA
Relative Expression
E2 (min)

Figure 3.7. p300, Mediator, and SRCs link ERα enhancer function to cell growth and clinical outcomes for ER-positive breast cancers.
(A) Growth curve showing the combinatorial effects of p300 inhibition and Med1 depletion on E2-dependent MCF-7 cell proliferation. Proliferation was measured after siRNA-mediated Med1-depletion ± the p300/CBP HAT inhibitor C646 in the presence of E2. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Points marked with different letters (*a*, *b, c*) are significantly different from each other ($p < 0.05$; two-way ANOVA).
(B) *SRC2* and *SRC3* are frequently amplified in ER-positive breast tumors. Bar graphs of somatic mutations and copy number variations identified for the *SRC2* and *SRC3* genes in ER-positive breast tumors based on data from TCGA.
(C) Amplifications of the *SRC2* and *SRC3* genes lead to overexpression of their cognate mRNAs in ER-positive breast tumors. Box plots showing *SRC2* and *SRC3* mRNA expression in ER-positive breast tumors binned based on copy number variations of *SRC2* and *SRC3* genes, respectively. Box plots marked with different letters (*a*, *b, c*) are significantly different from each other ($p < 0.005$; Wilcoxon rank-sum test).
(D) Kaplan-Meier plots for Luminal A ER-positive breast cancer patients using a set of genes (signature genes) whose expression is up-regulated in samples with *SRC2* or *SRC3* amplification compared to samples without amplification based on data from TCGA. The expression levels (high or low) of the signature genes determined in curated microarray datasets stratify patients into two groups. Their overall survival rates are shown in the plots.
(E) A model for two-step ERα enhancer activation, showing two distinct p300-dependent phases: (1) a "priming" phase with SRC-independent recruitment of p300 and (2) a "maintenance" phase with SRC-dependent recruitment of p300. See the text for details.

[Figure 3.7 is on the next page]


A
Proliferation
O.D. (562 nm)
Days
siControl, DMSO, + E2
siMed1, DMSO, + E2
siControl, C646, +E2
siMed1, C646, +E2
C
ER+ Breast Cancers (1,506)
SRC2 Expression
SRC3 Expression
z-Score
n.a.
Deep deletion
Shallow deletion
Diploid
Gain
Amplification
B
ER+ Breast cancers (1506) Somatic Mutations
Patient 1
1506
SRC1
SRC2
SRC3
Percent of Samples
No alteration
Amplification
Deletion
Missense mutation
D
Luminal A Breast Cancers
SRC2 Signature (Luminal A)
SRC3 Signature (Luminal A)
Overall Survival (%)
Low
High
p = 5.2 × 10-11
p = 6.1 × 10-10
Months
E
Inactive ERα Binding Site
"Enhancer Priming"
Initial Enhancer Formation
"Enhancer Maintenance"
Active Enhancer
ER
Bromo domain
Pol II
Mediator
p300/ CBP
Ac
L540Q
SRC
SRC-Independent Recruitment of p300 (20 min)
SRC-Dependent Recruitment of p300 (45 min)

CHAPTER FOUR

Role of BET Family Members in ERα Enhancer Function and Gene Regulation in Breast Cancer Cells

The study described in this chapter was conducted in collaboration with Rui L, Anusha Nagari, M.S., and Minho Chae, Ph.D. R.L. initiated the project, designed and performed the experiments. A.N. and M.C. executed the genomic analysis. I finalized the study, interpreted the data, and prepared the manuscript.

4.1 Summary

Estrogen (E2)-dependent gene regulation mediated by estrogen receptor alpha (ERα) plays a mitogenic role in ER-positive breast cancer cells. Although clinical applications of selective estrogen receptor modulators (SERMs), which directly interact with ERα to alter ERα activity, have been effective as a first line of treatment for breast cancer patients, a large subset of the patients will develop resistance after prolonged use of SERMs. Thus, there is a great need to develop alternative therapeutic strategies for SERM-resistant breast cancers. Here, we describe the potential use of the bromodomain family member protein (BRD) selective bromodomain inhibitor, JQ1, to alter E2-dependent gene expression program and inhibit E2-dependent growth of breast cancer cells. We show that each family member has partially redundant roles as ERα coregulators that are required for ERα-mediated gene transcription. Furthermore, we demonstrate the function of BRD3 as a molecular sensor of total BRD activity by the compensatory control of its protein levels. In addition, BRD3 colocalizes with a subset of ERα binding sites (ERBSs) that are enriched for active enhancer features and associated with highly E2-induced genes. Collectively, we illustrate a critical role of the BET family members in ERα dependent gene expression.

4.2 Introduction

Estrogen signaling plays a wide array of physiological roles in various reproductive and non-reproductive organs. Steroid hormone estrogen binds to its cognitive receptors, estrogen receptors (ERs) in the nucleus to exert its role through gene regulation (Burns and Korach, 2012; Nilsson et al., 2001; Simpson and Santen, 2015). Liganded ERs dimerize and bind to loci across the genome called transcription enhancers. A large part of ERBSs in the genome contain a specific DNA sequence motif called estrogen response element (ERE), which is often bound by other chromatin-associated proteins including FoxA1 prior to ER binding (Jozwik and Carroll, 2012; Nilsson et al., 2001; Welboren et al., 2009). Alternatively, activated ERs are recruited to the genome indirectly through other transcription factors (TF) such as activating protein-1 (AP-1) family proteins (Kushner et al., 2000). Upon binding to enhancers, ERα recruit additional transcription coregulators and chromatin remodelers to establish fully active transcription enhancers (Nilsson et al., 2001; Shao and Brown, 2004). Active ERα enhancers in turn recruit general transcription factors and RNA polymerase II (Pol II) by acting with a number of transcription coregulators to evoke transcriptional outputs (Nilsson et al., 2001).

Aside from its roles in normal physiology, diverse functions of estrogen signaling are implicated in pathological conditions including breast and uterine cancer, obesity, and osteoporosis (Burns and Korach, 2012; Simpson and Santen, 2015). In particular, estrogen signaling via ERα exhibits a tumorigenic role in breast cancer. Indeed, approximately 60 - 70 % of breast cancers express ERs, and SERMs, including partial agonists such as

tamoxifen and raloxifen as well as antagonist fulvestrant, are widely used as a first line of treatment in ER-positive breast cancer patients. Although the treatments targeting estrogen receptors are effective in many cases, prolonged use of SERMs, such as tamoxifen, ultimately promote the development of resistance in nearly all patients (Alluri et al., 2014b; Chang, 2012; Shao and Brown, 2004). The mechanisms of resistance includes acquisition of E2 signaling-independent cancer cell growth and adapting gain-of-function mutations on ERα (Alluri et al., 2014a; Chang, 2012). To achieve the latter mechanism, Several mutations in ERα found in SERM-resistant breast cancers are shown to cause ligand-independent activation of ERα by adapting an active conformation that allows unliganded mutant ERα to interact with transcriptional coregulators including steroid receptor coactivators (a.k.a. SRC1/2/3 or p160) (Herynk and Fuqua, 2004; Jeselsohn et al., 2015; Thomas and Gustafsson, 2015; Toy et al., 2013). Thus, there is an urging necessity to develop new clinical approaches to control ER-dependent gene expression in SERM-resistant breast cancer patients.

The BET family proteins are transcription coregulators that cooperate with a wide variety of transcription factors for proper transcription activation, including ERα (Nagarajan et al., 2014; Sanchez et al., 2014). The BET family is composed of four family members, BRD2, BRD3, BRD4, and BRDt, each of which contains two bromodomains that bind to acetylated lysine residues on histones (Filippakopoulos et al., 2012; Shi and Vakoc, 2014). BRD2, BRD3, and BRD4 are expressed in a wide range of tissues, while BRDt expression is limited to testis (Belkina and Denis, 2012; Filippakopoulos et al., 2012). Upon activation of transcription factors, histones are hyperacetylated by histone acetyltransferases recruited by

transcription factors, which in turn recruit BET family proteins to execute a transcription response. The BET family members recruited at enhancers are required for recruitment of p-TEFb that harbors CDK9, a kinase that phosphorylates NELF, allowing the release of Pol II from promoter-proximal pausing and transcription progression (Jang et al., 2005; Kanno et al., 2014; Wang and Filippakopoulos, 2015).

JQ1 and iBET are recently developed inhibitors for the BET family proteins that bind to the bromodomains to inhibit their recruitment to acetylated histones (Asangani et al., 2014; Dawson et al., 2011; Filippakopoulos et al., 2010). A growing body of literature supports the efficacy of these inhibitors in various diseases including many types of cancers by modulating the transcriptional networks that are crucial for the context-specific growth of cancer cells (Alluri et al., 2014a; Baratta et al., 2015). A previous study on the role of BRD4 in ERα-dependent gene transcription by Nagaragan *et al.* showed that JQ1 inhibits E2-dependent gene transcription in ER-positive breast cancer cells (Nagarajan et al., 2014). However, precise molecular mechanisms of BET family members in ERα-dependent transcription are not yet clear. Also, it is important to understand the functional relationships among the BET family members, to decipher through which family protein JQ1 modulates E2-dependent gene expression.

In this study, we investigate the role of the BET family proteins in ERα-mediated gene regulation. The inhibition of the BET family members by JQ1 attenuated E2-dependent growth of breast cancer cells. This effect of BET family inhibition was mediated by impaired E2-dependent gene transcription, supporting a critical role of the BET family proteins in ERα-dependent gene activation. Furthermore, we show that BRD3 plays a

prominent role by modulating the total activity of the family members. Collectively, our study demonstrates an important function of the BET family proteins in E2-dependent gene regulation.

4.3 Results

Expression of the BET Family Members Correlates With Breast Tumor Outcomes

The BET family protein members have been suggested as potent therapeutic targets in the treatment of various cancer types, including leukemia and prostate cancers (Asangani et al., 2014; Delmore et al., 2011; Zuber et al., 2011). We examined the potential roles of the BET family members in ER-positive breast cancers. We found that high expression of the family members correlates with poor overall survival rate for ER-positive breast tumor patients (Fig. 4.1A). In particular, BRD3 and BRD4 expression were retained in > 90 % and 100 % of breast tumor samples, respectively, while < 20 % of tumor samples exhibited detectable BRD2 levels (Fig. 4.1B and 4.1C) (Uhlen et al., 2005; Uhlen et al., 2015). We examined whether the BET family members play a role in ER-positive breast cancer cell growth. Consistent with previous studies, a potent BET family member inhibitor, JQ1, attenuated E2-dependent growth of MCF-7 cells (Fig. 4.1D) (Nagarajan et al., 2014).

BET Family Proteins Recruited to ERα Enhancers Are Required for E2-dependent Gene Expression

The BET family members recruited to active enhancers through acetylated histones further recruit CDK9-containing complex p-TEFb, which in turn phosphorylates NELF complex to release paused Pol II for productive elongation (Baratta et al., 2015; Kanno et al., 2014). We hypothesized that the effect of BET family inhibition on MCF-7 cell growth is due to dismissed? E2-dependent transcription regulation. Indeed, the presence of JQ1 impaired E2-induced gene expression in MCF-7 cells (Fig. 4.2A). Importantly, the protein levels of ERα and steroid receptor coactivators (SRCs) were comparable in the presence or absence of JQ1 (Fig. 4.8). Furthermore, JQ1 did not change the levels of ERα recruitment and histone acetylation at the enhancers for E2-induced genes (Fig. 4.2B). However, the E2-dependent recruitment of the BET family proteins was significantly reduced by JQ1 (Fig. 4.2C). Collectively, our data support the mechanism of JQ1 as an inhibitor of BET family member recruitment to ERα enhancers.

It has been previously shown that in some Myc-dependent cancers, JQ1 inhibits tumor growth by repressing *Myc* transcription (Chapuy et al., 2013; Fowler et al., 2014; Loven et al., 2013). Since *Myc* expression is elevated by E2, we tested whether the attenuated E2-dependent growth of MCF-7 by JQ1 is due to reduced E2-dependent induction of Myc. Interestingly, JQ1 did not affect E2-induced Myc expression (Fig. 4.9), suggesting a Myc-independent mechanism of JQ1 action on E2-induced growth of MCF-7 cells.

Inhibition of the BET Family Proteins Results in Dysregulation the E2-dependent Gene Expression Program

Based on the significant growth inhibitory effect of JQ1, we hypothesized that the BET family proteins may have a more extensive genome-wide effect on E2-dependent gene expression. To explore this, we performed RNA-seq assays using MCF-7 cells treated with E2 in the presence or absence of JQ1. We found that 21 % of genes expressed in at least one condition in MCF-7 cells were affected by treatment with JQ1. However, the number of genes affected by JQ1 increased to 38% when examining E2-regulated genes, suggesting an important role of the BET family members in the E2-responsive gene expression program (Fig. 4.3A). The expression of 345 out of 706 E2-upregulated genes and 178 out of 659 E2-downregulated genes were affected greater than 1.5 fold by JQ1 (Fig. 4.3B and 4.3C). Furthermore, gene ontology analysis identified *response to endogenous stimulus* and *hormones* as the top ontology terms among gene groups whose E2-induced gene activation was suppressed by JQ1. This supports the hypothesis that the inhibitory effect of JQ1 in E2-stimulated growth of MCF-7 cells was mediated through regulation of the E2-dependent gene expression program (Table 1).

BRD3 Is a Critical Coregulator for ERα-dependent Gene Expression in MCF-7 cells

The BET family bromodomain inhibitor JQ1 exhibits a range of affinities to each BET family protein. Each family member contains two bromodomains, each with different affinities toward JQ1. The N-terminal bromodomains in BRD3 and BRD4 have the highest affinity to JQ1, followed by the C-terminal bromodomains of the two proteins. Although the

affinity is lower, the bromodomains of BRD2 are also targeted by JQ1 (Filippakopoulos et al., 2010). Elucidating the differential functions and responses to JQ1 among the BET family members is crucial to understanding their roles in E2-dependent gene expression.

To dissect the differential functions of the BET family members, we performed siRNA-mediated knockdown of BRD2, BRD3, and BRD4 in MCF-7 cells and evaluated their contribution to E2-dependent gene activation (Fig. 4.4A). Depletion of a single family member did not affect E2-induced gene expression in MCF-7 cells. However, E2-responsive gene activation was impaired when BRD3 was depleted in combination with BRD2, BRD4, or both (Fig. 4.4B). Interestingly, BRD3 protein levels were elevated when either BRD2, BRD4 or both were depleted, suggesting a compensation mechanism that controls BRD3 expression levels. We did not find elevated expression of BRD2 or BRD4 under any condition tested, highlighting a BRD3-specific role in fine-tuning the overall activity of the BET family members (Fig. 4.4A). To test this hypothesis, we re-expressed siRNA-resistant BRD3 in MCF-7 cells with simultaneous knockdown of the BET family members (Fig. 4.5A). Re-expression of BRD3 restored E2-dependent gene expression at a level comparable to the siRNA control, illustrating that BRD3 alone is sufficient to function as a coregulator for ERα (Fig. 4.5B). Together, these results allude to a partial functional redundancy among the family members, and a key role of BRD3 as a molecular sensor that regulates the total activity of the BET family members.

BRD3 Co-occupies a Subset of ERα Enhancers

To further understand the role of BRD3 at ERα enhancers, we performed ChIP-seq assays for BRD3 and acetylated H4 in MCF-7 cells treated with E2 and analyzed with previously performed ERα ChIP-seq data (Franco et al., 2015). Interestingly, we found that BRD3 occupancy segregates ERBSs into two distinctive groups, ERBSs with or without BRD3 enrichment. The group of ERα BSs enriched for BRD3 recruitment was also significantly enriched for acetylated histone H4 upon E2 treatment (Fig. 4.6A, 4.6B, and 4.6C; Fig. 4.10), which serves as a platform for BRD3 binding (Belkina and Denis, 2012).

Enhancers are often characterized by specific features, including the enrichment of lineage-determining chromatin binding factors (e.g. FoxA1 and AP2γ), high DNAaseI sensitivity, the enrichment of acetylated histones and specific histone marks (e.g. H3K4me1/2 and H3K27ac), the enrichment of transcription coregulators, as well as generation of enhancer transcripts (eRNAs) (Calo and Wysocka, 2013; Ong and Corces, 2011; Shlyueva et al., 2014). To understand the role of BRD3 in ERα enhancer activation, we compared our ChIP-seq datasets with publicly available FoxA1 ChIP-seq and DNaseI hypersensitivity sites (DHSs) datasets of MCF-7 cells treated with E2. Histone acetylation (acH4), FoxA1 and DHSs are all enriched at ERBSs with BRD3 enrichment compared to ERBSs without BRD3 upon E2 treatment. While ERα binding and acetyl H4 levels were induced after E2 treatment at ERBSs without BRD3 enrichment, we did not observe E2-dependent increase of FoxA1 recruitment and DHSs at ERBSs without BRD3 enrichment (Fig. 4.6B). These results indicate that BRD3 recruitment occurs downstream of ERα binading and at least partial histone acetylation. In addition, our data also suggest that,

although acetylated histones might be required for BRD3 recruitment, acetylated histones are not sufficient to recruit BRD3. Lastly, our analysis also implies a possible role of BRD3 in E2-induced FoxA1 recruitment and DHSs at ERα enhancers.

DNA Sequences Contribute to BRD3 Recruitment at ERα Enhancers

From the ChIP-seq datasets, we found that ERα peaks associated with BRD3 are surrounded by multiple ERBSs compared to ERα peaks that do not associate with BRD3 (Fig. 4.6A). To test the relationship in genomic localization between the satellite ERα binding peaks and the central ERα binding peaks with or without BRD3 enrichment, we quantified the number of ERα peaks within 10 kb surrounding the central ERα peaks with or without BRD3 and binned into 500 bp intervals from the reference ERBSs. Interestingly, we found more ERα peaks within 5 kb of the reference ERα binding peaks that are associated with BRD3 compared to the ERα binding peaks without BRD3. In particular, there is a 30 times greater chance of finding another ERα peak within 1 kb if a given ERα binding site is associated with BRD3 compared to ERα peaks without BRD3 (Fig. 4.6D).

We also noticed that the intensity of ERα binding is higher when they surround ERα peaks associated with BRD3 compared to ERα peaks without BRD3. We asked if differences in ERα binding motifs (EREs) under the surrounding ERBSs could account for the differences in the binding intensities of the satellite ERα peaks. Interestingly, both full and half motifs under the satellite ERα peaks are nearly identical to the reference ERα peaks, regardless of BRD3 enrichment (Fig. 4.6E). Instead, the frequencies of full ERE occurrence was significantly higher within 10 kb around the reference ERBSs associated with BRD3 as

compared to the reference ERBSs without BRD3 enrichment (Fig. 4.6F). Taken together, our results imply a close genomic relationship between the clustered ERα enhancers with BRD3 recruitment and the effects of ERE occurrence surrounding the ERBSs that colocalized with BRD3.

ERα Enhancers Occupied by BRD3 Are Active Enhancers

In our previous study, we demonstrated that approximately half of distal ERBSs generate enhancer transcripts. ERBSs associated with eRNAs highly correlate with various active enhancer features, including enrichment of coregulators and H3K4me1 regardless of ERα binding intensities. These results indicate that only a subset of ERBSs serve as active enhancers, and eRNAs are a sensitive surrogate to measure enhancer activity (Fang et al., 2014; Franco et al., 2015).

To examine the role of BRD3 in ERα enhancer activation, we analyzed eRNA transcription levels using a GRO-seq dataset (Hah et al., 2011). Significantly higher levels of eRNA were transcribed upon E2 treatment at ERBSs with BRD3 enrichment compared to ERBSs without BRD3 enrichment (Fig. 4.7A), suggesting a close association of BRD3 with active enhancers. In line with this result, E2-regulated genes nearest to ERα/BRD3 enhancers are enriched for E2-upregulated genes that are more highly induced compared to E2-regulated genes nearest to ERBSs without BRD3 recruitment (Fig. 4.7B, 7C, 7D; Fig. 4.11). These results strongly support the role for BRD3 in ERα-mediated gene transcription.

4.4 Discussion

The BET Family Members Are an Integral Part of Active ERα Enhancers

In this study, we illustrate a critical role of the BET family proteins in E2-dependent gene expression. Inhibition of the family members attenuated E2-dependent growth of ER-positive breast cancer cells (Fig. 4.1D). The attenuated growth by JQ1 was in part due to the genome-wide impairment of E2-dependent gene expression mediated by ERα (Fig. 4.2A and 4.3B). As expected, JQ1 had no effect on ERα binding or histone acetylation levels at ERα enhancers, but rather prevented the recruitment of the BET family members at ERα enhancers (Fig. 4.2B and 4.2C). Our results corroborate the known role of the BET family members in signal-induced gene transcription, including a previously reported role of BRD4 in E2-dependent gene transcription (Alluri et al., 2014a; Asangani et al., 2014; Nagarajan et al., 2014; Nicodeme et al., 2010). Furthermore, expression of the BET family members highly correlates with clinical outcomes in ER-positive breast cancer patients, suggesting the BET family members as potential therapeutic targets (Fig. 4.1A).

BRD3 Orchestrates the Overall Activity of the BET Family Members

BRD2, BRD3, and BRD4 are all expressed at detectable levels in ER-positive MCF-7 breast cancer cells (Fig. 4.4A). Despite the low-selectivity of JQ1 among the family members (Filippakopoulos et al., 2010), a number of studies on the roles of the BET family proteins have blindly focused on BRD4 (Brown et al., 2014; Gopalakrishnan et al., 2016; Hong et al., 2016; Nagarajan et al., 2014) as it was the member protein used to design JQ1

(Filippakopoulos et al., 2010). Indeed, thorough studies dissecting the differential functions among the family members are limited (Belkina et al., 2013; Deeney et al., 2016; Delmore et al., 2011; Stonestrom et al., 2015).

In our current study, the depletion of any of the family members alone was not sufficient to impair E2-dependent gene activation. Interestingly, the co-depletion of BRD3 with either BRD2 or BRD4 or both was required to significantly alter E2-responsive gene expression (Fig. 4.4B). In addition, BRD3 protein levels were elevated when BRD2 and/or BRD4 were depleted, presumably to compensate the total activity of the BET family proteins. Importantly, the compensation mechanism was achieved only by modulating BRD3 protein levels, but not BRD2 or BRD4 (Fig. 4.4A). These results demonstrate the partial functional redundancy among the family members, at least in the ERα enhancers tested in this study, and emphasize the role of BRD3 as a molecular sensor that fine-tunes the overall activity of the family members.

Our findings on the differential and redundant roles of the BET family members are likely to be context-specific. For instance, a previous study demonstrated the critical role of BRD2, but not BRD3 or BRD4, for insulin secretion by pancreatic β-cells (Deeney et al., 2016). On the other hand, additional studies have shown that each family member is required for proper inflammatory response in macrophages, enhanced cell survival in multiple myeloma models, and AR-dependent growth of prostate cancers (Asangani et al., 2014; Belkina et al., 2013l; Delmore et al., 2011). Thus, further investigation is needed to understand the extent of functional redundancy among the BET family members in E2-dependent gene expression and the molecular mechanisms that control BRD3 levels in

response to changes in BRD2 and BRD4 activity in order to maximize the potential of the BET family members as therapeutic targets in breast cancer treatment.

BRD3 Marks Active ERα Enhancers

Given the central role of BRD3 in E2-responsive gene expression in MCF-7 cells, we examined the genome-wide localization of BRD3 in response to E2 treatment. Interestingly, BRD3 was enriched only at a subset of ERBSs upon E2 stimuation (Fig. 4.6A). This subset of ERBSs highly correlate with elevated levels of E2-dependent transcription at the enhancers as well as the target genes (Fig. 4.7A), supporting an integral role of BRD3 in E2-responsive gene expression. These results corroborate previous studies by our group and others which demonstrate that not all stable TF binding events result in transcription activation. Furthermore, these studies also reveal that functional TF binding sites strongly correlate with enrichment of other active enhancer marks, including coregulator recruitment and enhancer-enriched histone modification, and activation of the target genes (Hah et al., 2013; Kouwenhoven et al., 2015; Savic et al., 2015; Shlyueva et al., 2014).

Based on our ChIP-seq experiments, we find that BRD3-enriched ERBSs are surrounded by additional ERα binding at higher frequencies and intensities. Recent studies by Young group and others report clusters of cell-type specific transcription factor binding, namely super enhancers (SEs), which control highly cell-type specific genes and thus maintain cellular identities (Hnisz et al., 2013; Whyte et al., 2013). Subsequent studies also demonstrate that the genes controlled by SEs are disproportionally sensitive to perturbation of transcription coregulators, including the BET family members and the TFIIH component

CDK7 (Chapuy et al., 2013; Chipumuro et al., 2014; Loven et al., 2013). In agreement with these observations, the clusters of ERα enhancers enriched with BRD3 identified in our present study may be categorized as SEs that collectively activate key ERα regulated genes (Fig. 4.6 and 4.7). Together, our results indicate that BRD3 marks active ERα enhancers distinguished from the rest of ERBSs, implicating BRD3 in ERα enhancer activation (Hah et al., 2013).

Implication for Additional Mechanisms of BET Family Member Recruitment

Our current study shows the requirement of the bromodomains in BRD2, BRD3, and BRD4 for the recruitment at ERα enhancers (Fig. 4.2) (Filippakopoulos et al., 2012; Kanno et al., 2004; Zeng and Zhou, 2002). However, it is evident that ERα enhancers are acetylated at detectable levels compared to the surrounding regions even prior to E2 exposure when BRD3 recruitment is at a basal level. In addition, BRD3 is localized at ERBSs, while histone acetylation is more broadly distributed around ERα enhancers (Fig. 4.6A). Interestingly, ERBSs, regardless of BRD3 recruitment, exhibit an E2-dependent increase in histone acetylation. Only a subset of ERα binding sites, however, recruit BRD3 (Fig. 4.6B). Together, these observations suggest that 1) BRD3 recruitment happens in intermediate steps of ERα enhancer formation downstream of histone acetylation, 2) histone acetylation is not sufficient for BRD3 recruitment, and thus 3) there might be another layer of regulation for BET family member recruitment at partially formed ERα enhancer complexes. A deeper understanding is needed of the molecular mechanisms of BET family member recruitment

and function at ERα enhancers to fully assess the potential of the BET family members as therapeutic targets for breast cancer.

In summary, we demonstrate a critical role of BET family members in E2-responsive gene expression and the potential use of their inhibitors in treating ER-positive breast cancer. Among the three widely expressed BET family members, BRD3 is particularly interesting, given the presence of a compensatory mechanism that modulates its protein levels in order to orchestrate total activity of the BET family. We show that BRD3 is enriched for a subset of ERα associated with active enhancer features, such as elevated levels of E2-dependent coregulator recruitment and transcription. Collectively, our current study provides a better understanding of the function of the BET family members in ERα enhancers.

4.5 Materials and Methods

Kaplan-Meier Analyses

Kaplan-Meier estimators (Kaplan and Meier, 1958) were generated using the Gene Expression-Based Outcome for Breast Cancer Online (GOBO) tool (http://co.bmc.lu.se/gobo/) (Ringner et al., 2011). BRD2, BRD3 and BRD4 were provided as the input gene set to assess patient outcomes in ER-positive breast cancers.

Antibodies

The antibodies used for Western blotting and/or ChIP assays were as follows: ERα

(rabbit polyclonal generated in the Kraus Lab), BRD2 (Bethyl Lab, A302-583A), BRD3 (Bethyl Lab, A302-368A), BRD4 (Bethyl Lab, A301-985A), pan-acetyl H4 (Millipore, 06-866), Myc (Invitrogen, 13-2500), SRC2 (Santa Cruz Biotechnology, sc-8996), SRC3 (Santa Cruz Biotechnology, sc-7216), SNRP70 (Abcam, ab83306), β-tubulin (Abcam, ab6046).

Cell Culture and Treatments

MCF-7 breast cancer cells were kindly provided by Dr. Benita Katzenellenbogen (University of Illinois, Urbana-Champaign). They were maintained in Minimum Essential Medium (MEM) Eagle with Hank's salts (Sigma, M1018) supplemented with 5% HyClone calf serum (GE Healthcare, SH30072) and 20mM HEPES (ThermoFisher Scientific, BP310). Prior to gene expression and ChIP experiments, the cells were grown for three days in phenol red-free MEM Eagle medium supplemented with 5% charcoal-dextran-treated calf serum. Treatment conditions for cells were as follows: 17β-estradiol (E2), 100 nM (Sigma, E8875); (+)JQ1 (the active enantiomer of JQ1, referred to herein as JQ1), 500 nM unless otherwise stated (Cayman Chemical, 11187); and (-)JQ1 (the inactive enantiomer of JQ1), at the same concentrations as (+)JQ1 (Cayman Chemical, 11232). The cells were treated with (+)JQ1 or (-)JQ1 for 3 hours before treatment with E2. For gene expression analyses, the cells were collected after 3 hours of E2 treatment. For ChIP analyses, the cells were collected after 45 minutes of E2 treatment.

293T cells were purchased from the ATCC (CRL-3216) and maintained in high glucose Dulbecco's Modified Eagle's Medium (Sigma, D7777) supplemented with 10% fetal bovine serum.

Cell Proliferation Assays

MCF-7 cells were plated at a density of 4 x 10^4 cells per well in six well plates in standard grown medium and then switched to phenol red-free MEM Eagle medium supplemented with 5% charcoal-dextran-treated calf serum after attachment. The following day, the cells were treated with ethanol vehicle (Veh), E2 (100 nM), (+)JQ1 (62.5 or 250 nM), or both E2 + (+)JQ1, with fresh treatments added every 2 days. At selected time points over a six-day time course, the cells were fixed with 10% formaldehyde and stained with 0.1% crystal violet and 200 mM phosphoric acid. After washing away unincorporated stain, the crystal violet was extracted using 10% glacial acetic acid and the absorbance was read at 595 nm. The cell proliferation assays were run a minimum of three times with independent biological samples to ensure reproducibility.

siRNA-mediated Knockdown of BRD2, BRD3, and BRD4

MCF-7 cells were grown to 75% confluence in six well plates. The cells were then transfected with commercially available siRNA oligos directed against BRD2, BRD3 or BRD4 (Sigma) at a final concentration of 5 nM using Lipofectamine RNAiMAX reagent (Invitrogen, 13778150) per the manufacturer's instructions. MISSION universal negative control #2 (Sigma, SIC002) was used as a control siRNA. Treatments with E2 were performed 48 hours after siRNA transfection. The siRNA sequences are as follows:

- siBRD2 5'-GTTACAAGATGTCAGCGGA-3'
- siBRD3 5'-CCAAGGAAATGTCTCGGAT-3'
- siBRD4 5'-CTGGAATGCTCAGGAATGT-3'

Inducible Expression of BRD3

The lentiviral system for inducible expression of BRD3 is based on pINDUCER20 (Meerbrey et al., 2011), which was kindly provided by Dr. Thomas Westbrook (Baylor College of Medicine, Houston, TX). The "gateway cloning region" of the original pINDUCER20 vector was replaced with a multi-cloning site. The human BRD3 cDNA was cloned by reverse transcription-PCR from MCF-7 cell total RNA and then transferred into the modified pINDUCER20 vector with the addition of a sequence encoding an HA tag at the 3' end of the cDNA (for tagging of BRD3 at the carboxyl terminal end). The BRD3 cDNA was confirmed by sequencing.

The pINDUCER20-BRD3-HA plasmid was co-transfected into 293T cells with lentiviral packaging plasmid, envelope plasmid, and pAdVAntage (Promega) using GeneJuice transfection reagent (Novagen, 70967) according to the manufacturer's protocol. The supernatant containing the lentiviruses was collected 48 hours after transfection and used to infect MCF-7 cells. The infected MCF-7 cells were selected and maintained in 1 mg/ml Geneticin (Life Technologies, 11811031). For induction of BRD3 expression, doxycycline hyclate (Sigma, D9891) was added to the medium at a final concentration 50 ng/mL. Twenty-four hours later, the cells were collected for Western blotting or RT-qPCR.

Immunohistochemistry (IHC)

Immunohistochemical staining of patient samples were adapted from the Cancer Atlas of the Human Protein Atlas database, version 15 (www.proteinatlas.org). The

antibodies used for IHC were as follows: BRD2, HPA042816; BRD3, HPA051830; BRD4, CAB068177. The direct links to the original datasets are as follows:

BRD2: http://www.proteinatlas.org/ENSG00000204256-BRD2/cancer/tissue/breast+cancer

BRD3: http://www.proteinatlas.org/ENSG00000169925-BRD3/cancer/tissue/breast+cancer

BRD4: http://www.proteinatlas.org/ENSG00000141867-BRD4/cancer/tissue/breast+cancer

Western Blotting

Protein lysate from MCF-7 cells was prepared using lysis buffer [20 mM HEPES (pH 7.5), 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2mM EDTA, 25% Glycerol, 0.5% NP-40, 1 mM DTT, 1x complete protease inhibitor cocktail (Roche, 11697498001)], and SDS loading samples were prepared using the normalize amount of total protein. Protein expression was examined by western blotting with the antibodies stated above. The signals were developed using a chemilumenescent detection system. (ThermoFisher Scientific, 34080/34095).

mRNA Expression Analysis by Quantitative Real-time PCR (RT-qPCR)

Changes in the steady-state levels of target gene mRNAs were analyzed by RT-qPCR, as described previously (Luo et al., 2014) with a few modifications. MCF-7 cells were grown in six well plates and treated as described above (± 3 hours with 100 nM E2; ± 3 hour pretreatment with 500 nM JQ1). The cells were collected and total RNA was isolated using TRIzol reagent (Invitrogen, 15596) according to the manufacturer's protocol. Two micrograms of total RNA were reverse-transcribed using oligo(dT) or random hexmers using MMLV reverse transcriptase (Promega, M1701) according to the manufacturer's protocol.

The resulting cDNA was analyzed by qPCR using the primer sets listed below using a LightCycler 480 real-time PCR thermocycler (Roche) for 45 cycles. The expression levels were normalized to 18S ribosomal RNA as an internal standard. All experiments were conducted a minimum of three times with independent RNA isolations to ensure reproducibility.

- *TFF1* mRNA Forward 5'-TTGTGGTTTTCCTGGTGTCA-3'
- *TFF1* mRNA Reverse 5'-GCAGATCCCTGCAGAAGTGT-3'
- *GREB1* mRNA Forward 5'-AAACATCAGCTGCTCGGACT-3'
- *GREB1* mRNA Reverse 5'-CCTGACAGATGACACACAACG-3'
- *MYC* mRNA Forward 5'-TCGGATTCTCTGCTCTCCTC-3'
- *MYC* mRNA Reverse 5'-CCTGCCTCTTTTCCACAGAA-3'
- 18S ribosomal RNA Forward 5'-TACCACATCCAAGGAAGGCAGCA-3'
- 18S ribosomal RNA Reverse 5'-TGGAATTACCGCGGCTGCTGGCA-3'

Preparation of polyA+ RNA-seq Libraries

RNA-seq libraries were prepared as described previously (Zhong et al., 2011). Briefly, MCF-7 cells were grown in 10 cm diameter dish to the density of 8.8 x 10^6 cells and treated as described above (± 3 hours with 100 nM E2; ± 3 hour pretreatment with 500 nM JQ1). The cells were collected and total RNA was isolated using the RNeasy Plus kit (Qiagen, 74134). PolyA+ RNA was purified from the total RNA using Dynabeads Oligo(dT)25 (Life Technologies, 61002). Strand-specific libraries were prepared according to the "deoxyuridine triphosphate (dUTP)" method as described previously (Zhong et al.,

2011). After quality control analyses, the libraries were sequenced on an Illumina HiSeq 2000 (single-end sequencing, 50 nt).

Analysis of RNA-seq Data

Quality control of RNA-seq reads were performed using FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and reads were aligned using TopHat v2.0.10 (Langmead et al., 2009) on the hg19 reference genome. Using aligned reads as input, we used Cufflinks v.2.1.1 (Trapnell et al., 2010) and Cuffdiff v.2.1.1 (Trapnell et al., 2013) to assemble the reads into transcripts using RefSeq annotations and to call differentially regulated transcripts, respectively. All programs were run with default parameters. Expression of differentially expressed genes was visualized as heatmaps using Java Tree View (Saldanha, 2004) and as boxplots using the boxplot function in R. The read counts +1 to +13 kb from the 5' end of each gene were visualized in boxplots as Reads Per Kilobase of gene per Million mapped reads (RPKM). The statistical analysis on the effect of JQ1 upon E2 treatment for all E2-regulated genes was performed using 100 genes randomly chosen from each category.

Analysis of GRO-seq Data

GRO-seq data was analyzed as previously mentioned (Hah et al., 2011). The GRO-seq reads surrounding ± 2.5 kb of center of the ERα peaks or surrounding the 5' end of regulated genes nearest to the ERα peaks were visualized in boxplots as Reads Per Kilobase of gene per Million mapped reads (RPKM) using boxplot function in R.

Gene Ontology (GO) Analyses

Gene ontology analyses were performed using DAVID (Database for Annotation, Visualization, and Integrated Discovery) (Dennis et al., 2003). As input, the list of genes expressed in MCF-7 at least in one condition tested was used as a background. DAVID returns clusters of related ontological terms that are ranked according to an enrichment score. We listed the top term in each cluster (based on p-value) from the top ten clusters (based on enrichment score). The statistical analysis on E2-dependent gene transcription regulated by ERα enhancer with or without BRD3 enrichment was performed using 100 genes or ERα binding sties for eRNAs randomly chosen from each category.

Chromatin Immunoprecipitation (ChIP)

ChIP was performed as previously described (Franco et al., 2015b) with a few modifications. MCF-7 cells were plated at a density of 2.5 x 10^6 cells in a 15 cm diameter dish, grown for three days in estrogen-free medium, and treated as described above (± 45 minutes with 100 nM E2; ± 3 hour pretreatment with 500 nM JQ1). The cells were cross-linked with 1% formaldehyde for 10 min at 37^{o}C and quenched in 125 mM glycine for 5 min at 4^{o}C. The cells were then collected and lysed in Farnham lysis buffer [5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40, 1 mM DTT, and 1x complete protease inhibitor cocktail (Roche, 11697498001)]. For acetyl-histone H4 ChIP assays, 10 mM sodium butyrate was added to all buffers to prevent deacetylation of the histones. The crude nuclear pellet was collected by centrifugation, resuspended in lysis buffer [Tris-HCl pH 7.9, 1% SDS, 10 mM EDTA, 50 mM, 1 mM DTT, and 1x complete protease inhibitor cocktail], and incubated on ice for 10

minutes. The chromatin was sheared by sonication at 4°C using a Bioruptor UC200 at the highest setting for five 5-minute cycles of 30 seconds on and 60 seconds off to generate chromatin fragments of ~200-400 bp in length. The soluble chromatin was diluted 1:10 with dilution buffer [20 mM Tris-HCl pH 7.9, 0.5% Triton X-100, 2 mM EDTA, 150 mM NaCl, 1 mM DTT and 1x complete protease inhibitor cocktail] and pre-cleared with protein A agarose beads. Five percent of the material was removed and saved as input, and the rest of the pre-cleared supernatant was incubated overnight at 4^{o}C with the specified antibodies of interest or without antibody as a control (each 15 cm diameter dish yielded two immunoprecipitations) with continuous mixing.

After the incubation, the immune complexes were collected by adding protein A agarose beads and incubating for 2 hours at 4°C. The immunoprecipitated material was washed once with low salt wash buffer [20 mM Tris-HCl pH 7.9, 2 mM EDTA, 125 mM NaCl, 0.05% SDS, 1% Triton X-100, and 1x complete protease inhibitor cocktail], once with high-salt wash buffer [20 mM Tris-HCl pH 7.9, 2 mM EDTA, 500 mM NaCl, 0.05% SDS, 1% Triton X-100, and 1x complete protease inhibitor cocktail], once with LiCl wash buffer [10 mM Tris-HCl pH 7.9, 1 mM EDTA, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, and 1x complete protease inhibitor cocktail], and twice with Tris-EDTA (TE) containing 1x complete protease inhibitor cocktail. The immunoprecipitated material was eluted at room temperature in elution buffer [100 mM NaHCO3, 1% SDS], and reverse crosslinked by adding 100 mM NaCl and incubating at 65^{o}C overnight. The eluted material was then digested with proteinase K and RNase A to remove protein and RNA, respectively, and the enriched genomic DNA was extracted with phenol:chloroform:isoamyl alcohol, followed by

isopropanol precipitation. The ChIPed DNA was dissolved in water and analyzed by qPCR using the primer sets listed below using a LightCycler 480 real-time PCR thermocycler (Roche). All experiments were conducted a minimum of three times with independent RNA isolations to ensure reproducibility.

- *TFF1* enhancer Forward 5'-TGGTTGCAGATCTTGTTGGA-3'
- *TFF1* enhancer Reverse 5'-TTCTCACACACATCCCCTCA-3'
- *GREB1* enhancer Forward 5'-GAGCTGACCTTGTGGTAGGC-3'
- *GREB1* enhancer Reverse 5'-CAGGGGCTGACAACTGAAAT-3'

Preparation of ChIP-seq Libraries

ChIP-seq libraries were generated using two biological replicates the ChIPed DNA described above for each condition. The DNA was purified using a MiniElute PCR Purification Kit (Qiagen, 28004). After purification, 50 ng of ChIPed DNA for each condition was used to generate libraries for deep sequencing, as previously described (Franco et al., 2015b; Quail et al., 2008), with some modifications. Briefly, the DNA was end-repaired and a single "A"-base overhang was added using the Klenow fragment of *E. coli* DNA polymerase. The A-modified DNA was ligated to Illumina sequencing adaptors using the Illumina TruSeq DNA Sample Prep Kit. The ligated DNA (300-500 bp) was size-selected by agarose gel electrophoresis and purified using a QIAquick Gel Extraction kit (Qiagen, 28704). The size-selected fragments were PCR amplified using Illumina TruSeq P5 and P7 PCR primers, and purified using AMPure beads (Beckman Coulter, A63881). After

quality control analyses, the libraries were sequenced on an Illumina HiSeq 2000 (single-end sequencing, 50 nt).

Analysis of ChIP-seq data

Quality Control and Alignment. The quality of ChIP-seq reads was analyzed by FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). ChIP-seq reads were aligned to the hg19 reference genome using Bowtie 2 v2.2.2 using the default parameters (Langmead et al., 2009), and visualized on UCSC genome browser using BbigWig files generated by BEDTools (Quinlan and Hall, 2010) and custom R scripts.

Peak Calling and Data Representation. ERα peak dataset in E2-treated MCF-7 was employed from a previous study by Franco *et al.* in which peaks were called using the input as a control (Franco et al., 2015b). ChIP-seq read densities surrounding ± 5 kb of ERα peaks for ERα, BRD3, FOXA1, and DNase1, and ± 10 kb of ERα peaks for acetyl H4 were calculated using annotatePeaks.pl function in HOMER software (Heinz et al., 2010) and visualized as heatmaps using Java Tree View (Saldanha, 2004). The read counts within ± 500 bp of top 50% of ERα peaks with BRD3 enrichment and within ± 500 bp of top 10% of ERα without BRD3 enrichment was plotted as boxplots using the boxplot function in R. The statistical analysis on the recruitment of ERα, BRD3, and FoxA1 and the levels of acH4 and DNaseI hypersensitivity at ERBSs with or without BRD3 enrichment was performed using 100 ERBSs randomly chosen from each category.

Motif Analysis. Directed motif search was performed on ± 5 kb window around the center of all the ERα peaks with and without BRD3 enrichment using the command-line

version of FIMO software (Grant et al., 2011). A p-value of 1e-4 was used to identify the genomic locations with significant half ERE or full ERE in the ERα peaks as mentioned above. As a control, directed motif search was also performed on a set of random sequences of 10 kb length generated using BEDTools v2.16.2. The position weight matrix (PWM) for ESR1 was obtained from JASPAR database (Mathelier et al., 2014). The motif logos were generated using WEBLOGO online tool (Crooks et al., 2004).

4.6 References

Alluri, P.G., Asangani, I.A., and Chinnaiyan, A.M. (2014a). BETs abet Tam-R in ER-positive breast cancer. Cell Res *24*, 899-900.

Alluri, P.G., Speers, C., and Chinnaiyan, A.M. (2014b). Estrogen receptor mutations and their role in breast cancer progression. Breast Cancer Res *16*, 494.

Asangani, I.A., Dommeti, V.L., Wang, X., Malik, R., Cieslik, M., Yang, R., Escara-Wilke, J., Wilder-Romans, K., Dhanireddy, S., Engelke, C.*, et al.* (2014). Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. Nature *510*, 278-282.

Baratta, M.G., Schinzel, A.C., Zwang, Y., Bandopadhayay, P., Bowman-Colin, C., Kutt, J., Curtis, J., Piao, H., Wong, L.C., Kung, A.L.*, et al.* (2015). An in-tumor genetic screen reveals that the BET bromodomain protein, BRD4, is a potential therapeutic target in ovarian carcinoma. Proceedings of the National Academy of Sciences of the United States of America *112*, 232-237.

Belkina, A.C., and Denis, G.V. (2012). BET domain co-regulators in obesity, inflammation and cancer. Nat Rev Cancer *12*, 465-477.

Belkina, A.C., Nikolajczyk, B.S., and Denis, G.V. (2013). BET protein function is required for inflammation: Brd2 genetic disruption and BET inhibitor JQ1 impair mouse macrophage inflammatory responses. J Immunol *190*, 3670-3678.

Brown, J.D., Lin, C.Y., Duan, Q., Griffin, G., Federation, A.J., Paranal, R.M., Bair, S., Newton, G., Lichtman, A.H., Kung, A.L.*, et al.* (2014). NF-kappaB directs dynamic super enhancer formation in inflammation and atherogenesis. Mol Cell *56*, 219-231.

Burns, K.A., and Korach, K.S. (2012). Estrogen receptors and human disease: an update. Arch Toxicol *86*, 1491-1504.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Chang, M. (2012). Tamoxifen resistance in breast cancer. Biomol Ther (Seoul) *20*, 256-267.

Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G.*, et al.* (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell *24*, 777-790.

Chipumuro, E., Marco, E., Christensen, C.L., Kwiatkowski, N., Zhang, T., Hatheway, C.M., Abraham, B.J., Sharma, B., Yeung, C., Altabef, A.*, et al.* (2014). CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. Cell *159*, 1126-1139.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Dawson, M.A., Prinjha, R.K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W.I., Robson, S.C., Chung, C.W., Hopf, C., Savitski, M.M*., et al.* (2011). Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature *478*, 529-533.

Deeney, J.T., Belkina, A.C., Shirihai, O.S., Corkey, B.E., and Denis, G.V. (2016). BET Bromodomain Proteins Brd2, Brd3 and Brd4 Selectively Regulate Metabolic Pathways in the Pancreatic beta-Cell. PloS one *11*, e0151329.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., Qi, J*., et al.* (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell *146*, 904-917.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology *4*, P3.

Fang, B., Everett, L.J., Jager, J., Briggs, E., Armour, S.M., Feng, D., Roy, A., Gerhart-Hines, Z., Sun, Z., and Lazar, M.A. (2014). Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. Cell *159*, 1140-1152.

Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Muller, S., Pawson, T*., et al.* (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. Cell *149*, 214-231.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I*., et al.* (2010). Selective inhibition of BET bromodomains. Nature *468*, 1067-1073.

Fowler, T., Ghatak, P., Price, D.H., Conaway, R., Conaway, J., Chiang, C.M., Bradner, J.E., Shilatifard, A., and Roy, A.L. (2014). Regulation of MYC expression and differential JQ1 sensitivity in cancer cells. PloS one *9*, e87003.

Franco, H.L., Nagari, A., and Kraus, W.L. (2015). TNFα signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome

. Molecular Cell *(in press)*.

Gopalakrishnan, R., Matta, H., Tolani, B., Triche, T., Jr., and Chaudhary, P.M. (2016). Immunomodulatory drugs target IKZF1-IRF4-MYC axis in primary effusion lymphoma in a cereblon-dependent manner and display synergistic cytotoxicity with BRD4 inhibitors. Oncogene *35*, 1797-1810.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res *23*, 1210-1223.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Herynk, M.H., and Fuqua, S.A. (2004). Estrogen receptor mutations in human disease. Endocr Rev *25*, 869-898.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947.

Hong, S.H., Eun, J.W., Choi, S.K., Shen, Q., Choi, W.S., Han, J.W., Nam, S.W., and You, J.S. (2016). Epigenetic reader BRD4 inhibition as a therapeutic strategy to suppress E2F2-cell cycle regulation circuit in liver cancer. Oncotarget *7*, 32628-32640.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol Cell *19*, 523-534.

Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M., and Schiff, R. (2015). ESR1 mutations-a mechanism for acquired endocrine resistance in breast cancer. Nat Rev Clin Oncol *12*, 573-583.

Jozwik, K.M., and Carroll, J.S. (2012). Pioneer factors in hormone-dependent cancers. Nat Rev Cancer *12*, 381-385.

Kanno, T., Kanno, Y., LeRoy, G., Campos, E., Sun, H.W., Brooks, S.R., Vahedi, G., Heightman, T.D., Garcia, B.A., Reinberg, D.*, et al.* (2014). BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. Nat Struct Mol Biol *21*, 1047-1057.

Kanno, T., Kanno, Y., Siegel, R.M., Jang, M.K., Lenardo, M.J., and Ozato, K. (2004). Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. Mol Cell *13*, 33-43.

Kaplan, E.L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. J Amer Statist Assn *53*, 457-481.

Kouwenhoven, E.N., Oti, M., Niehues, H., van Heeringen, S.J., Schalkwijk, J., Stunnenberg, H.G., van Bokhoven, H., and Zhou, H. (2015). Transcription factor p63 bookmarks and regulates dynamic enhancers during epidermal differentiation. EMBO Rep *16*, 863-878.

Kushner, P.J., Agard, D.A., Greene, G.L., Scanlan, T.S., Shiau, A.K., Uht, R.M., and Webb, P. (2000). Estrogen receptor pathways to AP-1. J Steroid Biochem Mol Biol *74*, 311-317.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320-334.

Luo, X., Chae, M., Krishnakumar, R., Danko, C.G., and Kraus, W.L. (2014). Dynamic reorganization of the AC16 cardiomyocyte transcriptome in response to TNFalpha signaling revealed by integrated genomic analyses. BMC genomics *15*, 155.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H.*, et al.* (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res *42*, D142-147.

Meerbrey, K.L., Hu, G., Kessler, J.D., Roarty, K., Li, M.Z., Fang, J.E., Herschkowitz, J.I., Burrows, A.E., Ciccia, A., Sun, T.*, et al.* (2011). The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. Proceedings of the National Academy of Sciences of the United States of America *108*, 3665-3670.

Nagarajan, S., Hossan, T., Alawi, M., Najafova, Z., Indenbirken, D., Bedi, U., Taipaleenmaki, H., Ben-Batalla, I., Scheller, M., Loges, S.*, et al.* (2014). Bromodomain protein BRD4 is required for estrogen receptor-dependent enhancer activation and gene transcription. Cell Rep *8*, 460-469.

Nicodeme, E., Jeffrey, K.L., Schaefer, U., Beinke, S., Dewell, S., Chung, C.W., Chandwani, R., Marazzi, I., Wilson, P., Coste, H.*, et al.* (2010). Suppression of inflammation by a synthetic histone mimic. Nature *468*, 1119-1123.

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

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283-293.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., and Turner, D.J. (2008). A large genome center's improvements to the Illumina sequencing system. Nature methods *5*, 1005-1010.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ringner, M., Fredlund, E., Hakkinen, J., Borg, A., and Staaf, J. (2011). GOBO: gene expression-based outcome for breast cancer online. PloS one *6*, e17911.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Sanchez, R., Meslamani, J., and Zhou, M.M. (2014). The bromodomain: from epigenome reader to druggable target. Biochim Biophys Acta *1839*, 676-685.

Savic, D., Roberts, B.S., Carleton, J.B., Partridge, E.C., White, M.A., Cohen, B.A., Cooper, G.M., Gertz, J., and Myers, R.M. (2015). Promoter-distal RNA polymerase II binding discriminates active from inactive CCAAT/ enhancer-binding protein beta binding sites. Genome Res *25*, 1791-1800.

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

Shi, J., and Vakoc, C.R. (2014). The mechanisms behind the therapeutic activity of BET bromodomain inhibition. Mol Cell *54*, 728-736.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Simpson, E., and Santen, R.J. (2015). Celebrating 75 years of oestradiol. J Mol Endocrinol *55*, T1-20.

Stonestrom, A.J., Hsu, S.C., Jahn, K.S., Huang, P., Keller, C.A., Giardine, B.M., Kadauke, S., Campbell, A.E., Evans, P., Hardison, R.C.*, et al.* (2015). Functions of BET proteins in erythroid gene expression. Blood *125*, 2825-2834.

Thomas, C., and Gustafsson, J.A. (2015). Estrogen receptor mutations and functional consequences for breast cancer. Trends Endocrinol Metab *26*, 467-476.

Toy, W., Shen, Y., Won, H., Green, B., Sakr, R.A., Will, M., Li, Z., Gala, K., Fanning, S., King, T.A.*, et al.* (2013). ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. Nat Genet *45*, 1439-1445.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol *31*, 46-53.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C.A., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C.*, et al.* (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics *4*, 1920-1932.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A.*, et al.* (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

Wang, C.Y., and Filippakopoulos, P. (2015). Beating the odds: BETs in disease. Trends Biochem Sci *40*, 468-479.

Welboren, W.J., van Driel, M.A., Janssen-Megens, E.M., van Heeringen, S.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J *28*, 1418-1428.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

Zeng, L., and Zhou, M.M. (2002). Bromodomain: an acetyl-lysine binding domain. FEBS Lett *513*, 124-128.

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y., Xiang, J.Z., Fei, Z., and Giovannoni, J.J. (2011). High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harbor protocols *2011*, 940-949.

Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M.*, et al.* (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature *478*, 524-528.

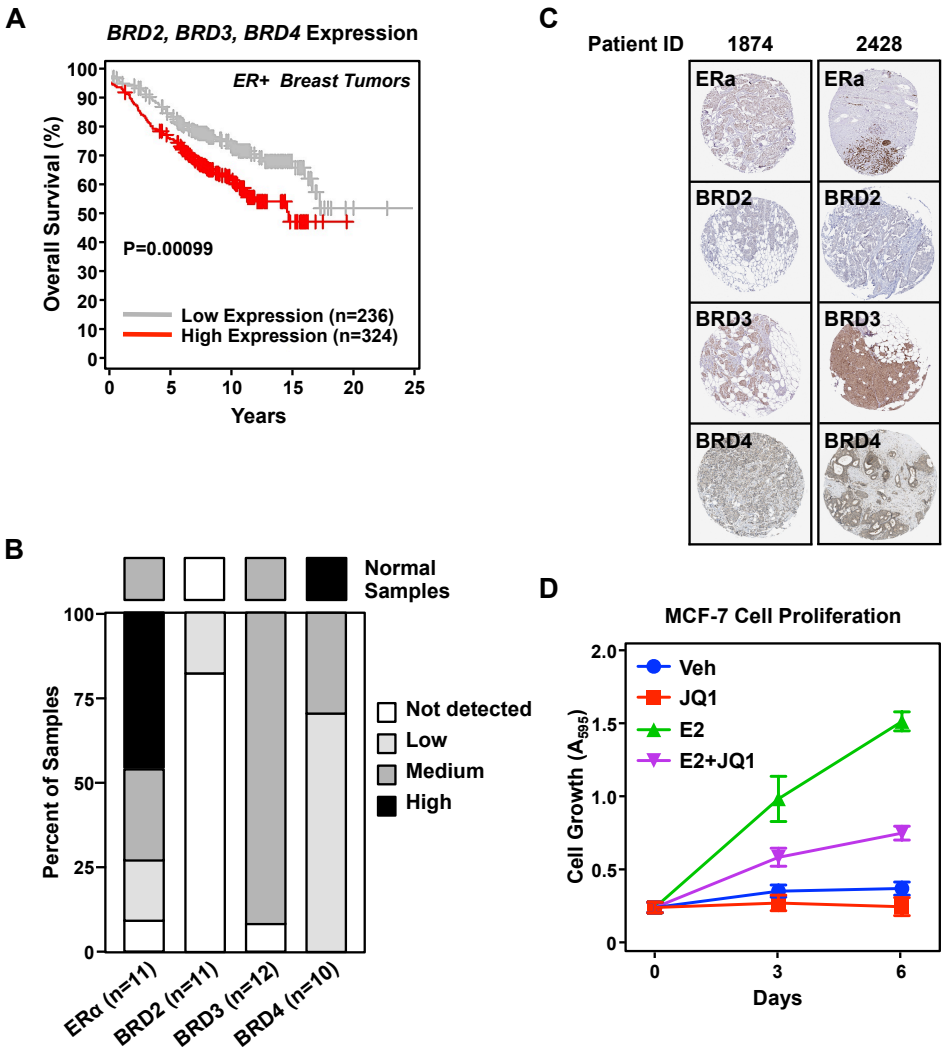


Figure 4.1. The expressions of the BET family members correlate with clinical outcomes for ER-positive breast cancer patients, and the BRD family proteins are required for E2-dependent cell proliferation.

(A) High expression of the BET family proteins correlates with negative clinical outcomes in ER-positive breast cancer patients. Kaplan-Meier plots for overall survival of ER-positive breast tumor patients with high (red) or low (gray) BRD2, BRD3, and BRD4 expression.

(B) Majority of breast tumor samples express BRD3 or BRD4. Quantification of immunohistochemical standing for ERα, BRD2, BRD3, and BRD4 in breast tumor specimens in the Human Protein Atlas. Normal tissues herein represent samples of non-neoplastic and morphologically normal part of tissues that were surgically removed from three individuals with breast cancers.

(C) Representative immunohistochemical staining for ERα, BRD2, BRD3, and BRD4 in ER-positive ductal carcinoma patient samples adapted from the Human Protein Atlas.

(D) JQ1 attenuates E2-depedent growth of MCF-7 cells. Proliferation assays of ER-positive MCF-7 breast cancer cells with vehicle or E2 treatment in the presence or absence of 250 nM JQ1.

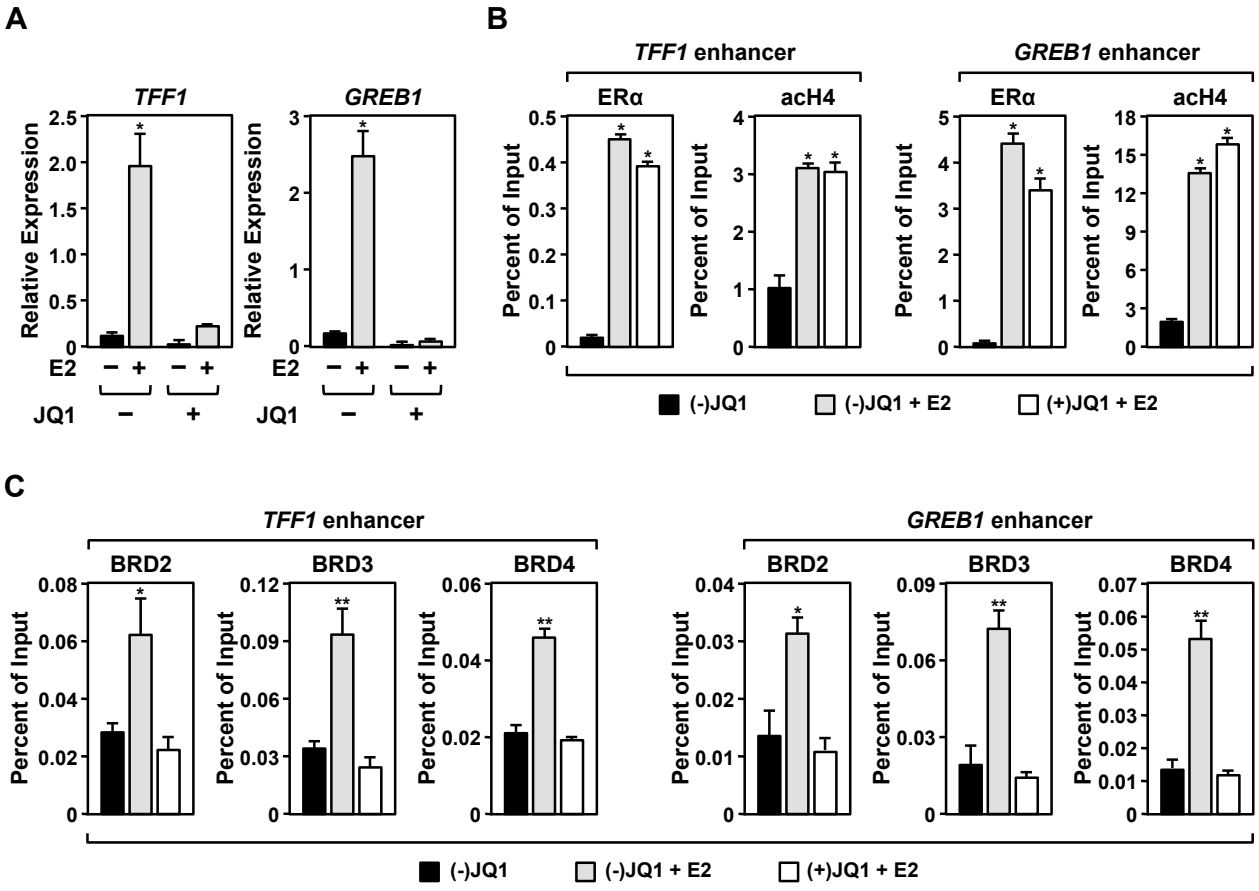


Figure 4.2. The BET family members are recruited at ERα binding sites upon E2-treatment and are required for E2-dependent gene expression.
(A) JQ1 impairs E2-depnedent gene induction in MCF-7 cells. The expression of E2-induced genes in the presence of active (+) or inactive (-) form of JQ1 determined by RT-qPCR. The asterisks indicate significant differences (two-way ANOVA, $p < 0.001$).
(B) JQ1 does not affect neither ERα binding or histone acetylation levels at ERα enhancers. The enrichment of ERα and acetyl H4 at enhancers for E2-responsitve genes in the presence of active (+) or inactive (-) form of JQ1 determined by ChIP-qPCR. The asterisks indicate significant differences from the control condition (one-way ANOVA, $p < 0.001$).
(C) JQ1 inhibits the E2-dependent recruitment of the BET family proteins at ERα enhancers. The effect of JQ1 on BRD2, BRD3, and BRD4 recruitment at ERα enhancers upon E2 treatment determined by ChIP-qPCR. The asterisks indicate significant differences from the control condition (one-way ANOVA, *; $p < 0.05$, **; $p < 0.005$).

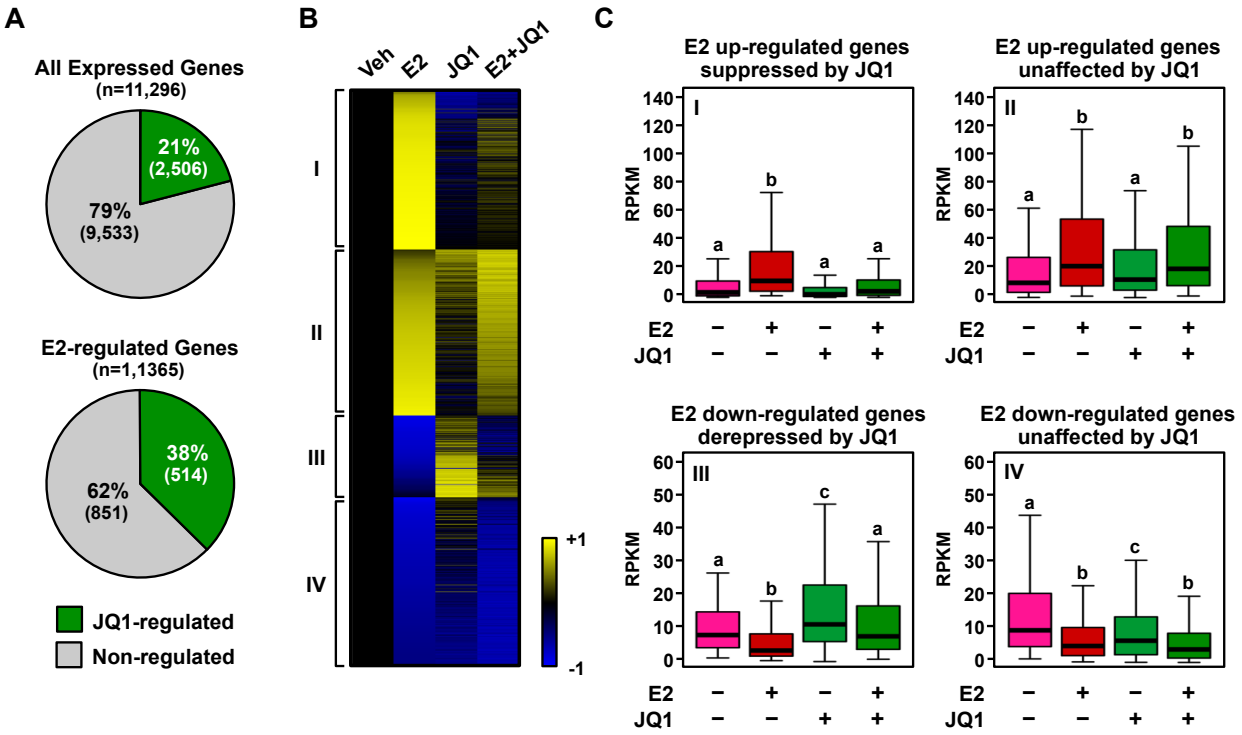


Figure 4.3. The BET family members regulate E2-dependengt gene expression.
(A) E2-regulated genes are disproportionally affected by JQ1. Pie charts showing the fraction of genes whose expression levels were affected by JQ1 among all expressed or E2-regulated genes in MCF-7 cells. The expression levels were determined by RNA-seq.
(B) A heatmap representation of gene expression regulated by E2 in the presence or absence of JQ1. Genes are called statistically significantly regulated when the fold change of expression levels to the vehicle without JQ1 are greater than 2 or less than 0.25 with FDR < 0.1%. Roman numerous indicate four clustered gene groups.
(C) Boxplots of expression levels for the top 50% of highly regulated genes within each group indicated in (B). Letters on the bars indicate significant differences (Wilcoxon rank sum test, $p < 5 \times 10^{-10}$).

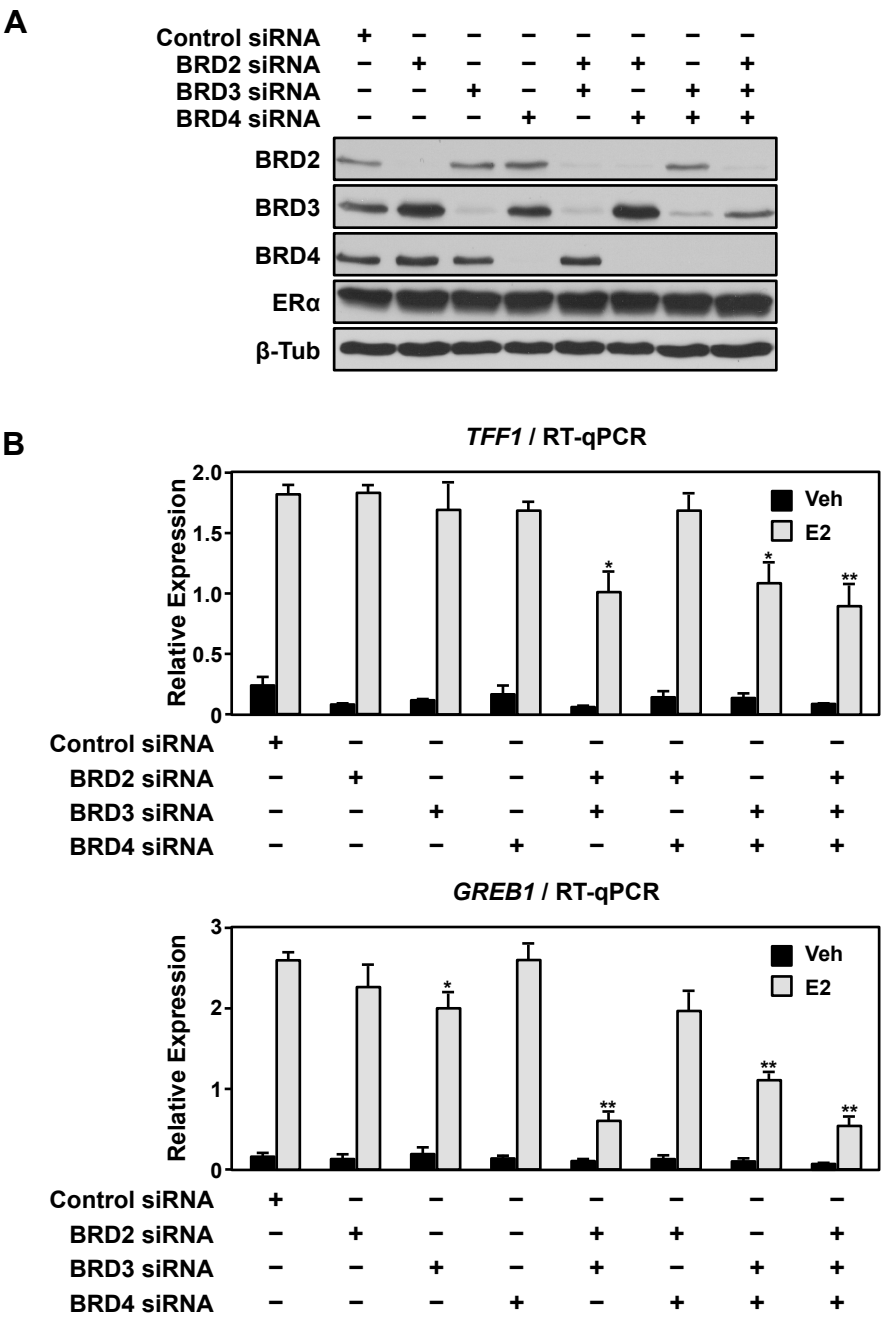


Figure 4.4. BRD3 is a critical member of the BET family proteins for E2-dependent gene induction in MCF-7 cells.

(A) siRNA-mediated knockdown of BRD2, BRD3, and BRD4 individually or in combination. Protein expression was determined by western blot.

(B) E2-induced gene expression is attenuated only when BRD3 is depleted in combination with BRD2, BRD4, or both. E2-dependent gene expression determined by RT-qPCR under the condition that BET family proteins were depleted by siRNA individually or in combination. The asterisks indicate significant differences from control siRNA under the same condition (two-way ANOVA, *; $p < 0.05$, **; $p < 0.005$).

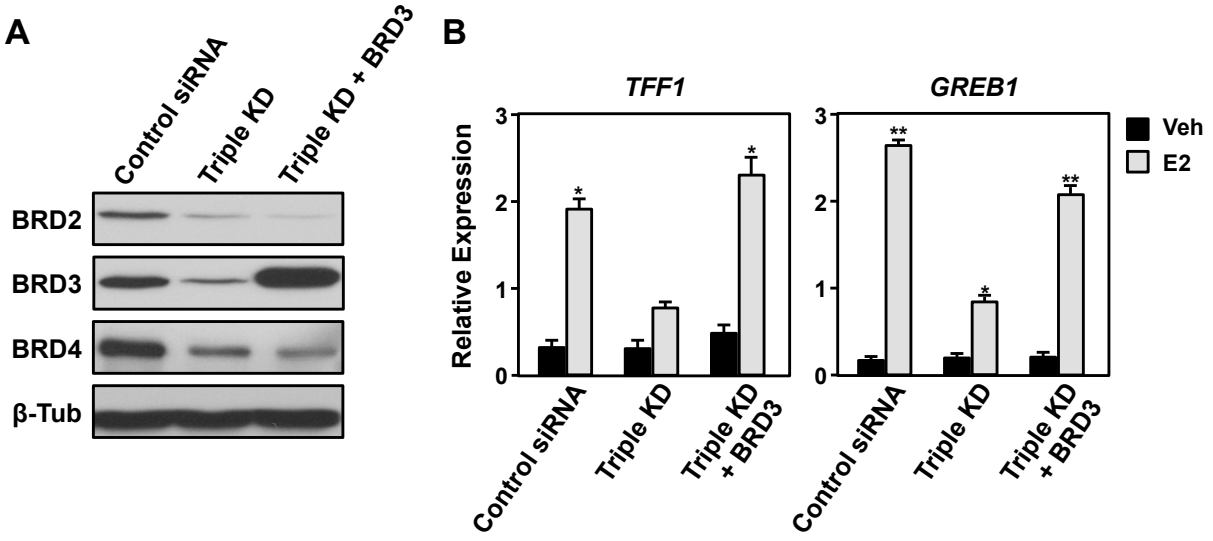


Figure 4.5. BRD3 is sufficient to restore E2-dependent gene expression suppressed by depletion of the BET family proteins.
(A) Western blot showing re-expression of siRNA-resistant BRD3 in MCF-7 cells with the simultaneous knockdown of BRD2, BRD3, and BRD4 by siRNA.
(B) Re-expression of BRD3 rescues E2-induced gene expression impaired by the depletion of the BET family member proteins. E2-induced gene expression in MCF-7 cells determined by RT-qPCR when siRNA-resistant BRD3 was expressed in triple knockdown of BRD2, BRD3, and BRD4. The asterisks indicate significant differences from control siRNA without treatment (two-way ANOVA, *; $p < 0.01$, **; $p < 0.0001$).

Figure 4.6. BRD3 is enriched for a subset of ERα binding sites associated with active enhancer features.
(A) Heatmap representations of ERα, BRD3, acetyl H4, FOXA1, and DNaseI ChIP-seq surrounding ERα peaks. ERα binding sites were classified based on co-localization of BRD3. The genomic scales were indicated on the bottom of − E2 columns. The same scales were applied for + E2 conditions of the corresponding ChIP-seq dataset.
(B) Boxplots of ERα, BRD3, and acetyl H4, FOXA1 ChIP-seq and DNaseI-seq reads 1 kb surrounding ERα peaks with or without BRD3 enrichment. Letters indicate significant differences compared to – E2 condition. (Wilcoxon rank sum test, $p < 0.05$).
(C) Genome browser tracks of ERα, BRD3, and acH4 ChIP-seq and GRO-seq data at GREB1 locus representing ERα binding sites with BRD3.
(D) Number of ERα peaks surrounding ERα binding sites colocalized with BRD3 binned based on the distance from the BRD3-associated or –dissociated ERα binding sites.
(E) Enriched motifs under satellite ERα binding sites surrounding the central ERα binding sites colocalized with or without BRD3.
(F) Boxplots showing the frequencies of ERα motifs within 10 kb from ERα binding sites with or without BRD3. Boxplots of the number of full and half EREs within 10 kb surrounding ERα peaks coocupied with or without BRD3. Letters indicate significant differences compared to – E2 condition. (Wilcoxon rank sum test, $p < 2.2 \times 10^{-16}$).

[Figure 4.6 is on the next page]

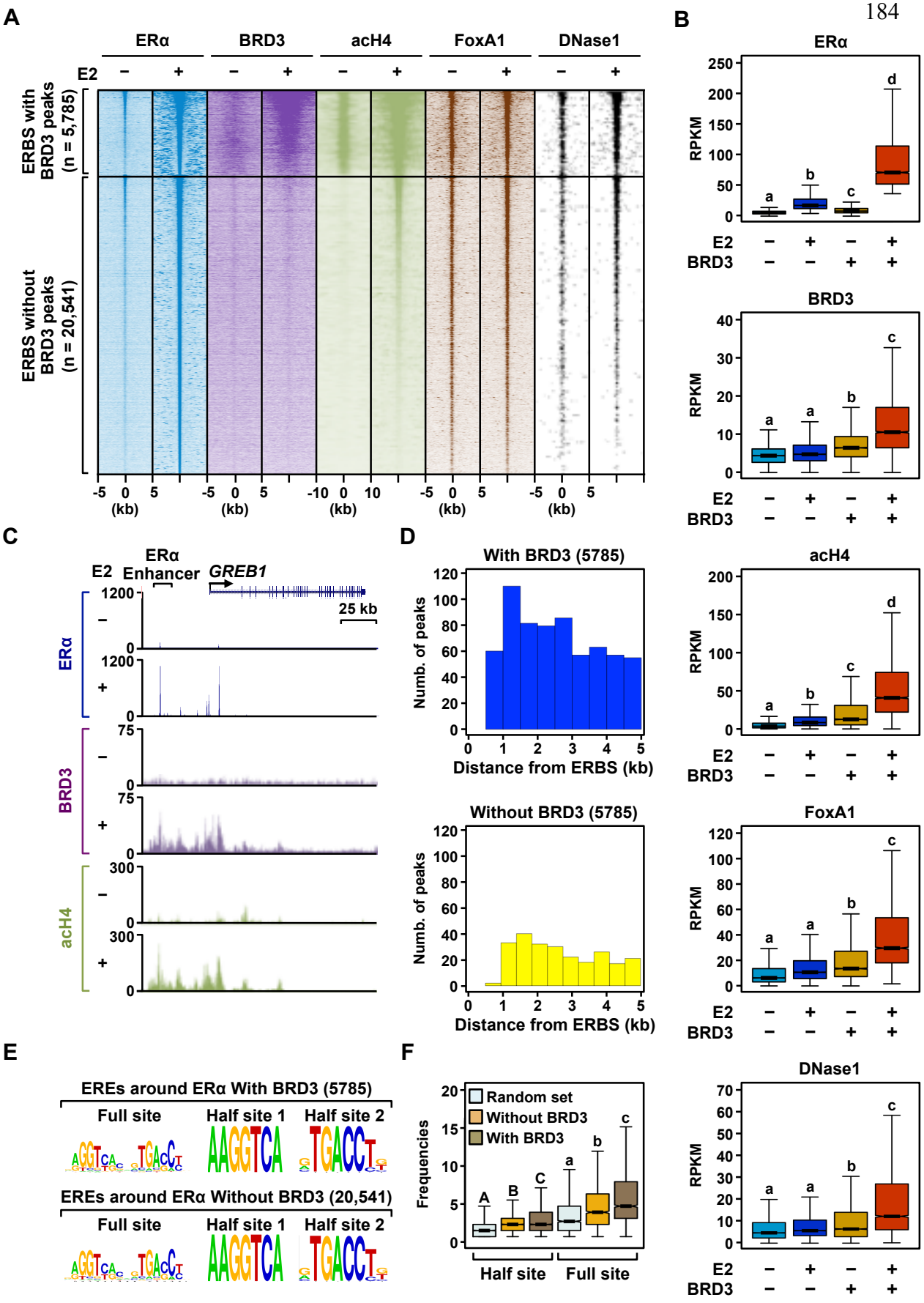
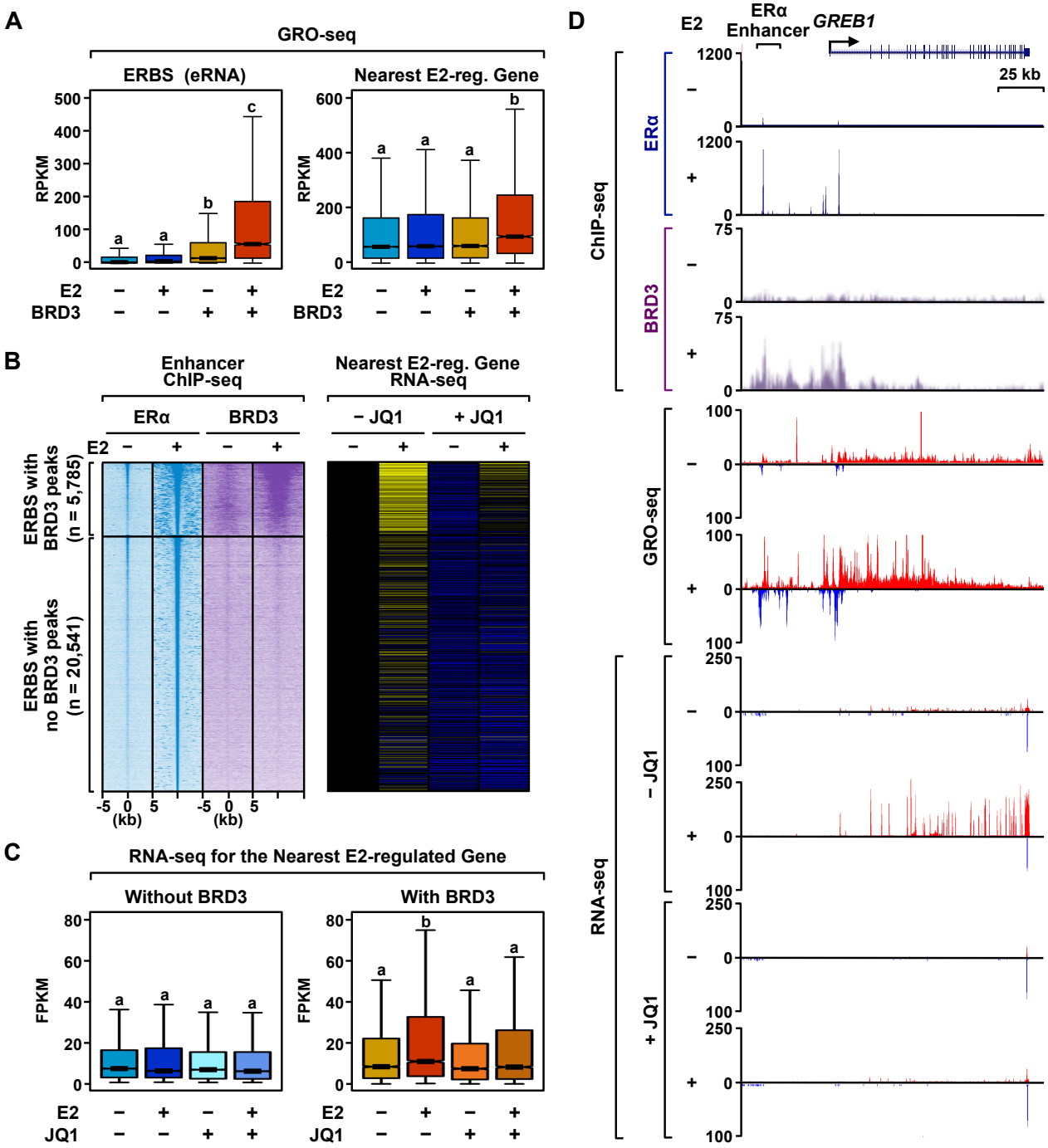

A
ERα
BRD3
acH4
FoxA1
DNase1
E2
ERBS with BRD3 peaks (n = 5,785)
ERBS without BRD3 peaks (n = 20,541)
(kb)
B
ERα
BRD3
acH4
FoxA1
DNase1
RPKM
E2
BRD3
C
ERα
E2
Enhancer
GREB1
25 kb
ERα
BRD3
acH4
D
With BRD3 (5785)
Without BRD3 (5785)
Numb. of peaks
Distance from ERBS (kb)
E
EREs around ERα With BRD3 (5785)
EREs around ERα Without BRD3 (20,541)
Full site
Half site 1
Half site 2
F
Random set
Without BRD3
With BRD3
Frequencies
Half site
Full site

Figure 4.7. ERα binding sites with high BRD3 occupancy are associated with E2-induced genes that are affected by JQ1.
(A) Boxplots showing the transcription levels determined by GRO-seq at ERα enhancers colocalized with or without BRD3 and the nearest E2-regulated genes. Letters indicate significant differences compared to – E2 condition. (Wilcoxon rank sum test, $p < 0.05$).
(B) JQ1-affected E2-induceed genes are enriched for the target genes of ERα enhancers with BRD3 colocalizaiton. Heatmap representations of ERα and BRD3 as shown in Figure 4.6A and RNA-seq of the nearest genes upon E2-treatment in the absence or presence of JQ1.
(C) Boxplots showing RNA-seq reads of E2-regulated genes nearest to the ERα binding sites with or without BRD3 enrichment upon E2-treatment in the absence or the presence of JQ1. Letters on the bars indicate significant differences (Wilcoxon rank sum test, $p < 2.2 \times 10^{-16}$).
(D) Genome browser tracks of ChIP-seq for ERα and BRD3, GRO-seq, and RNA-seq data at GREB1 locus representing ERα binding sites with BRD3.

[Figure 4.7 is on the next page]


A
GRO-seq
ERBS (eRNA)
Nearest E2-reg. Gene
RPKM
E2
BRD3
B
Enhancer ChIP-seq
Nearest E2-reg. Gene RNA-seq
ERα
BRD3
− JQ1
+ JQ1
ERBS with BRD3 peaks (n = 5,785)
ERBS with no BRD3 peaks (n = 20,541)
(kb)
C
RNA-seq for the Nearest E2-regulated Gene
Without BRD3
With BRD3
FPKM
JQ1
D
E2
ERα Enhancer
GREB1
25 kb
ChIP-seq
ERα
BRD3
GRO-seq
RNA-seq
− JQ1
+ JQ1

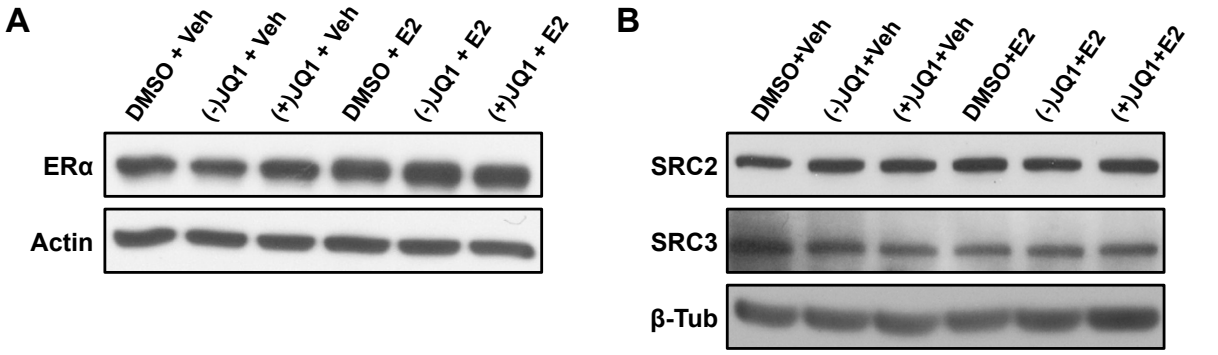


Figure 4.8. JQ1 does not affect the expression of ERα, SRC2, and SRC3.
Western blot showing protein expression of ERα, SRC2, and SRC3 in MCF-7 cells treated with E2 for 45 min. in the absence or presence of JQ1.

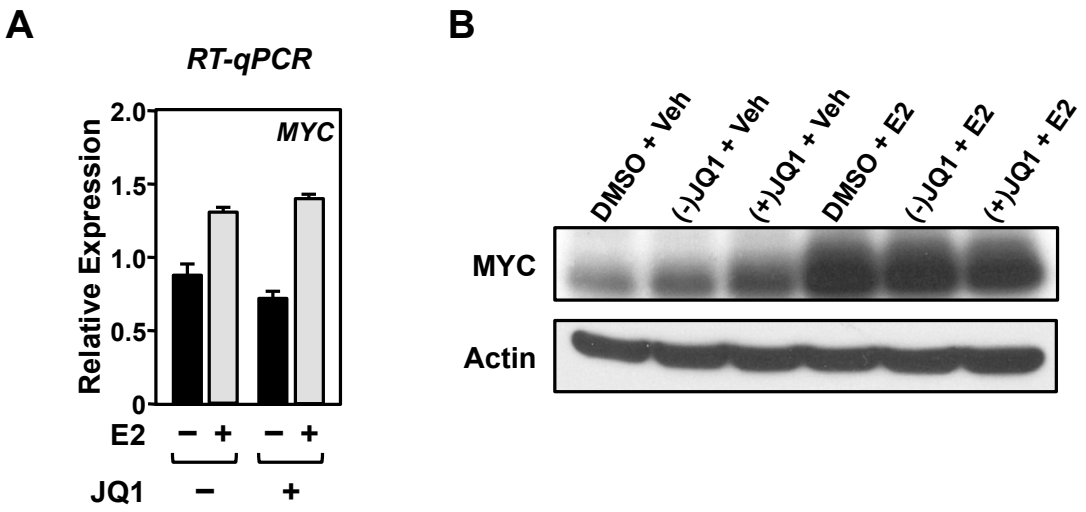


Figure 4.9. E2-dependent MYC upregulation is not affected by JQ1.
(A) RT-qPCR assays showing MYC expression at RNA levels upon E2-treatment in the presence or absence of JQ1.
(B) Western blotting showing MYC protein levels in MCF-7 cells treated with E2 for 6 hrs in the presence or absence of JQ1.

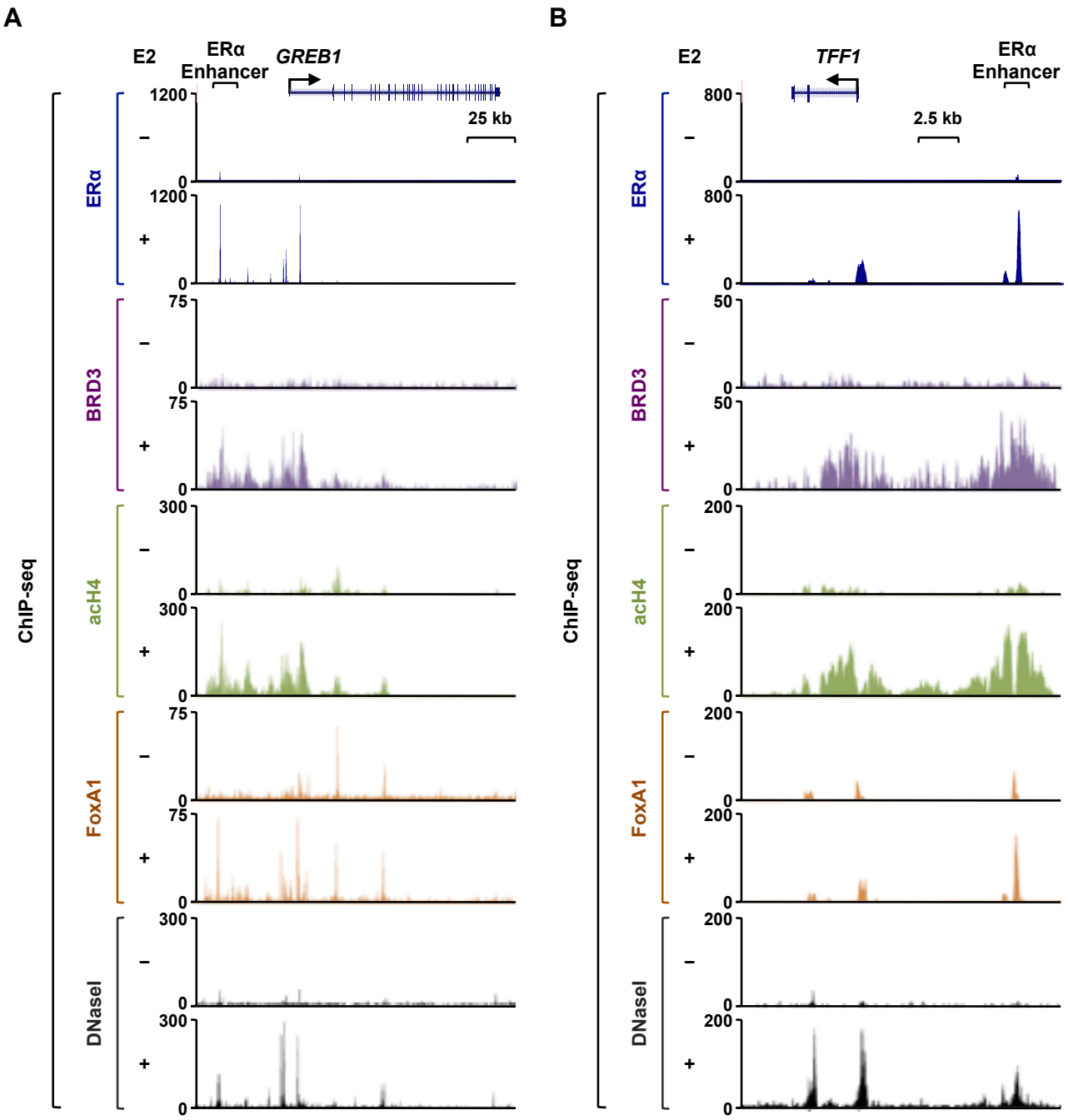


Figure 4.10. BRD3 colocalizes with ERα enhancers.
Genome browser tracks of ERα, BRD3, acetyl H4, FOXA1 ChIP-seq and DNaseI-seq datasets at GREB1 **(A)** and TFF1 **(B)** loci representing ERα enhancer colocalized with BRD3 enrichment.

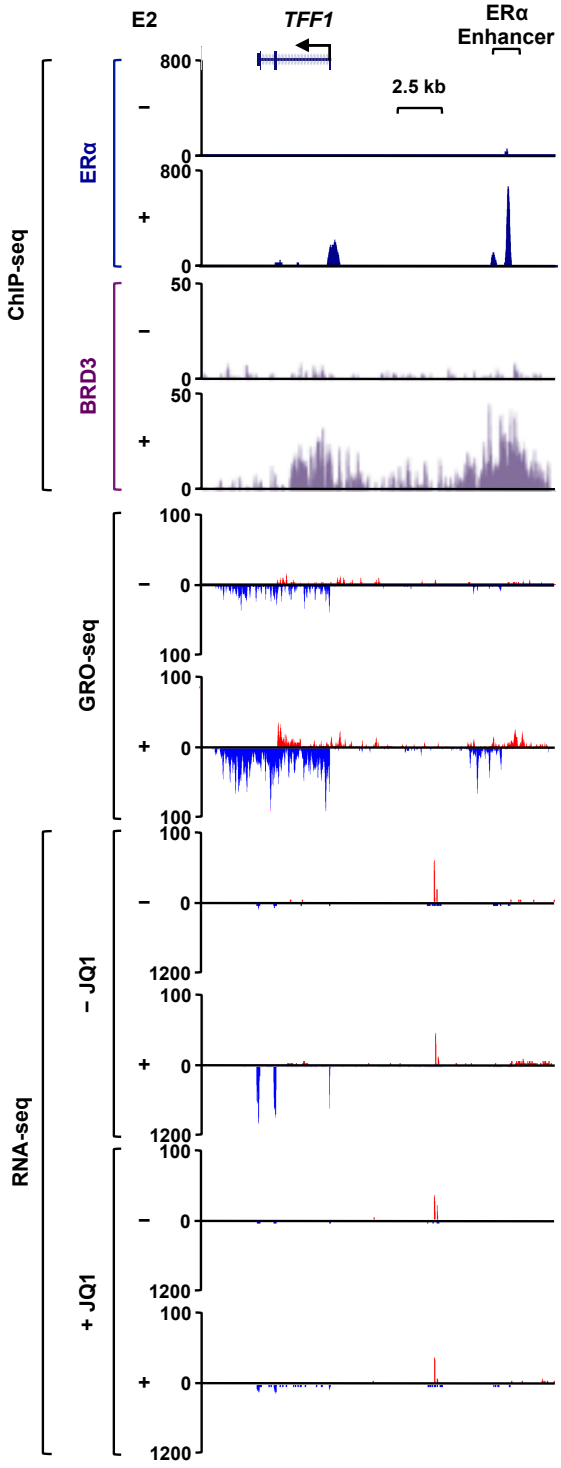


Figure 4.11. E2-induced genes affected by JQ1 are associated with ERα binding sites with BRD3.

Genome browser tracks of ChIP-seq for ERα and BRD3, GRO-seq, and RNA-seq data at TFF1 locus representing ERα binding sites with BRD3 enrichment.

E2 up-regulated genes suppressed by JQ1 (suppression >2 fold)

Category	Term	p Value
GOTERM_BP_FAT	GO:0009719 response to endogenous stimulus	1.2E-5
GOTERM_BP_FAT	GO:0009725 response to hormone	2.5E-5
GOTERM_BP_FAT	GO:0001568 blood vessel development	1.9E-4
GOTERM_BP_FAT	GO:0001944 vasculature development	2.2E-4
GOTERM_MF_FAT	GO:0008092 cytoskeletal protein binding	5.8E-4
GOTERM_BP_FAT	GO:0042445 hormone metabolic process	6.3E-4
GOTERM_BP_FAT	GO:0010033 response to organic substance	2.4E-3
GOTERM_BP_FAT	GO:0048514 blood vessel morphogenesis	2.5E-3
GOTERM_BP_FAT	GO:0010817 regulation of hormone levels	3.0E-3
GOTERM_BP_FAT	GO:0043434 response to peptide hormone	3.3E-3

E2 down-regulated genes derepressed by JQ1(derepression >1.5 fold)

Category	Term	p Value
GOTERM_BP_FAT	GO:0007050 cell cycle arrest	1.7E-5
GOTERM_BP_FAT	GO:0042127 regulation of cell proliferation	1.4E-4
GOTERM_MF_FAT	GO:0005160 transforming growth factor beta receptor binding	3.3E-4
GOTERM_BP_FAT	GO:0045787 positive regulation of cell cycle	1.1E-3
GOTERM_BP_FAT	GO:0060284 regulation of cell development	1.2E-3
GOTERM_BP_FAT	GO:0051726 regulation of cell cycle	1.2E-3
GOTERM_BP_FAT	GO:0007389 pattern specification process	1.3E-3
GOTERM_BP_FAT	GO:0010770 positive regulation of cell morphogenesis involved in differentiation	1.3E-3
GOTERM_BP_FAT	GO:0010718 positive regulation of epithelial to mesenchymal transition	1.3E-3
GOTERM_BP_FAT	GO:0008219 cell death	1.6E-3

Table 1. Gene ontology analysis of JQ1-affected genes

E2-induced genes that are sensitive to JQ1 are enriched for terms related to hormone-dependent signaling including hormone response, hormone metabolism, and hormone regulation. Gene ontology analysis of E2-regulated genes affected and unaffected by JQ1. Top ten GO terms with p-values are listed for each category

CHAPTER FIVE

Development of Single-cell GRO-seq

This chapter describes the outline of experimental designs for the development of single-cell GRO-seq technology. I design the experiments and prepare libraries in collaboration with the Gary Hon laboratory in the Green Center for Reproductive Biology Sciences, who have set up Drop-seq equipment. Once library preparation is optimized and libraries are prepared and sequenced, the data analysis will be performed in the Hon laboratory.

5.1 Summary

Heterogenic gene expression is an intrinsic nature of a population of cells. It is a fundamental question in biology why heterogeneous gene expression occurs and how it is controlled (Elowitz et al., 2002; Levsky et al., 2002; Raser and O'Shea, 2004). The current technologies available, however, have limitations in observing the link between heterogenic gene expression and heterogenic usage of regulatory elements at the single-cell level (Kolodziejczyk et al., 2015; Marr et al., 2016). In this chapter, I describe strategies for the development of an innovative technology, single-cell global run-on sequencing (scGRO-seq), which will allow us to determine enhancer activities and their target gene transcription on the same cell. scGRO-seq is a promising strategy for better understanding the molecular basis of transcription regulation.

5.2 Introduction

Transcription Enhancers

Transcription regulation is a central regulatory mechanism for controlling a wide array of biological processes (Heinz et al., 2015; Szutorisz et al., 2005). Complex transcription profiles are orchestrated by relatively small numbers of transcription factors (TFs) (Spitz and Furlong, 2012; Young, 2011). TFs bind to thousands of DNA regulatory elements, called transcription enhancers, to form enhancer complex by recruiting transcription coregulators and histone modifiers, resulting in the establishment of enhancer-specific chromatin environments such as open chromatin states and enrichment of enhancer-enriched histone modifications including H3K4me1 and H3K27ac (Calo and Wysocka, 2013; Ong and Corces, 2011). Enhancer activation in turn induces transcription not only at the target gene promoters but also at the enhancers, which leads to the production of non-coding transcripts named enhancer transcripts (eRNAs) (De Santa et al., 2010; Kim et al., 2010).

Enhancer Annotation Based on eRNAs

There are several suggested functions of enhancer transcripts (Hsieh et al., 2014; Melo et al., 2013; Mousavi et al., 2013; Pnueli et al., 2015; Schaukowitch et al., 2014); however, we are still missing a unifying model of the functions and mechanisms. Nonetheless, it has became widely accepted that GRO-seq, which detects the genomic location and orientation of actively transcribing RNA polymerases, is the best available method to measure eRNAs expression (Fang et al., 2014; Lai and Shiekhattar, 2014), partly

due to the nature of eRNAs as unstable transcripts (Kim and Shiekhattar, 2016). Using GRO-seq datasets, our previous study and others demonstrated enhancer transcription at TF binding sites across the genome (De Santa et al., 2010; Franco et al., 2015; Hah et al., 2013; Kim et al., 2010). Furthermore, we showed that enhancer transcription robustly marks active enhancers that highly correlate with active enhancer features, including enhancer-enriched histone modifications and coregulators, enhancer-promoter chromatin interactions, and the target gene transcription (Hah et al., 2013). To this end, we previously developed a computational pipeline to annotate active enhancers by determining eRNA production without knowing the transcription factor that is responsible for the eRNA and the target gene expression (Hah et al., 2013; Nagari et al., 2017).

Heterogeneity in Expression Regulation

Single-cell gene expression analysis has demonstrated that gene expression is highly heterogeneous even among a pure population of cells (Kurimoto and Saitou, 2010; Leek and Storey, 2007; Saliba et al., 2014; Tang et al., 2010). Drop-seq and its derivatives has become a popular method to study steady-state RNA levels in individual cells (Junker and van Oudenaarden, 2015; Klein et al., 2015; Kolodziejczyk et al., 2015; Macosko et al., 2015; Shekhar et al., 2016). In this method, each cell is captured within a single droplet containing a single bead attached with an adaptor oligo with a barcode sequence that is unique to each bead. By capturing transcripts from a single cell to adaptor oligonucleotides on a single bead, transcripts from a given cell can be identified by their unique barcodes that are

different from other cells. Transcripts uniquely barcoded for each cell are pooled and subjected to RNA-seq library preparation (Klein et al., 2015; Macosko et al., 2015).

Similarly, a few methods have been developed to study regulatory mechanisms of heterogeneous gene expression at the single-cell level, including single-cell ATAC-seq and ChIP-seq. These studies revealed that heterogenic activity of DNA regulatory elements across individual cells in a population (Buenrostro et al., 2015; Smallwood et al., 2014; Wu et al., 2016).

A significant limitation in the currently available tools that determine enhancer activity and the target gene expression lies on the fact that one cannot obtain the both types of information in the same single cell (Gawad et al., 2016; Macaulay and Voet, 2014; Marr et al., 2016). The insufficiency on these methods makes it unfeasible to understand the connection between enhancer activity and the target gene transcription in the single-cell level. Although we have demonstrated a strong correlation between enhancer activity and the target gene transcription using populations of cells, we are still missing definitive evidence for the connectivity between the two events at the single cell level. If the two events are synchronized, we hypothesize that heterogenic gene expression results from heterogenic enhancer usage among a cell population. Also another hypothetic scenario will be where a given gene is controlled by differential enhancers within a cell population. In addition, activity of a given enhancer may not reflect to the target gene transcription in the same manner among individual cells in a population (Figure 5.1). By using enhancer transcription as a surrogate for enhancer activity, scGRO-seq will allow us to simultaneously determine enhancer activity and the target gene transcription at the single-cell level.

Comprehensive analyses of scGRO-seq data promise to uncover the missing link between enhancer activity and the target gene transcription.

Moreover, scGRO-seq provides additional advantages compared to the currently available single-cell sequencing analyses. Compared to single-cell RNA-seq, scGRO-seq will provide direct measurement of transcription outcomes at high temporal resolution. In addition, the methods for regulatory element analysis such as single-cell ATAC-seq and ChIP-seq, severely suffer from the limited dynamic ranges of measurement and the high false negative error rates. These methods are based on sequencing fragmented genomic DNA (gDNA). These problems arise since there are only two copies of genomic DNA in a given diploid cell available as the starting material for each genomic loci (Buenrostro et al., 2015; Smallwood et al., 2014; Wu et al., 2016). Enhancer activity determined by scGRO-seq will reside in a wider dynamic range of measurement compared to these methods. Each transcription unit is transcribed by multiple RNA polymerases at a given moment each of which produces a nascent transcript that serves as a starting material for scGRO-seq library preparation. Library preparation starting with greater numbers of molecules per loci helps to avoid high false negative error rates associated with the gDNA-based single-cell sequencing methods.

By unrevealing transcription states at high-temporal resolution at the single-cell level, scGRO-seq will become a powerful tool to define biological processes in more details, to identify rear population of cells, and to understand key regulatory mechanisms of transcription (Grun et al., 2015; Tirosh et al., 2016; Treutlein et al., 2016; Zhou et al., 2016).

5.3 Significance and Hypothesis

In this chapter, I describe a strategy for developing a new technology, scGRO-seq, which will allow us to determine active transcription at the single-cell level by measuring nascent transcripts. Using scGRO-seq data analyzed by rigorous bioinformatic pipelines, we will annotate active enhancers based on eRNA transcription at the single-cell level. In addition, integrate analyses on enhancer activities and the target gene transcription at the single-cell level will reveal the molecular mechanisms of enhancer function. Collectively, scGRO-seq will expand our understanding on the molecular mechanisms of enhancer functions by bringing light on the missing link between enhancer activity and the target gene transcription.

5.4 Approaches and Experimental Plans

5.4.1 Biological Systems

For the development of scGRO-seq, we will use MCF-7 ERα-positive breast cancer cells treated in the time course of E2 for reasons (1) MCF-7 cells allow us to isolate run-on competent nuclei for GRO-seq (Hah et al., 2011), (2) estrogen provokes rapid and extensive transcription responses in MCF-7 cells (Hah et al., 2011), and (3) the Kraus lab has prepared time course GRO-seq libraries for E2-treated MCF-7 cells that allow us to compared the transcriptome observed in single-cell levels and in a population of cells (Danko et al., 2013).

These advantages with MCF-7 cells will allow me to avoid potential obstacles at least in part during the development of scGRO-seq technology, but yet to perform informative analysis with biological significance.

Upon completion of successful development of the protocol, scGRO-seq will be also applied to mouse embryonic stem cells (mESCs) during a time course of differentiation in embryonic bodies to examine high temporal resolution of transcriptional regulation during embryonic differentiation. Additional urging question that can be answered using scGRO-seq is allele-specific regulation of target gene transcription by enhancers. In this case, scGRO-seq will be prepared from mouse tissues of an offspring mouse generated by crossing two different strains. The data will be analyzed utilizing single nucleotide polymorphisms (SNPs) unique to each strain.

5.4.2 Overview and Experimental Plans

The strategies described herein for the development of scGRO-seq take an approach by bridging GRO-seq and modified Drop-seq using a click chemistry reaction (Fig. 5.2). Clickable cytidine triphosphate analog (rCTP) will be used in nuclear run-on in a population of intact nuclei (Birts et al., 2014; Core et al., 2008). Single nucleus will then be isolated in droplets, and rCTP analog-incorporated nascent transcripts will be clicked by copper-catalyzed alkyne-azide cycloaddition (CuAAC) to 5'-azide of the custom synthesized adaptor oligonucleotides that is pre-annealed to the barcoded Drop-seq oligonucleotides conjugated to a bead in individual droplets. Nascent transcripts uniquely barcoded for each single-nuclei will be pooled and converted to sequencing libraries using modified Drop-seq protocol (Fig.

5.2) (Klein et al., 2015; Macosko et al., 2015). Sequenced date will be analyzed to understand the molecular bases of enhancer function.

Nuclear Run-on and Encapsulation of a Single Nucleolus

Nuclei will be isolated, and nuclear run-on will be performed as a population of intact nuclei in a reaction solution containing adenosine/uridine/guanosine triphosphate (rATP, rUTP, rGTP), and 3'-(O-Propargyl)-rCTP (Fig. 5.2) (Birts et al., 2014; Core et al., 2008; El-Sagheer et al., 2011; Hah et al., 2011). 3'-(O-Propargyl)-rCTP incorporation into nascent transcripts by RNA polymerases will stall transcription due to the lack of the 3' hydroxyl group of ribose sugar required for elongation (Fig. 5.3). This transcription stalling enables us to determine active transcription events at a near-nucleotide resolution. Upon completion of nuclear run-on and incorporation of 3'-(O-Propargyl)-rCTP, single nuclei will be isolated and encapsulated into individual droplets containing a single bead using the Drop-seq equipment. Each bead are conjugated to Drop-seq oligonucleotides with unique barcodes that is pre-annealed through poly-dT sequences to the custom synthesized oligonucleotides of which 5'-terminal is modified with 5'-azide-deoxythymidine (5'-azide-dT) (Table 2, Fig. 5.2) (Klein et al., 2015; Macosko et al., 2015).

Capturing Nascent Transcripts Using Click Chemistry

Nascent transcripts will be released from the chromatin in individual droplets by a cocktail of restriction enzymes (FspEI, MspJI, McrBC, and DnpI) that specifically digest methylated DNA (Siwek et al., 2012; Venitt et al., 1976; Zheng et al., 2010). Methylation-

specific endonucleases will fragment genomic DNA without digesting the synthetic adapter oligonucleotides linked to the beads in droplets. In each droplet, the alkyne group of 3'-(O-Propargyl)-rCTP incorporated on the 3' end of newly synthesized RNA molecules will be clicked to the azide group of 5'-azide-dT on the 5' end of the custom synthesized adapter oligonucleotides by the CuAAC click chemistry reaction (Fig. 5.2; Table 2) (Gierlich et al., 2006; Seela and Sirivolu, 2006). The downstream of the 5'-azide-dT, the custom synthesized adaptor oligonucleotides also contain poly(dA) sequences that is pre-annealed to poly(dT) sequences of the Drop-seq oligonucleotides, which is conjugated to beads (Figure 5.2) (Birts et al., 2014; Gibson et al., 2016). The Drop-seq oligonucleotides contain barcodes identical on a given bead but unique to each bead as in Drop-seq. This barcoding strategy will allow us to cluster sequencing reads from individual nuclei by encapsulating a single beads and a single nucleus in each droplet through the Drop-seq equipment (Klein et al., 2015; Macosko et al., 2015).

Library Preparation Using Modified Drop-seq Protocol

The captured nascent transcripts will be subjected to library preparation by following the Drop-seq protocol with some modifications as described below (Klein et al., 2015; Macosko et al., 2015). After the click reaction, nascent transcripts linked to the barcoded beads will be released from the droplets and reverse transcribed in a pool using pre-annealed Drop-seq oligonucleotides for priming and Nextera TSO oligonucleotides for template switching in place of the standard Drop-seq TSO (Table 2) (Birts et al., 2014; Chen et al., 2014). In the standard Drop-seq protocol, P7 adaptor required for sequencing will be

incorporated by Tn5 tagmentation reaction. During this process, Tn5 favors DNA fragments > 600 bp as their substrates. However, a great portion on nascent transcripts will be considerably shorter than 600 bp, including transcripts generated from promoter-proximally paused polymerases. The use of Nextera TSO, which contains P7 adaptor sequence, in place of the Drop-seq standard TSO ensures incorporation of P7 adaptor on the 3' terminal of cDNA independent of the tagmentation reaction. The first PCR reaction following reverse transcription and Exonuclease I treatment will be performed using Nextera TSO PCR primer in addition to SMART PCR primers at the equimolar ratio. The second PCR reaction following to the tagmentation will be performed using New-P5-SMART PCR hybrid oligo and Nextera N70X oligo as in the standard Drop-seq protocol, where "X" in N70X represents an integer for barcode identifier.

The quality of the PCR amplified libraries will be accessed by Agilent TapeStation. Once passing the quality control, the libraries will be quantified by Qubit Fluormeter and pooled for sequencing on Illumina NextSeq 400 instrument. For the analysis, another set of libraries will be prepared by the same strategy using a population of nuclei without the single nuclei isolation step as a control.

5.4.3 Data Analysis

Alignment and Transcript Calling

The origin of nuclei for each sequencing read will be identified by de-barcoding the barcode sequences unique to each bead. De-barcoded reads will be subjected to proper quality control steps, followed by mapping to the reference genome. Uniquely mapped reads

will be used for the subsequent analysis (Klein et al., 2015; Macosko et al., 2015). First, a universe of transcripts will be called *de novo* by groHMM, a computational analysis tool for GRO-seq datasets developed in the Kraus lab (Chae et al., 2015), using the datasets from a population of nuclei. Next, putative active enhancers will be annotated based on short, bidirectional intergenic transcripts, a characteristic of eRNAs, using a computational pipeline for GRO-seq-based enhancer calling (Nagari et al., 2017). Finally, datasets from single nuclei will be used to determine transcription levels of each transcription unit in the universe of *de novo* called transcripts (Hah et al., 2013). The analysis described above will provide us with a list of genes with their expression levels and a list of enhancers with their activity levels measured as a production of eRNAs in individual nuclei.

Integrative Analysis of Genes Expression and Enhancer Activity

Using the list of expressed genes and activate enhancers, we will focus on our further analysis to determine (1) the extent of heterogeneity on enhancer usage across cells and across conditions, (2) how enhancer avidity is connected to the target gene transcription at the single cell level, and (3) the transcription regulatory mechanisms at the high temporal resolution to uncover previously unknown regulatory mechanisms. We will first determine differential gene/eRNA expression across cells and conditions using statistical analysis package edgeR (Robinson et al., 2010). Differential enhancer activities will be analyzed based on differential eRNA expression. In addition, the activity of each enhancer will be linked to the expression of the nearest neighboring genes as the target gene expression (Danko et al., 2013, Franco, 2015 #32; Franco et al., 2015; Hah et al., 2013). The analyzed

data will be visualized in appropriate forms including heatmaps, boxplots, and dot plots (Franco et al., 2015; Hah et al., 2013).

5.5 Selected Results of Optimization Steps

5.5.1 Nuclear Run-on Incorporates 3-(O'-Propargyl)-CTP in Nascent Transcripts

Objective

One of the key aspects of the scGRO-seq strategy is the use of 3'-(O-propargyl)-CTP. This nucleotide analog was chosen because of its biocompatible nature. In particular, the triazole linkage formed by CuAAC reaction between 3'-(O-propargyl)-ribonucleotie and 5' azide-containing RNA has been shown to be reverse transcribed by (1) MMLV-derived reverse transcriptase *in vitro* (Figure 5.3) (Chen et al., 2014) and (2) RNA Polymerase II in mammalian cells (Figure 5.3) (Birts et al., 2014). Moreover, a plasmid vector composed of DAN fragments conjugated by CuAAC reaction between 3'-(O-propargyl)-dCTP and 5' azide is amplified in *Escherichia coli* upon transformation (Figure 5.3) (El-Sagheer et al., 2011). 3'-(O-propargyl)-CMP is commercially available at ChemGenes (RP-3302). The nucleoside, 3'-(O-propargyl)-CMP, was converted to the nucleotide, 3'-(O-propargyl)-CTP, by the custom synthesis service at ChemGenes.

Equally important characteristics of 3'-(O-propargyl)-CTP was the fact that RNA polymerases will be stalled upon tis incorporation due to the lack of 3' hydroxyl group (Figure 5.3). This feature provides an advantage to determine the genomic localization of

actively engaged RNA polymerases at near-nucleotide resolution. 3'-(O-propargyl)-NTP-mediated transcription stalling ensures that run-on reaction time will minimally contribute as a source of experimental variability. In addition, the immediate stalling of transcription will allow us to improve the protocol in the future to determine the transcription snapshot at the single nucleotide resolution in the single-cell level using 3'-(O-propargyl)-NTP. To achieve this goal, Drop-seq beads needs to be custom synthesized so that it contains P5 sequence, bead barcode, molecular barcode, followed by a non-genomic unique sequence in place of the poly(A) sequence of the current Drop-seq beads that serves as an annealing site for an adaptor that contains an 5'-azide modification. 3'-(O-propargyl)-NTP-incorporated nascent transcripts will be the adaptor oligo with 5' azide and library preparation will proceed as the original scGRO-seq protocol.

The following experiment was performed to test (1) if 3'-(O-propargyl)-CTP can be incorporated in nascent transcripts during run-on reactions using isolated nuclei, and (2) the integrity of RNA upon incubation in Cu^{+}-containing click reaction mix. Briefly, nuclear run-on was performed in the presence of 3'-(O-propargyl)-CTP for the test reaction or biotin-11-CTP (CTP analog used in PRO-seq) for a positive control (Mahat et al., 2016). Biotin-azide was clicked to 3'-(O-propargyl)-CTP by CuAAC reaction using run-on nuclei in the presence or absence of copper catalyst. Total RNA was isolated and subjected to dot blotting with streptavidin-HRP. Additionally the RNA was examined on RNA Tape Station to determine the integrity after incubation in CuAAC reaction mix at the elevated temperature for the prolonged time.

Materials and Methods

MCF-7 cells growing on a 150 mm^2 plate were washed twice with ice-cold PBS and scraped in PBS containing 0.02 U/μL SUPERase•In (Ambion) and collected by centrifugation at 500x g for 5 min. at 4°C. The cells were resuspended in Sucrose Permeabilization Buffer (10 mM Tris•HCl pH 7.4, 300 mM Sucrose, 10 mM KCl, 5 mM $MgCl_2$, 1 mM EDTA, 0.05% Tween-20, 0.1% NP-40, 1 mM DTT, 0.02 U/μL SUPERase•In, and 1x complete protease inhibitor cocktail) and equilibrated on ice for 5 min. The cells were collected by centrifugation and washed with 1 mL Sucrose Permeabilization Buffer once. The permeabilized cells were resuspended in 1 mL Sucrose Permeabilization Buffer and counted using hemocytometer. Cells were collected by centrifugation and resuspended in Nuclei Freezing Buffer to 10 x 10^6 nuclei/100 μL (50 mM Tris•HCl pH 8.3, 5 mM MgCl2, 40% glycerol, 0.1 mM EDTA, and 0.02 U/μL SUPERase•In), aliquoted 50 μL per tube, and stored at -80°C.

The permeabilized cells stored at -80°C were thawed on ice, mixed with pre-heated 50 μL reaction mix [2x Nuclear Run-on Buffer (20 mM Tris•HCl pH 8.0, 300 mM KCl, 5 mM $MgCl_2$, and 0.1% Tween-20), 50 μM Biotin-11-CTP or 3'-(O-propargyl)-CTP (ChemGenes, custom synthesized), 250 μM rATP/rUTP/rGTP each, 1 mM DTT, 0.8 U/μL SUPERase•In, and 1x complete protease inhibitor cocktail], and incubated at 37°C for 5 min. After run-on reaction, 900 mL ice-cold Glycerol Lysis Buffer (10 mM Tris•HCl pH 7.4, 10% glycerol, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 1 mM EDTA, 0.5% NP-40, 1 mM DTT, 0.02 U/μL SUPERase•In, and 1x complete protease inhibitor cocktail) was added to the run-on reaction and the nuclei were collected by centrifugation at 200x g for 5 min. at 4°C. The nuclei were

washed twice with Glycerol Lysis Buffer, resuspended in 100 µL click reaction mix [±3 µL/reaction pre-mixed Catalyst Solution Mix (33.3 mM $CuSO_4$, 166.7 mM THPTA, and 166.7 mM sodium ascorbate), 1x K^+ Tris Lysis Buffer (50 mM Tris•OAc pH 7.9, 50 mM KOAc, 0.1% Sarcosyl, and 3% Ficoll PM-400), 1 mM GTP, 1x Activator Solution (NEB R0661S), 50 µL biotin-azide, 5 mM Aminoguanidine, 1 U FspEI (NEB), 1 U MspJI (NEB), 2 U McrBC (NEB), 4 U DpnI (NEB), and 0.4 U/µL SUPERase•In], and incubated at 37°C for 3 hr in 0.2 mL tube while shaking.

After clicking reaction, total RNA was isolated using 300 µL TRIzol LS (Invitrogen) by following the manufacture's instruction. The total RNA was ran through P-30 column (BioRad) to remove unincorporated nucleotide, normalized, and subjected to dot blot. Serial dilution of RNA was placed on a nitrocellulose membrane, blocked in 5% BSA/TBST, and blotted with Streptavidin-HRP (ThermoFisher, N100) in 5% BSA/TBST. RNA integrity was examined using RNA Tape (Agilent, 5067-5576) on 2200 TapeStation (Agilent).

Results

To examine if 3'-(O-propargyl)-CTP is incorporated during the nuclear run-on reaction, I performed run-on reaction using permeabilized MCF-7 cells in the presence of the CTP analog. 3'-(O-propargyl)-CTP-incorporated nascent transcripts were clicked to biotin-azide by CuAAC reaction, which was then detected using streptavidin-HRP. Nuclear run-on reaction performed in the presence of biotin-11-CTP, the CTP analog used in PRO-seq (Mahat et al., 2016), served as a positive control because (1) it had been shown to incorporate

in nascent transcripts, and (2) its detection by streptavidin-HRP is independent of the clicking reaction (Figure 5.4A).

The level of biotin content in the RNA sample incorporated with 3'-(O-propargyl)-CTP and clicked to biotin in the presence of copper catalyst was similar to that in the RNA samples incorporated with biotin-11-CTP, indicating the efficient incorporation of 3'-(O-propargyl)-CTP in the nascent transcripts and the efficient CuAAC reaction-mediated biotin conjugation (Figure 5.4B). Importantly the 3'-(O-propargyl)-CTP-incorporated RNA samples clicked in the absence of copper exhibited significantly lower signal, demonstrating that the detected biotin levels was dependent on CuAAC reaction (Figure 5.4B).

Another key finding from this experiment was that RNA was degraded after CuAAC reaction as determined using RNA Tape on TapeStation (Figure 5.4C), presumably by Cu^{2+}-mediated RNA fragmentation. This result leads to a modification of the scGRO-seq protocol. In the original experimental plan, 3'-(O-propargyl)-CTP-incorporated nascent transcript was assumed to maintain the integrity, leading to reverse transcription all the way to the 5'-end of nascent transcripts. The 3' adaptor was planed to be incorporated in cDNA through a template switching reaction by reverse transcriptase, which depends on the 5' m7G cap of RNA. The intact nascent transcripts were going to be amplified by PCR with low cycle numbers. Double-stranded DNA was going to be fragmented by Tn5, which incorporates P7 adapter on the 3'end of DNA generated from >600 nt transcripts during fragmentation (Macosko et al., 2015). However, with the Cu^{2+}-catalyzed fragmentation of RNA, the experimental plan was modified so that P7-containing RNA adapter (VRA5) will be ligated to the 5' end of nascent transcripts following to CuAAC reaction, RppH-mediated

5’ cap removal, and T4 Polynucleotide kinase-mediated phosphorylation of 5’end (Figure 5.2) (Mahat et al., 2016).

5.5.2 Capture of Alkyne-containing Small Molecules by 5’Azide-modified Adapter Oligo through Copper-catalyzed Click Reaction on Beads

Objective

The experiment in this section was performed to ensure that the 5’ azide-containing adapter oligo that is hybridized to the Drop-seq oligo captures alkyne-containing molecules thought CuAAC reaction while maintaining hybridization in the conditions compatible to scGRO-seq library preparation.

Materials and Methods

5’ Biotinylated Drop-seq oligo was custom synthesized by IDT. 42 μM 5’ biotinylated Drop-seq oligo and dT-polyT adapter oligo with or without 5’ azide modification (Gene Link) was annealed in 1x Annealing Buffer (10 mM Tris•HCl pH 7.9, 10 mM NaCl, and 1 mM EDTA) by incubating at 95°C for 2 min., 94°C for 45 sec., and decreasing by 0.5°C every cycle for 139 cycles. The annealed oligos were then diluted to 20 μM in 1x Annealing Buffer.

100 μL Dynabeads MyOne Streptavidin T1 beads (Invitrogen) per reaction was washed once with 300 mM Salt Wash Buffer (10 mM Tris•HCl pH 7.4, 300 mM NaCl, 1 mM EDTA, and 0.1% NP-40), once with Base Wash Solution (0.1 N NaOH and 50 mM NaCl), and once with 100 mM NaCl. The beads were resuspended in 200 μL Biotin-binding

Buffer (10 mM Tris•HCl pH 7.8, 100 mM NaCl, and 1 mM EDTA). 50 µL of 20 µM annealed oligo was mixed with 150 µL Biotin-binding Buffer. The diluted oligo and washed beads were mixed together and incubated for 30 min. at RT while rotating. After binding reaction, the beads were washed once with 500 µL High Salt Wash Buffer (10 mM Tris•HCl pH 7.4, 2 M NaCl, 1 mM EDTA, and 0.1% NP-40) and twice with 500 µL K^+ Wash Buffer (50 mM Tris•OAc pH 7.9, 50 mM KOAc, and 0.1 % NP-40). At the last wash, beads susupended in 500 µL wash buffer were divided into 1x 100 µL and 1x 400 µL and placed on a magnetic stand, and the wash buffer was aspirated.

Catalyst solution was prepared by mixing 1 µL of 100 mM $CuSO_4$, 1 µL of 500 mM THPTA, and 1 µL of 500 mM sodium ascorbate. A click reaction mix were prepared by mixing ±1.2 µL pre-mixed Catalyst solution, 1x K^+ Tris Lysis Buffer, 50 µM Cy3-alkyne, and 5 mM Aminoguanidine to the total volume of 40 µL. The biotin-oligo-bound T1 beads were resuspended in the click reaction mix, and transferred to 0.2 mL reaction tube for incubation at 37°C for 1 hr with mild vortexing. After clicking reaction, the beads were washed twice with High Salt Wash Buffer, once with K^+ Wash Buffer, and twice with Low Salt Wash Buffer (5 mM Tris•HCl pH 7.4, 100 mM NaCl, and 0.1% NP-40). The beads were resuspended with 100 µL 1x RQ-1 DNase buffer containing 20 µL RQ-1 DNase (Promega), and incubated at 37°C for 1 hr. Supernatant containing released Cy3 was transferred to a 96-well plate (sample labeled as Off Beads). Beads were washed twice with 1x RQ-1 DNase buffer, resuspended with 100 µL 1x RQ-1 DNase buffer, and transferred to a 96-well plate (sample labeled as On Beads). A serial dilution of Cy3-alkyne was prepared to generate a slandered curve. Fluorescent signal was determined using a plate reader and the

percent recovery was calculated. Note that in the clicking reaction, Cy3-alkyne was at 50 μM, while the theoretical concentration of annealed oligos are at 10 μM with 100 % annealing efficiency and at the binding efficiency to T1 beads suggested by the manufacture; thus the maximum theoretical values for recovery rate is 16% for the reaction using 80% of the beads.

Results

To test if the 5' azide-containing adapter oligo captures alkyne-containing molecules while annealed to the Drop-seq beads, I performed a Cu^{+}-catalyzed click reaction using the adapter oligo annealed to the Drop-seq oligo as an azide donor and Cy3-alkyne as an alkyne donor. First, the 5' azide-containing adapter oligo was annealed to another oligo of the sequence identical to the Drop-seq oligo with 5' biotin modification. After annealing, the double-stranded (ds) oligs were immobilized on streptavidin-coated beads. Upon oligo immobilization, the beads were divided into 20 and 80 % of the total beads, and subjected to a Cu^{+}-catalyzed click reaction with a saturating amount of Cy3-alkyne. Once capturing Cy3-alkyne, the beads were washed and DNA was digested by DNase to release Cy3 from the beads for measurement of fluorescent signals. Equivalent reaction was performed alongside using an adapter oligo without 5' azide modification as a negative control (Figure 5.5A).

Significantly higher fluorescent signaling was detected when the reaction was curried out with the adapter oligo with 5' azide modification, suggesting CuAAC-dependent capture and release of Cy3-alkyne (Figure 5.5B). In addition, fluorescent signal was approximately three times stronger in the reaction using 80% beads compared to that using 20% of the beads

when the 5' azide adapter oligo was used. This dose dependent increase of fluorescent signal was not evident when the adapter oligp lacked 5' azide modification, further supporting the specificity of the readout (Figure 5.5B). In conclusion, the experiment in this section shows that the 5' azide-containing adapter oligo captures alkyne-containing molecules through CuAAC reaction while maintaining the annealing to the Drop-seq oligo on beads.

5.5.3 Copper-mediated Fragmentation of gDNA

Objective

Fragmentation of genomic DNA in Droplets is a critical step in the workflow of scGRO-seq protocol. Based on RNA-seq from chromatin-associated fractions, nascent transcripts are expected to maintain its association with chromatin in the nuclei. The nascent transcripts need to be released from the chromatin for their capture by the Drop-seq beads through CuAAC reaction; otherwise, sticky chromatin may congregate on the beads, leading to high background from the genomic DNA contamination. It is also important to note that the fragmentation of the genomic DNA needs to be accomplished without damaging the oligos directly or indirectly attached to the Drop-seq beads. To achieve these goals, it was planed to supplement a cocktail of methylation-dependent endonucleases (FspEI, MspJI, McrBC, and DpnI) in the lysis buffer that flows with beads in the microfluidic device in the original scGRO-seq protocol. Methylation-dependent endonucleases will digest genomics DNA while keeping the synthetic oligos on the beads intact.

Materials and Methods

Digestion of gDNA by a cocktail of methylation-dependent endonucleases. 60 ng of genomic DNA isolated and purified from MCF-7 cells was incubated at 37°C for 2 hours in a mock click reaction mix [3 µL/reaction pre-mixed Catalyst Solution Mix (33.3 mM $CuSo_4$, 166.7 mM THPTA, and 166.7 mM sodium ascorbate), 1x K^+ Tris Lysis Buffer with or without Mg^{2+} (50 mM Tris•OAc pH 7.9, 50 mM KOAc, 0.1% Sarcosyl, 3% Ficoll PM-400, and ± 5 mM $MgCl_2$), 1 mM GTP, 1x Activator Solution (NEB R0661S), and 5 mM Aminoguanidine].

gDNA digestion with additional nucleases. 5 million nuclei were thawed on ice. 500 µL Glycerol Lysis Buffer was added to the nuclei and aliquoted into 5 x 100 µL. The nuclei was collected to by centrifugation and the nuclei was resuspended in 52 µL mock click reaction mix [±3 µL/reaction pre-mixed Catalyst Solution Mix (33.3 mM $CuSo_4$, 166.7 mM THPTA, and 166.7 mM sodium ascorbate), 1x K^+ Tris Lysis Buffer (50 mM Tris•OAc pH 7.9, 50 mM KOAc, 0.1% Sarcosyl, and 3% Ficoll PM-400), 1 mM GTP, 1x Activator Solution (NEB R0661S), 5 mM Aminoguanidine, and 0.4 U/µL SUPERase•In]. 1.75 µL of each reaction mix was transferred and pooled to a tube containing 200 µL Buffer D (100 mM Tris•HCl pH 8.0, 4 M Guanidine Thiocyanate, 25 mM Sodium Citrate, pH 7.9, 0.5% Sarkosyl, 100 mM β-Mercaptoethanol). The indicated combination of 4 µL methylation-dependent restriction enzyme mix [1 U FspEI (NEB, R0662S), 1U MspJI (NEB, R0661S), 2 U McrBC (NEB, M0272S), 4 U DpnI (NEB, R0176S)], 2 µL λ Exonuclease (NEB, M0293S), 2 µL Exonuclease I (NEB, M0262S), and/or H_2O to the total volume of 8 µL of enzyme mix was added to each reaction tube. The reaction mix was incubated at 37°C for 5

hr while shaking. 10 μL of the reaction mix was transferred to a tube containing 200 μL Buffer D every hour. After 5 hr incubation, 200 μL phenol-chloroform-isoamyl acid was added to each tube of sample in 200 μL Buffer D, vortexed at low setting, and centrifuged at the maximum speed at 4°C for 15 min. The aqueous layer was transferred to a tube containing 200 μL chloroform, vortexed at low setting, and centrifuged at the maximum speed at 4°C for 15 min. The aqueous layer was transferred to a new tube. 12 μL of 5 M NaCl and 1 μL GlycoBlue (Ambion, AM9515) were added and mixed by tapping the tube. 600 μL ethanol added, incubate on ice for 10 min., and centrifuged at the maximum speed at 4°C for 20 min. The pellet was washed with 75% ethanol, air dried, and resuspended in 12 μL H_2O. 5 μL of the sample was treated with RNase (Roche) in NEB buffer 2 (NEB) for 37°C for 1hr. Size of genomic DNA was analyzed on 1% agarose gel.

Results

CuAAC click reaction will be carried out in the buffer containing the copper chelator THPTA. To avoid the competition for THPTA binding, it is ideal that the click reaction buffer does not contain cations other than copper. However, it is expected that the activities of restriction endonucleases for gDNA digestion is magnesium dependent. To test the magnesium dependency, DNA digestion was first tested in the presence or absence of magnesium using isolated gDNA.

To test if methylation-dependent restriction enzyme requires magnesium for their catalytic activity, isolated gDNA was incubated with click reaction buffer in the presence or absence of magnesium. To our surprise, the gDNA was sufficiently digested in the absence

of magnesium (Figure 5.6A). I speculate that copper is able to replace the coenzymatic function of magnesium for the endonucleases tested here. Additional unanticipated observation from this experiment is that gDNA is degraded even in the absence of the endonuclease cocktail (Figure 5.6A). As Cu^{2+} is known to damage biomolecules, it might be possible that the copper in the reaction is degrading gDNA in the absence of enzyme.

After several experiments to optimize the scGRO-seq protocol, I found that the methylation-dependent restriction enzyme cocktail was insufficient for adequate extent of gDNA digestion. To increase the digest efficiency, I included λ Exonuclease, 5' phosphate-dependent 5' to 3' double-strand DNA nuclease, and Exonuclease I, 3' to 5' single-strand DNA exonuclease (substrates generated upon λ Exonuclease-dependent gDNA digestion), in addition to the methylation-dependent endonuclease cocktail to increase digestion efficiency.

To test a suitable condition for gDNA digestion, additional digestion test was performed using nuclei in the click reaction buffer supplemented with the methylation-dependent endonuclease cocktail, λ Exonuclease, and/or Exonuclease I. A time course digestion experiment confirmed the time-dependent fragmentation of gDNA in all the conditions (Figure 5.6B). As consistent with the pervious experiment using isolated gDNA, the reaction in the absence of any nucleases showed the most efficient gDNA digestion (Figure 5.6B). To determine if the enzyme storage buffer in which the enzymes are delivered in has any effects on copper-dependent DNA degradation, I reconstituted the enzyme storage buffer without enzymes and performed gDNA test with or without the storage buffer. The addition of the enzyme storage buffer impaired copper-mediated DNA degradation (data not shown). To this end, I decided not to use any enzymes to digest gDNA and release nascent

transcript by copper-mediated chromatin fragmentation. Since copper-mediated DNA fragmentation does not discriminate gDNA and adapter oligos, it further needs to be tested the effect of the CuAAC click reaction on the adapter oligos.

5.6 Expected Outcomes and Interpretations

In this chapter, I described an integrative approach incorporating state-of-art scGRO-seq technology and a well-tuned computational pipeline to address key biological questions. The proposed technology will uncover previously unknown the relationship between enhancer activities and the target gene transcription at the single-cell level. We expect that scGRO-seq technology will reveal the molecular basis on heterogeneous gene expression by bridging the heterogeneous enhancer usage and gene transcription. Collectively, this approach will significantly expand our knowledge on the molecular basis of gene transcription.

5.5 References

Birts, C.N., Sanzone, A.P., El-Sagheer, A.H., Blaydes, J.P., Brown, T., and Tavassoli, A. (2014). Transcription of click-linked DNA in human cells. Angew Chem Int Ed Engl *53*, 2362-2365.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486-490.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Chae, M., Danko, C.G., and Kraus, W.L. (2015). groHMM: a computational tool for identifying unannotated and cell type-specific transcription units from global run-on sequencing data. BMC Bioinformatics *16*, 222.

Chen, X., El-Sagheer, A.H., and Brown, T. (2014). Reverse transcription through a bulky triazole linkage in RNA: implications for RNA sequencing. Chem Commun (Camb) *50*, 7597-7600.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol Cell *50*, 212-222.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

El-Sagheer, A.H., Sanzone, A.P., Gao, R., Tavassoli, A., and Brown, T. (2011). Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli. Proc Natl Acad Sci U S A *108*, 11338-11343.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Fang, B., Everett, L.J., Jager, J., Briggs, E., Armour, S.M., Feng, D., Roy, A., Gerhart-Hines, Z., Sun, Z., and Lazar, M.A. (2014). Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. Cell *159*, 1140-1152.

Franco, H.L., Nagari, A., and Kraus, W.L. (2015). TNFalpha signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. Mol Cell *58*, 21-34.

Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat Rev Genet *17*, 175-188.

Gibson, B.A., Zhang, Y., Jiang, H., Hussey, K.M., Shrimp, J.H., Lin, H., Schwede, F., Yu, Y., and Kraus, W.L. (2016). Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. Science *353*, 45-50.

Gierlich, J., Burley, G.A., Gramlich, P.M., Hammond, D.M., and Carell, T. (2006). Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. Organic letters *8*, 3639-3642.

Grun, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251-255.

Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res *23*, 1210-1223.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol *16*, 144-154.

Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S.*, et al.* (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc Natl Acad Sci U S A *111*, 7319-7324.

Junker, J.P., and van Oudenaarden, A. (2015). Single-cell transcriptomics enters the age of mass production. Mol Cell *58*, 563-564.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, T.K., and Shiekhattar, R. (2016). Diverse regulatory interactions of long noncoding RNAs. Curr Opin Genet Dev *36*, 73-82.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol Cell *58*, 610-620.

Kurimoto, K., and Saitou, M. (2010). Single-cell cDNA microarray profiling of complex biological processes of differentiation. Curr Opin Genet Dev *20*, 470-477.

Lai, F., and Shiekhattar, R. (2014). Enhancer RNAs: the new molecules of transcription. Curr Opin Genet Dev *25*, 38-42.

Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet *3*, 1724-1735.

Levsky, J.M., Shenoy, S.M., Pezo, R.C., and Singer, R.H. (2002). Single-cell gene expression profiling. Science *297*, 836-840.

Macaulay, I.C., and Voet, T. (2014). Single cell genomics: advances and future perspectives. PLoS Genet *10*, e1004126.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M.*, et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nature protocols *11*, 1455-1476.

Marr, C., Zhou, J.X., and Huang, S. (2016). Single-cell gene expression profiling and cell state dynamics: collecting data, correlating data points and connecting the dots. Curr Opin Biotechnol *39*, 207-214.

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Leveille, N., Kalluri, R.*, et al.* (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell *49*, 524-535.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell *51*, 606-617.

Nagari, A., Murakami, S., Malladi, V.S., and Kraus, W.L. (2017). Computational Approaches for Mining GRO-Seq Data to Identify and Characterize Active Enhancers. Methods Mol Biol *1468*, 121-138.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283-293.

Pnueli, L., Rudnizky, S., Yosefzon, Y., and Melamed, P. (2015). RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin alpha-subunit gene. Proc Natl Acad Sci U S A *112*, 4369-4374.

Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811-1814.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Saliba, A.E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic acids research *42*, 8845-8860.

Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell *56*, 29-42.

Seela, F., and Sirivolu, V.R. (2006). DNA containing side chains with terminal triple bonds: Base-pair stability and functionalization of alkynylated pyrimidines and 7-deazapurines. Chem Biodivers *3*, 509-514.

Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M.*, et al.* (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell *166*, 1308-1323 e1330.

Siwek, W., Czapinska, H., Bochtler, M., Bujnicki, J.M., and Skowronek, K. (2012). Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. Nucleic acids research *40*, 7563-7572.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nature methods *11*, 817-820.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet *13*, 613-626.

Szutorisz, H., Dillon, N., and Tora, L. (2005). The role of enhancers as centres for general transcription factor recruitment. Trends Biochem Sci *30*, 593-599.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K., and Surani, M.A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nature protocols *5*, 516-535.

Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M.G.*, et al.* (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M.*, et al.* (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391-395.

Venitt, S., Kirkland, D.J., Searle, C.E., and Harnden, D.G. (1976). Letter: Safety of hair dyes. Lancet *1*, 641-642.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W.*, et al.* (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature *534*, 652-657.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940-954.

Zheng, Y., Cohen-Karni, D., Xu, D., Chin, H.G., Wilson, G., Pradhan, S., and Roberts, R.J. (2010). A unique family of Mrr-like modification-dependent restriction endonucleases. Nucleic acids research *38*, 5527-5534.

Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z.*, et al.* (2016). Tracing haematopoietic stem cell formation at single-cell resolution. Nature *533*, 487-492.

Name	Sequence (5' to 3')	Note
Drop-seq oligonucleotides	5'-Bead-Linker-TTTTTTTGAATTCAAGCAGTGGTATCAACGCAGAGTAAGCTTACJJJJJJJJJJJNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	J = bead barcode; N = moleuclar barcode
Custom synthesized adapter oligonucleotides	(5' azide)TAAAAAAAAAAAAAAAAAAAAAAAAAAAA*A	(5'azide)T, 5'-azide-deoxythymidine ; *,phosphorothioate bond
Nextera TSO	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGrGrGrG	Replace with TSO
Nextera TSO PCR primer	GTCTCGTGGGCTCGGAGATG	Supplement in PCR reaction 1 with SMART PCR primer
5' RNA adapter (VRA5)	CCUUGGCACCCGAGAAUUCCA	5' RNA adapter
RP1 PCR Primer	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCG*A	Universal PCR primer
RPI1 PCR Primer	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCC*A	Sample barcord #1 contained PCR primer

Table 2. Oligonucleotide sequences for scGRO-seq

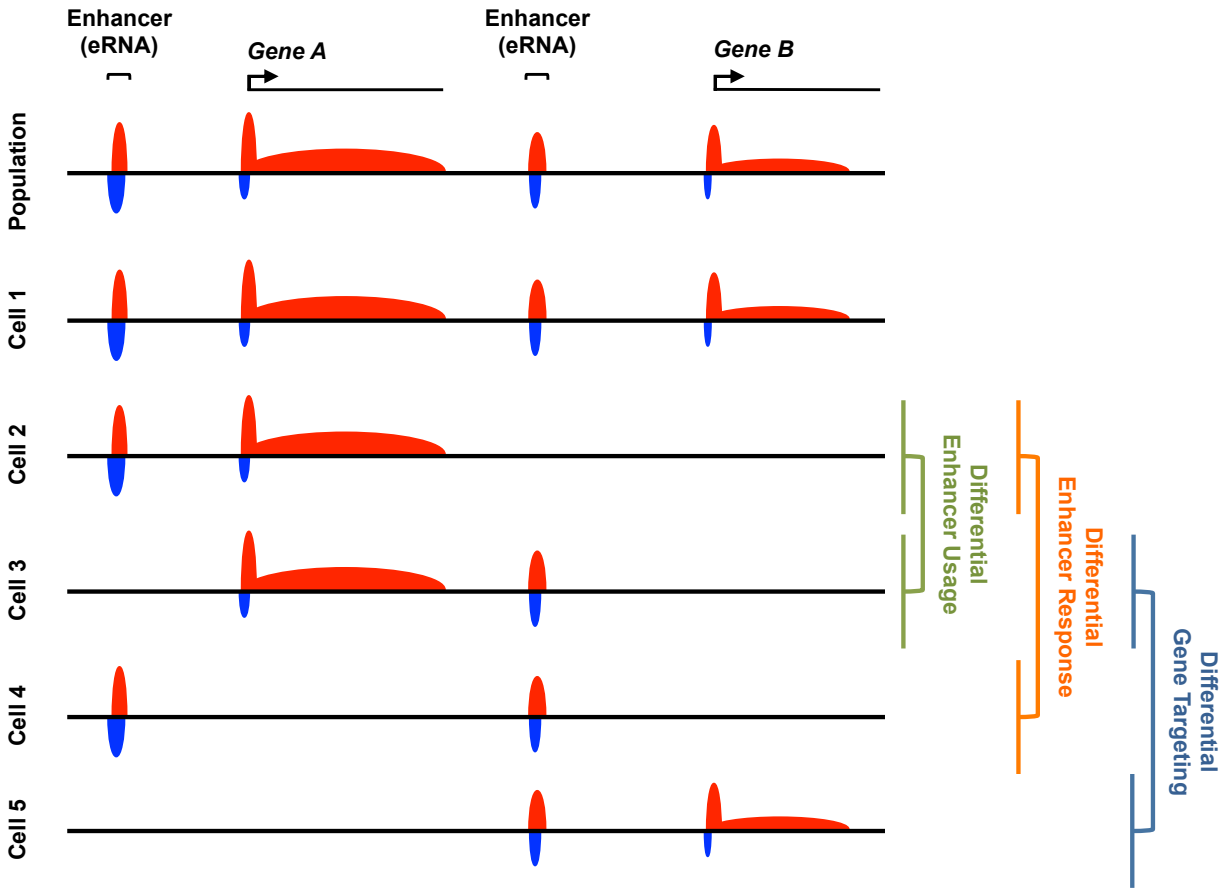


Figure 5.1. Schematics comparing patterns of enhancer activity and gene transcription in a population and single cells.
Schematics of genome browser tracks in a model locus with GRO-seq data prepared from a population of cells versus single cells. While it is possible that some cells show patterns of transcription similar to that in a population, many cells may express a subset of transcripts as exemplified in Cell 2 to 5.

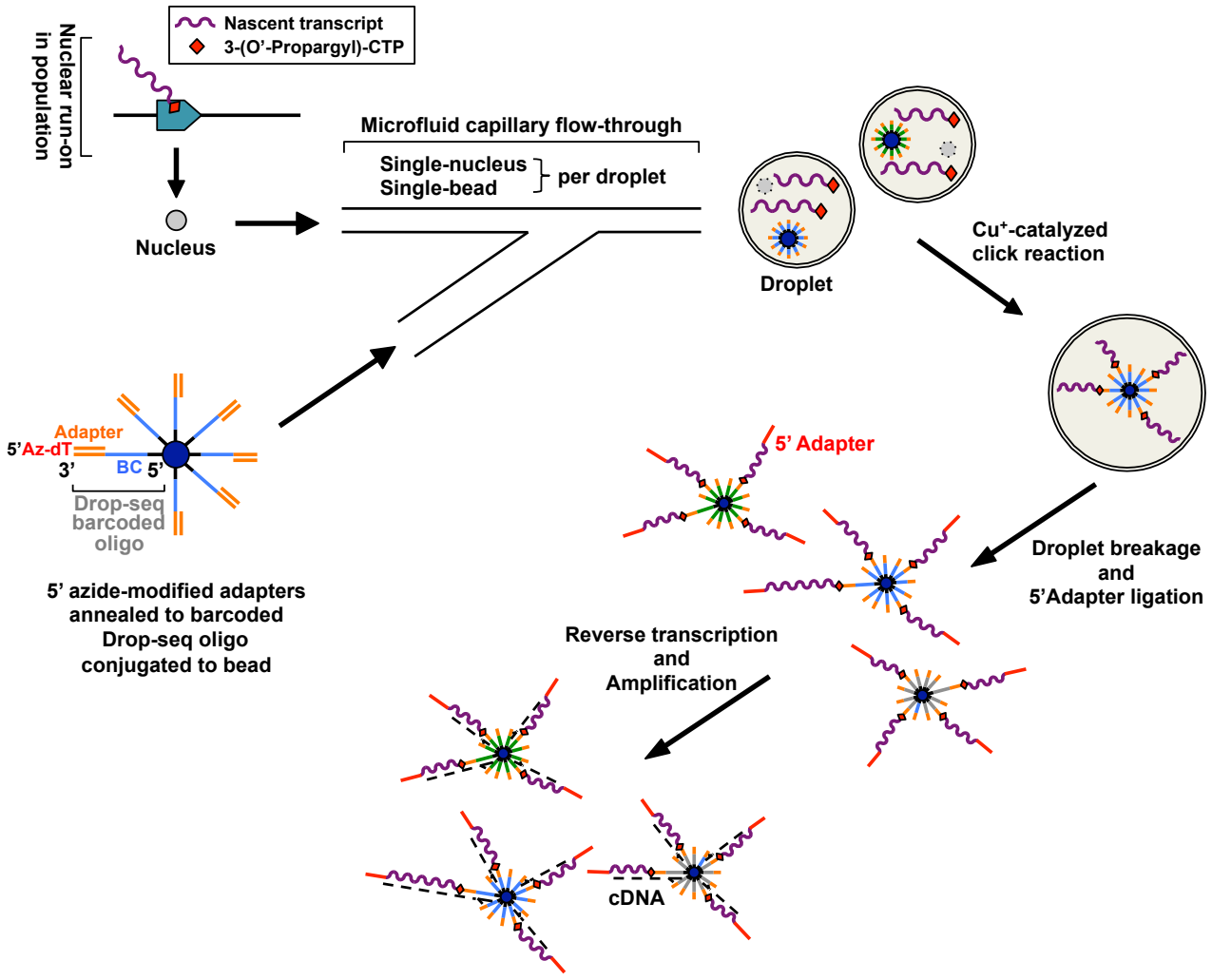


Figure 5.2. The workflow of scGRO-seq library preparation protocol.
Isolated nuclei are used for nuclear run-on in the presence of 3'-(O-propargyl)-CTP along with rATP/rGTP/rUTP. In a separate tube, the 5'-azide-modified adapter oligo is annealed to the Drop-seq oligo that is directly conjugated to the beads. The Drop-seq oligo on a given bead contains a unique barcode sequence that allows tracking sequence reads originated from each bead. Post-run-on nuclei and beads are introduced in separate channels on a microfluidic capillary device in such a way that a single nucleus and a single bead are encapsulated in one droplet. Upon capturing nascent transcripts in droplets through CuAAC click reaction between O-propargyl group of the CTP analog and 5' azide on the adapter oligo, the droplets are broken and nascent RNA on the beads are pooled. 5' RNA oligo are ligated to the nascent transcripts followed by the first strand synthesis by reverse transcription primed from the Drop-seq oligo on the beads. The cDNA library is amplified and subjected to sequencing on the Illumina platform using custom sequencing primers as in the Drop-seq protocol.

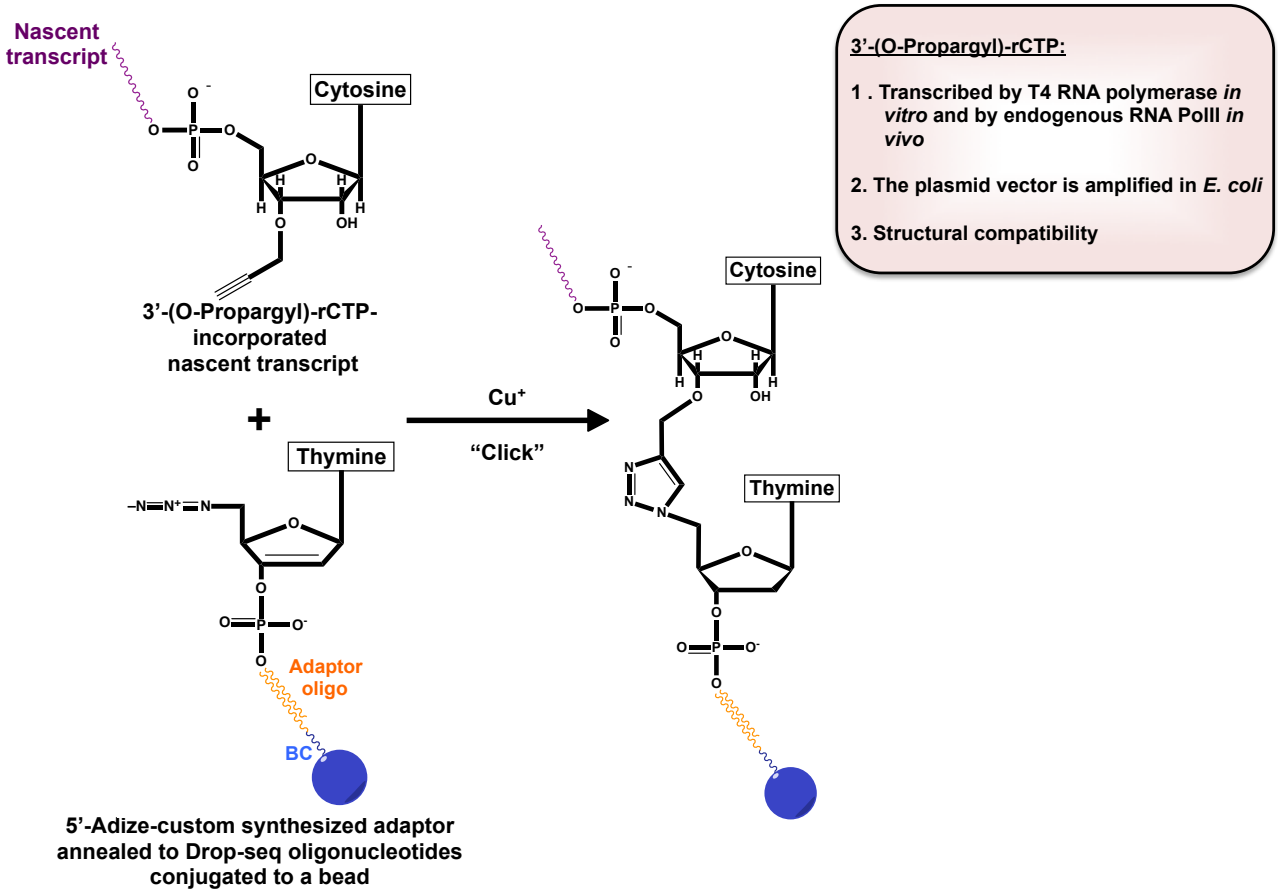


Figure 5.3. Schematics of CuAAC reaction between 3'-(O-propargyl)-CTP and 5' azide of the adapter oligo.
3'-(O-propargyl)-CTP is incorporated in nascent transcripts during nuclear run-on. The adapter oligo with 5' azide modification is annealed to the Drop-seq oligo in a separate tube. The 3'-(O-propargyl)-CTP-incorporated nascent transcript is captured to the adapter oligo through a copper-catalyzed click reaction between the alkyne group in the CTP analog and the azide group on the oligo. The resulting DNA/RNA hybrid contains a triazole ring.

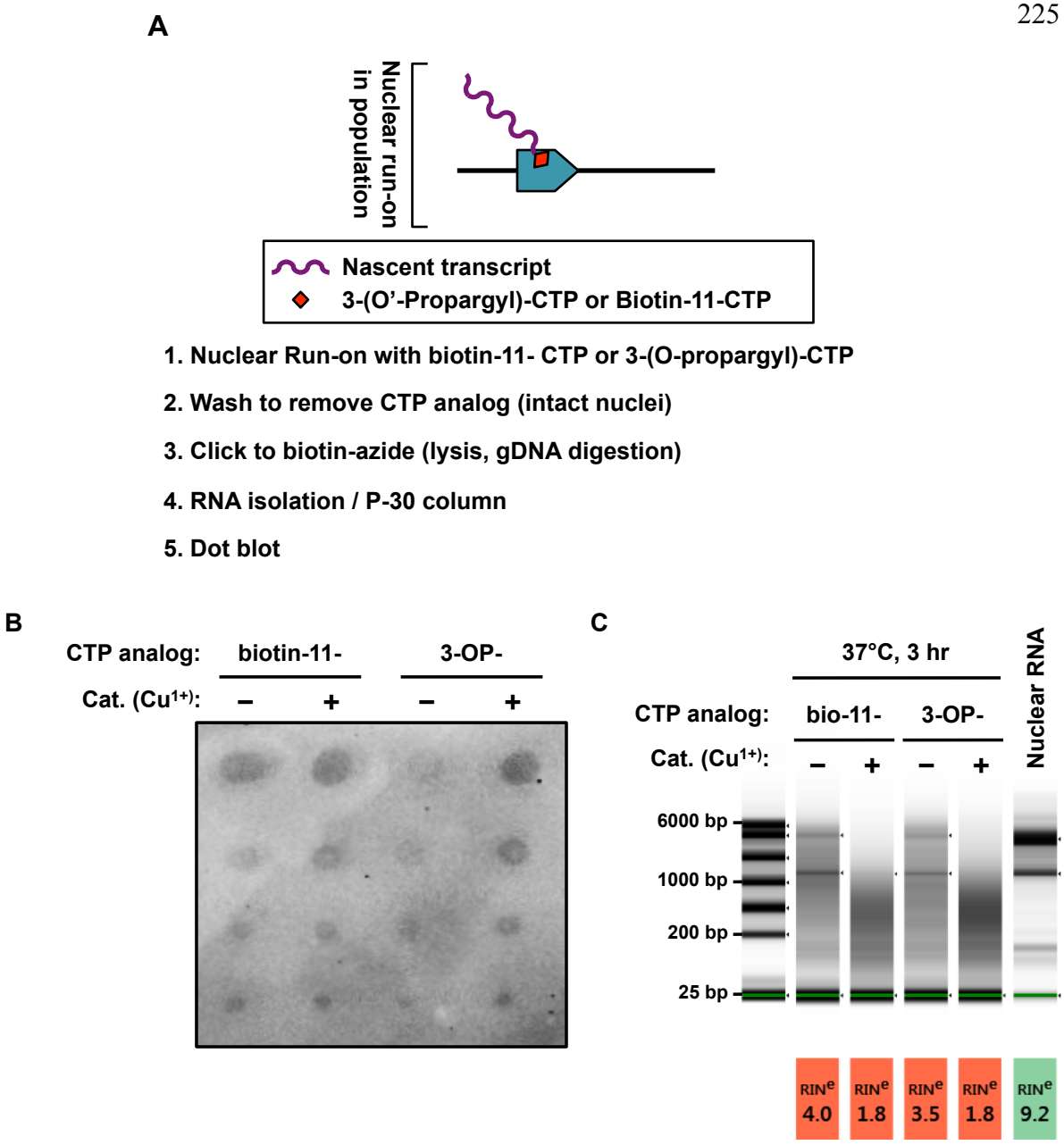
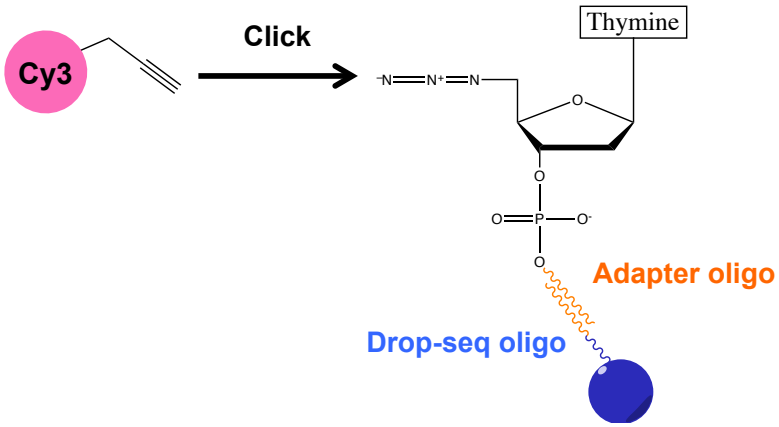


Figure 5.4. Nuclear run-on incorporates 3'-(O-propargyl)-CTP in nascent RNA.
(A) A schematic and steps of the experiment to test incorporation of 3'-(O-propargyl)-CTP in nascent transcripts during nuclear run-on.
(B) A dot blot showing the biotin incorporation levels in each test condition. 3'-(O-propargyl)-CTP or biotin-CTP incorporated in nascent RNA was clicked to biotin-azide, and the biotin levels were compared by a dot blot.
(C) The RNA integrity was determined using an RNA Tape on TapeStation. RAN was degraded through incubation for the click reaction. The presence of copper catalysis further enhanced the RNA degradation.

Figure 5.5. The 5' azide-containing adapter oligo captures alkyne-containing molecules on beads through a CuAAC reaction.
(A) A schematic and steps of the experiment to test if the 5 azide modified adapter oligo captures Cy3-alkyne thought a CuAAC reaction while maintaining annealing to the Drop-seq oligo.
(B) Fluorescent intensities form Cy3 using titrating amount of adapter oligos with or without 5' azide modification. Off beads, Cy3 released from the beads; on beads, Cy3 background on the beads after release of Cy3.

[Figure 5.5 is on the next page]

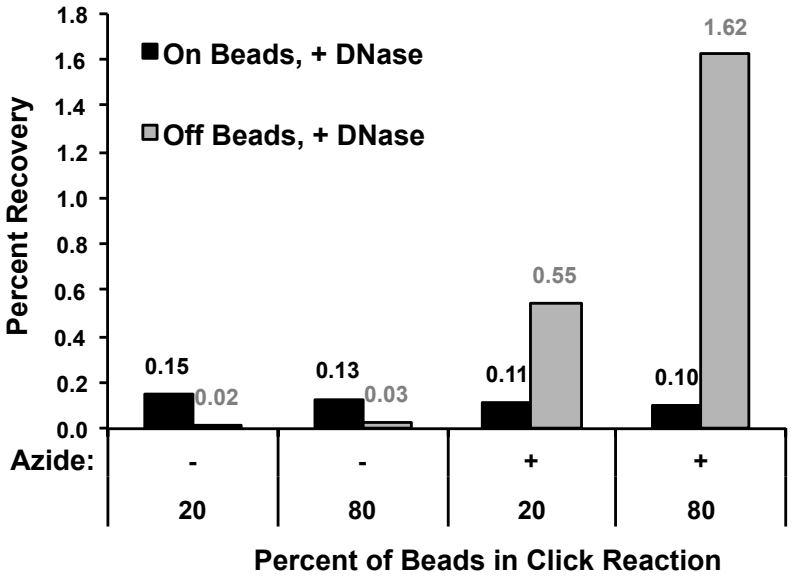
A



1. Anneal Azide-adaptor oligo to biotin-Drop-seq oligo
2. Immobilize annealed Oligo to streptavidin beads
3. Click Cy3-alkyne
4. Wash beads
5. Release Cy3 by DNase treatment
6. Collect the beads and take the elute

   Supe = "Off Beads"
7. Wash beads
8. Resuspend the beads = "On Beads"

B



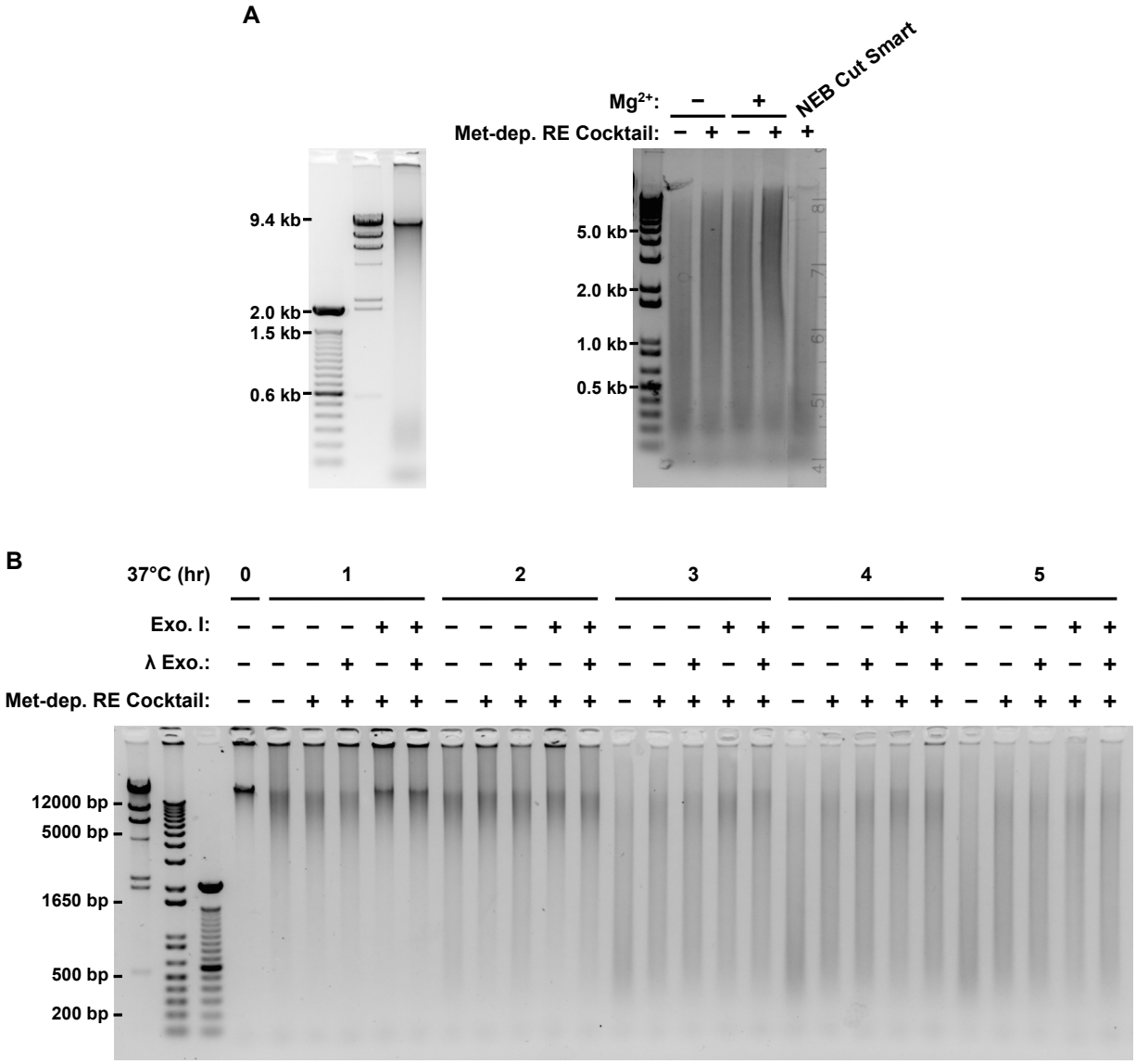


Figure 5.6. Copper-mediated genomic DNA degradation occurs during CuAAC reaction without nucleases.
(A) An agarose gel showing undigested gDNA (*left*) and gDNA digested by methylation-dependent restriction enzyme (Met-dep. RE cocktail) in the presence or absence of Mg^{2+} (*right*).
(B) An agarose gel showing time-dependent gDNA digestion in the presence of indicated combination of nucleases. Copper-mediated gDNA digestion with 3 hr incubation sufficiently degrades gDNA.

CHAPTER SIX

Summary and Future Directions

6.1 Summary

Transcription is a fundamental biological process involving the regulation of a wide array of physiological and pathological activities in the cell. DNA regulatory elements called enhancers provide a focal point for transcription factor binding, which in turn triggers assembly of enhancer complexes and enhancer activation. This entire process consists of layers of molecular events involving numbers of transcription coregulators.

In my dissertation studies, I explored selected aspects on enhancer activation using ligand-dependent, DNA-binding nuclear receptor ERα as a model transcription factor. Upon ligand and chromatin binding, ERα recruits numbers of coregulators including SRCs, Mediator, p300, and BRDs to form an enhancer complex. Using genomic, molecular, and cellular assays, my studies provide insights on the dynamic and cooperative nature of ERα enhancer complex formation. In particular, p300 and Mediator are recruited to ERα cooperatively independent of SRCs in the initial phase of enhancer activation, which is sufficient to evoke the transcriptional response. However, ERα enhancer requires SRCs and SRC-mediated maintenance of p300 recruitment to maintain transcription outputs. In another words, p300 recruitment in any mechanisms results in transcription activation at enhancer as well as the target gene, indicating a central role of p300/CBP in the enhancer complex. As active enhancers nucleated by a wide range of TFs share many common features, dynamic regulation and assembly such

as coregulator switching as described herein may be a common underlining mechanism for enhancer complex formation by many TFs.

Successful enhancer complex formation results in transcription activation. Using enhancer transcripts, we characterized genome-wide ERα binding sites and their enhancer activities. Interestingly, only as much as half of ERα binding events resulted in the production of enhancer transcription in our study. ERα binding sites accompanied with eRNA production correlated with other active enhancer features such as enrichment of coregulator including BRDs and p300/CBP and enhancer-enriched histone marks. Collectively, my dissertation studies indicate that (1) TF binding alone is not sufficient for enhancer activation, (2) enhancer complex formation involving coregulators occurs in a dynamic and cooperative manner, and (3) the production of enhancer transcription reside as the end-point of enhancer activation. Altogether, these aforementioned studies deliver better understanding of the mechanisms and functions of transcription enhancers.

6.2 Future Directions

6.2.1 The Roles and Mechanisms of Enhancer Features

As presented in this dissertation studies, enhancer formation and activation is a dynamic process in which numbers of coregulators influence each other. Although the recent work by us and others have dissected some of the key events in dynamic enhancer formation, it requires ample amount of additional studies to comprehend the full picture of enhancer biology. Central questions yet to be answered include differential functions of each coregulators in enhancer complexes and the functional relationships among coregulators. In parallel, it is also important to determine minimal requirement for an active enhancer complex. In my dissertation studies, we demonstrated that the recruitment of p300 and Mediator independent of active ERα was sufficient to induce gene expression. What are coregulators beside p300 and Mediator that were recruited by the two factors and which of them are essential for enhancer activation? Is the minimal requirement for enhancer activity for a given TF consistent or variable across biological contexts? Moreover, do enhancer complexes formed by various TFs share a common requirement or they claim TF-specific minimal units?

Development of microarrays and high-throughput sequencing technologies accelerated our understanding on transcription regulations by enhancers. Series of studies in the past two decades have accumulated gnomic observations that correlate TF binding and enhancer activity with active enhancer features such as enhancer-enriched histone marks and extensive chromatin looping to the target genes. However, these

studies are typically descriptive and limited to correlations. Only in recent years, the field has started focusing on molecular mechanisms and functions of the events associated enhancers beyond correlations (Chen et al., 2017; Murakami et al., 2017; Stampfel et al., 2015). Long list of questions yet to be answered suggests a hope for the bright future of the field.

As aforementioned, one of cryptic pieces of enhancer features is the formation of enhancer-promoter chromatin looping. It has been speculated as a mechanism for distal enhancers to modulate the target gene expression; however, direct evidence to support this hypothesis is still underway. In my dissertation studies using 3C-qPCR assays, we demonstrated that enhancer-promoter chromatin looping is not sufficient for transcription activation by two unrelated experiments (in MCF-7 cells treated with Flavopiridol where enhancer-promoter chromatin looping is maintained in absence of transcription; MDA-MB-231 cells expressing ERα LQ where E2-dependent transcription is impaired while E2 induces enhancer-promoter chromatin interaction is maintained). In addition, these results align with other studies using different biological systems where enhancer-promoter chromatin looping is formed without transcription activation (Krivega et al., 2014; Schaukowitch et al., 2014; Therizols et al., 2014). Nevertheless, it is still unclear if enhancer-promoter chromatin looping is necessary for transcription. Further characterization of the molecular mechanisms on chromatin looping formation will aid our understanding for its role in gene regulation.

6.2.2 The Molecular Basis of Enhancer Activity and Transcriptional Outcomes in Single-cell

The accumulated observations based on the conventional genomic experiments (e.g. ChIP-on-chip, Next Generation Sequencing) revealed numbers of enhancer features, including enhancer-enriched histone modifications, coregulator recruitment, and chromatin looping. One caveat of the conventional genomic techniques using a population of cells is a lack of evidence for the occurrence of enhancer activation and the target gene expression in the same cells. If any of the enhancer features are required for transcription of the target genes, one should observe both events in the same cell.

For long time, the drawback of the population-based genomic techniques had been complemented with studies using imaging technologies (Chen and Larson, 2016; Shachar et al., 2015; Symmons and Raj, 2016). Imaging-based studies early on showed that signal-dependent transcription activation is intrinsically stochastic events at the single-cell level (Kaern et al., 2005; Raj and van Oudenaarden, 2008). Moreover, further studies using single-cell imaging technologies suggested discrepancies between enhancer features and the target gene transcription (Fukaya et al., 2016; Rahman et al., 2017). For instance, Fukaya *et al.* showed that a given enhancer shared by two reporter genes turns on the reporter genes simultaneously rather than stochastically choosing one of the two reporters at a time. Their results contradict with the paradigm in which the promoter-enhancer interaction determines the enhancer-target gene specificity. In addition, Rahman *et al*. reported their observations suggesting disconnect between active enhancer features and the transcriptional outcomes. They determined the number of loci that

express enhancer transcripts (eRNAs) as well as the target genes in polyploid cells using fluorescence in situ hybridization (FIHS). Although the number of alleles that express eRNA and the target gene mRNA correlate across experimental conditions, these evens are observed predominantly in separate alleles. Furthermore, they show that when eRNA and mRNA are produced from two different alleles, the distance between the two alleles are greater than what would be observed when they are in close proximity through chromatin looping. Like the study by Fukaya *et al.*, they suggested an un-linked relationship between eRNA and mRNA transcription and the promoter-enhancer chromatin looping.

While imaging technology is effective, it is not as suitable to observe thousands of transcripts in a single experiment as next-generation sequencing. Also detection of eRNAs by FISH may inherit some limitations because of the lack of systematic annotation of eRNAs as well as their unstable nature as transcripts. Single-cell GRO-seq (scGRO-seq) promises an alternative highly sensitive, dynamic, and high-throughput approach that overcome these issues. By capturing enhancer transcription as a surrogate for enhancer activity along with target gene transcription, scGRO-seq will help us to uncover new principles in molecular mechanisms and functions of transcription enhancer at the single-cell level.

6.3 References

Chen, F.X., Xie, P., Collings, C.K., Cao, K., Aoi, Y., Marshall, S.A., Rendleman, E.J., Ugarenko, M., Ozark, P.A., Zhang, A., et al. (2017). PAF1 regulation of promoter-proximal pause release via enhancer activation. Science 357, 1294-1298.

Chen, H., and Larson, D.R. (2016). What have single-molecule studies taught us about gene expression? Genes Dev 30, 1796-1810.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell 166, 358-368.

Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet 6, 451-464.

Krivega, I., Dale, R.K., and Dean, A. (2014). Role of LDB1 in the transition from chromatin looping to transcription activation. Genes Dev 28, 1278-1290.

Murakami, S., Nagari, A., and Kraus, W.L. (2017). Dynamic assembly and activation of estrogen receptor alpha enhancers through coregulator switching. Genes Dev 31, 1535-1548.

Rahman, S., Zorca, C.E., Traboulsi, T., Noutahi, E., Krause, M.R., Mader, S., and Zenklusen, D. (2017). Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. Nucleic Acids Res 45, 3017-3030.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135, 216-226.

Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell 56, 29-42.

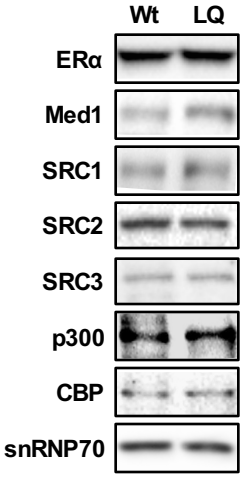
Shachar, S., Voss, T.C., Pegoraro, G., Sciascia, N., and Misteli, T. (2015). Identification of Gene Positioning Factors Using High-Throughput Imaging Mapping. Cell 162, 911-923.

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature 528, 147-151.

Symmons, O., and Raj, A. (2016). What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. Mol Cell 62, 788-802.

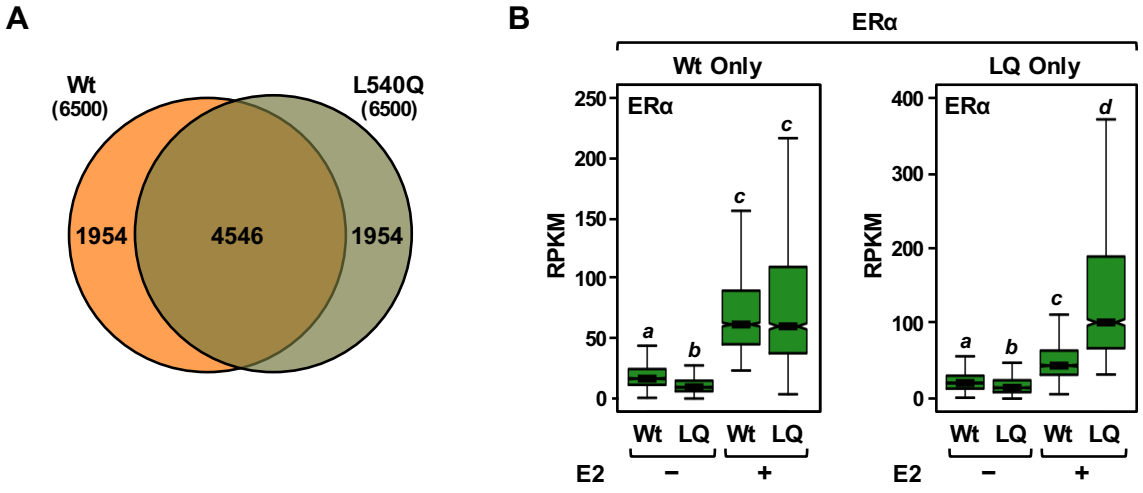
Therizols, P., Illingworth, R.S., Courilleau, C., Boyle, S., Wood, A.J., and Bickmore, W.A. (2014). Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. Science 346, 1238-1242.

APPENDIX S1
Supplemental Figures Related To Chapter 3



Appendix S1. ERα coregulators of interest in this study are expressed at similar levels in 231/ERαWt and 231/ERαLQ cells.
Western blots showing the expression levels of ERα and its coregulators in 231/ERαWt and 231/ERαLQ cells.

APPENDIX S2
Supplemental Figures Related To Chapter 3

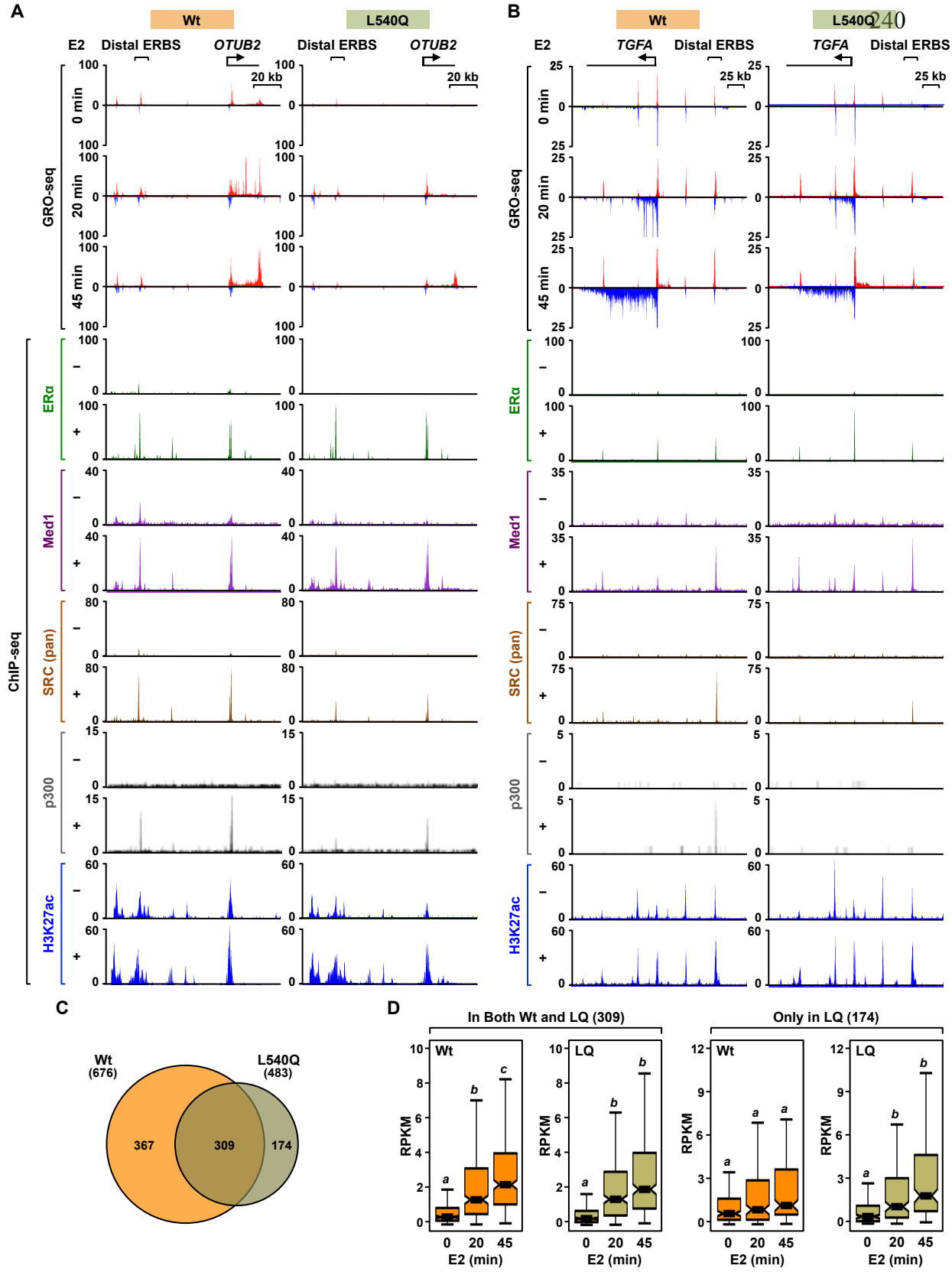


Appendix S2. Considerable overlap between ERαWt and ERαLQ binding sites in 231/ERα cells.
(A) A large subset of ERαWt and ERαLQ binding sites overlap in 231/ERα cells. Venn diagram showing the overall of significant ERαWt and ERαLQ peaks upon 45 min. of E2 treatment in 231/ERαWt and 231/ERαLQ cells, as determined by ChIP-seq, for the top 6,500 peaks.
(B) Loci with significant peaks of ERαWt in 231/ERαWt cells are enriched in read counts for ERαLQ in 231/ERαLQ, and vice versa. Box plots of ERα ChIP-seq read counts ± 45 min. of E2 treatment at ERα binding sites unique to either ERαWt (Wt only) or ERαLQ (LQ only) after 45 min. of E2 treatment in 231/ERαWt and 231/ERαLQ cells. Box plots marked with different letters (*a*, *b*, *c, d*) are significantly different from each other ($p < 2.2 \times 10^{-16}$; Wilcoxon rank-sum test).

APPENDIX S3
Supplemental Figures Related To Chapter 3

Appendix S3. Reduced SRC recruitment at ERα L540Q binding sites impairs enhancer formation and target gene transcription *(continued)*.
(A and B) Impaired recruitment of SRC and p300 at ERα L540Q binding sites results in reduced E2-stimulated transcription. GRO-seq browser tracks with a time course of E2 treatment (0, 20, and 45 min.) and ChIP-seq browser tracks for ERα, Med1, SRC (pan), p300, and H3K27ac ± 45 min. E2 at the *OTUB2* (A) and *TGFA* (B) loci in MBA-MB-231 cells expressing ERα wild-type or ERα L540Q (231/ERαWt and 231/ERαLQ cells, respectively).
(C) ERα L540Q exhibits impaired transcriptional activity. Venn diagram showing the number of genes significantly up-regulated by 45 min. of E2 treatment in 231/ERαWt and 231/ERαLQ cells as measured by GRO-seq. FDR = 0.05.
(D) Box plots showing the read counts for 309 genes up-regulated by both wild-type and L540Q ERα (*left*) or 174 genes up-regulated by the L540Q mutant, but not by the wild-type, (*right*) upon E2 treatment, as shown in panel (C). Box Plots marked with different letters (*a, b, c*) are significantly different from each other ($p < 1.54 \times 10^{-05}$; Wilcoxon rank-sum test).

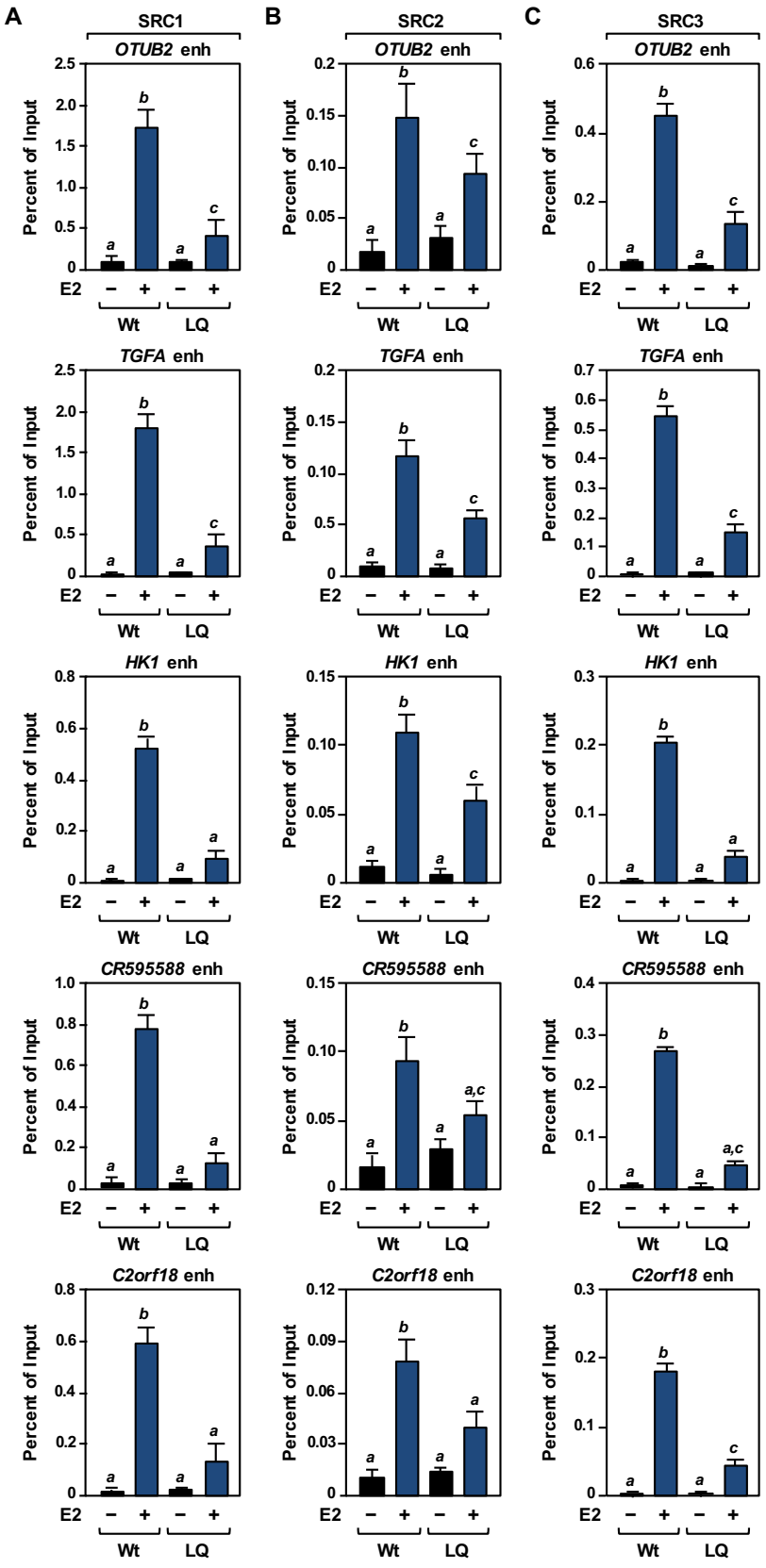
[Appendix S3 is on the next page]


A
Wt
L540Q
E2
Distal ERBS
OTUB2
20 kb
GRO-seq
0 min
20 min
45 min
ChIP-seq
ERα
Med1
SRC (pan)
p300
H3K27ac
B
TGFA
25 kb
C
Wt
(676)
L540Q
(483)
367
309
174
D
In Both Wt and LQ (309)
Only in LQ (174)
LQ
RPKM
E2 (min)

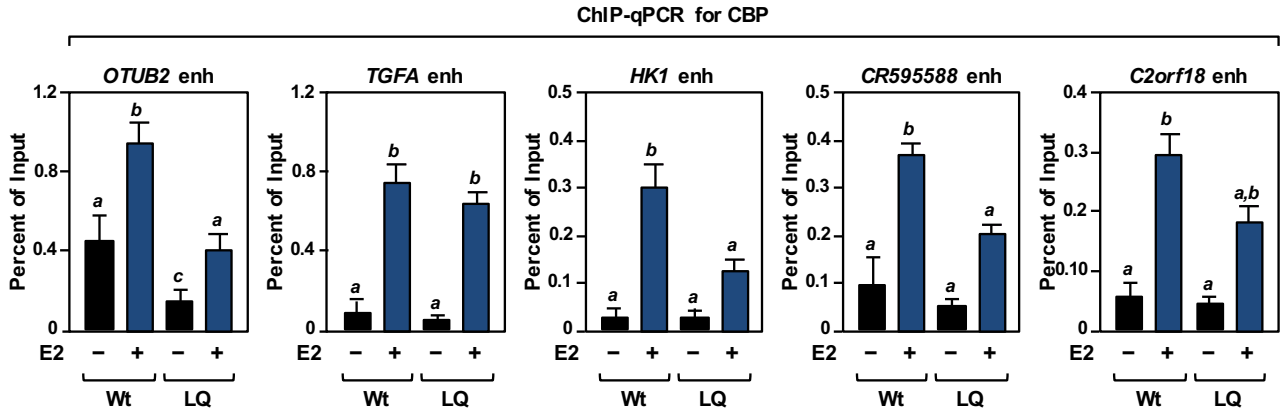
APPENDIX S4
Supplemental Figures Related To Chapter 3

Appendix S4. The recruitment of all three SRC family members is impaired at ERα L540Q binding sites.
(A-C) ChIP-qPCR assays for (A) SRC1, (B) SRC2, and (C) SRC3 in 231/ERαWt and 231/ERαLQ cells treated ± E2 for 45 min. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Bars marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 0.05$; two-way ANOVA).

[Appendix S4 is on the next page]

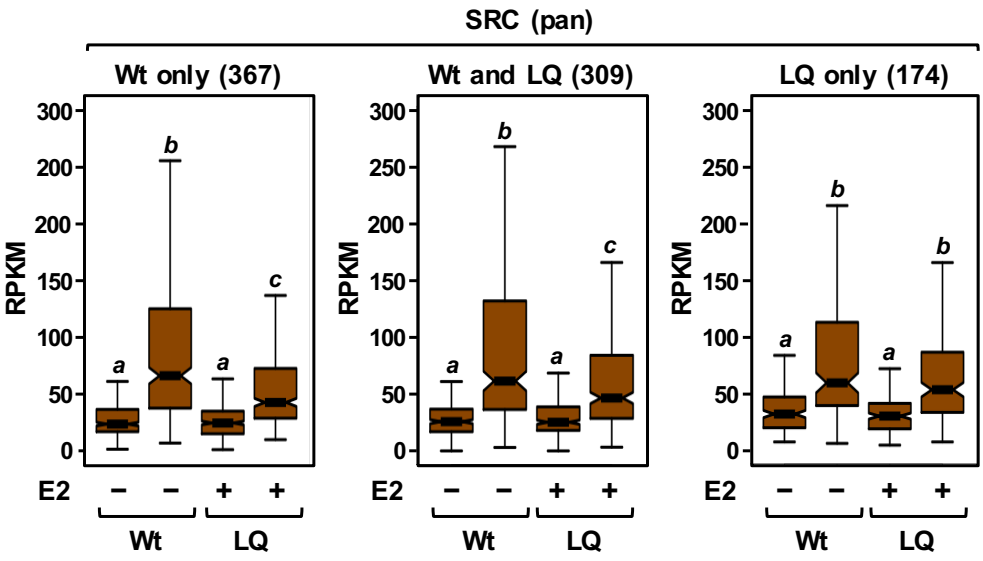

A
B
C
SRC1
SRC2
SRC3
OTUB2 enh
TGFA enh
HK1 enh
CR595588 enh
C2orf18 enh
Percent of Input
E2
Wt
LQ

APPENDIX S5
Supplemental Figures Related To Chapter 3



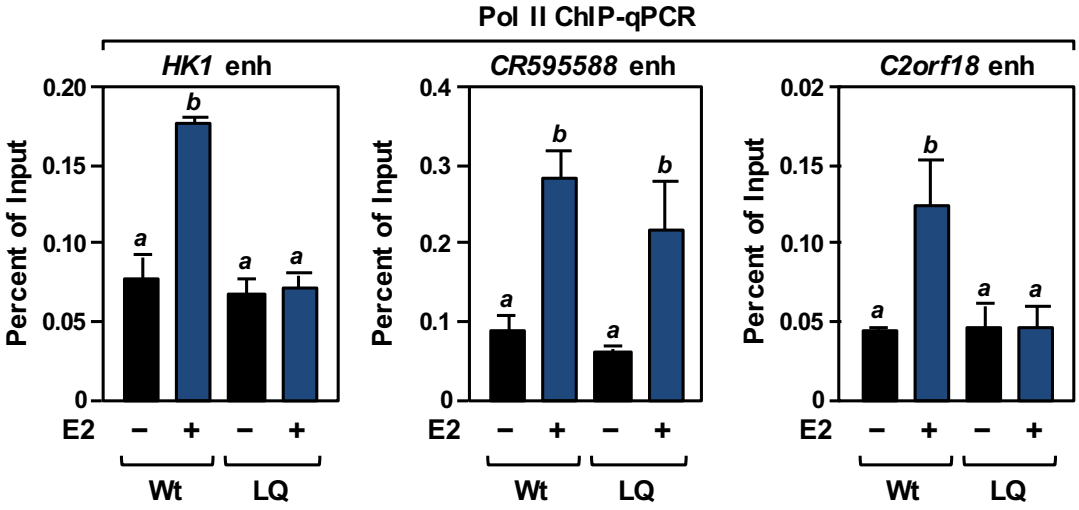
Appendix S5. Recruitment of CBP is impaired at ERα L540Q binding sites.
ChIP-qPCR assays for CBP in 231/ERαWt and 231/ERαLQ cells treated ± E2 for 45 min. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Bars marked with different letters (*a*, *b*, c) are significantly different from each other ($p < 0.05$; two-way ANOVA).

APPENDIX S6
Supplemental Figures Related To Chapter 3



Appendix S6. Recruitment of SRC to various types of ERα L540Q genomic binding sites.
Box plots of SRC (pan) ChIP-seq read counts ± 45 min. of E2 treatment at ERα binding sites located nearby genes (1) uniquely up-regulated by ERαWt (*left*), (2) commonly up-regulated by ERαWt and ERαLQ (*middle*), or (3) uniquely up-regulated by ERαLQ (*right*). Box plots marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 2.2 \times 10^{-5}$; Wilcoxon rank-sum test).

APPENDIX S7
Supplemental Figures Related To Chapter 3



Appendix S7. Pol II recruitment is impaired at ERα L540Q binding sites.
ChIP-qPCR assays for Pol II in 231/ERαWt and 231/ERαLQ cells treated ± E2 for 45 min. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Bars marked with different letters (*a*, *b*) are significantly different from each other ($p < 0.05$; two-way ANOVA).

APPENDIX S8
Supplemental Figures Related To Chapter 3

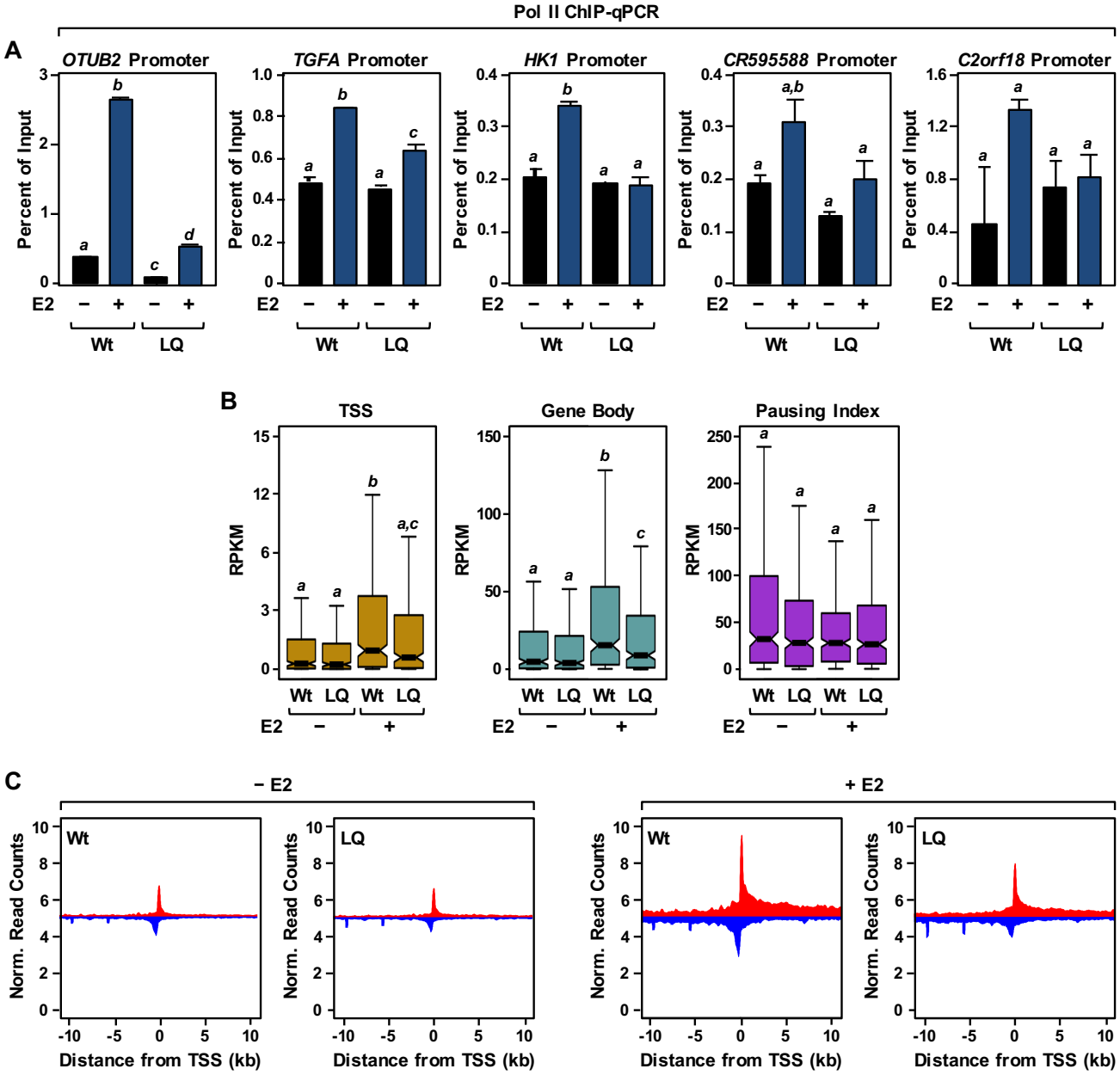
Appendix S8. Impaired E2-dependent Pol II loading and transcription initiation with ERα L540Q.

(A) Pol II recruitment is impaired at the promoters of ERα L540Q target genes. ChIP-qPCR assays for Pol II at the promoters of the indicated genes in 231/ERαWt and 231/ERαLQ cells treated ± E2 for 45 min. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Bars marked with different letters (*a*, *b*, *c*, *d*) are significantly different from each other ($p < 0.05$; two-way ANOVA).

(B) ERαLQ shows impaired initiation for E2-dependent gene activation, which ultimately affects elongation. (*Left and middle*) Box plots showing the read counts for the transcription start site (TSS) (*left*) and the gene body (*middle*) for 367 genes up-regulated by ERαWt, but not by ERαLQ, upon E2 treatment as shown in Figure 3A. (*Right*) Box plots showing the pausing indexes for the same set of genes. Box plots marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 2.2 \times 10^{-03}$; Wilcoxon rank-sum test).

(C) Metaplots of GRO-seq data ± 10 kb around the TSS of the same set of genes described in (B).

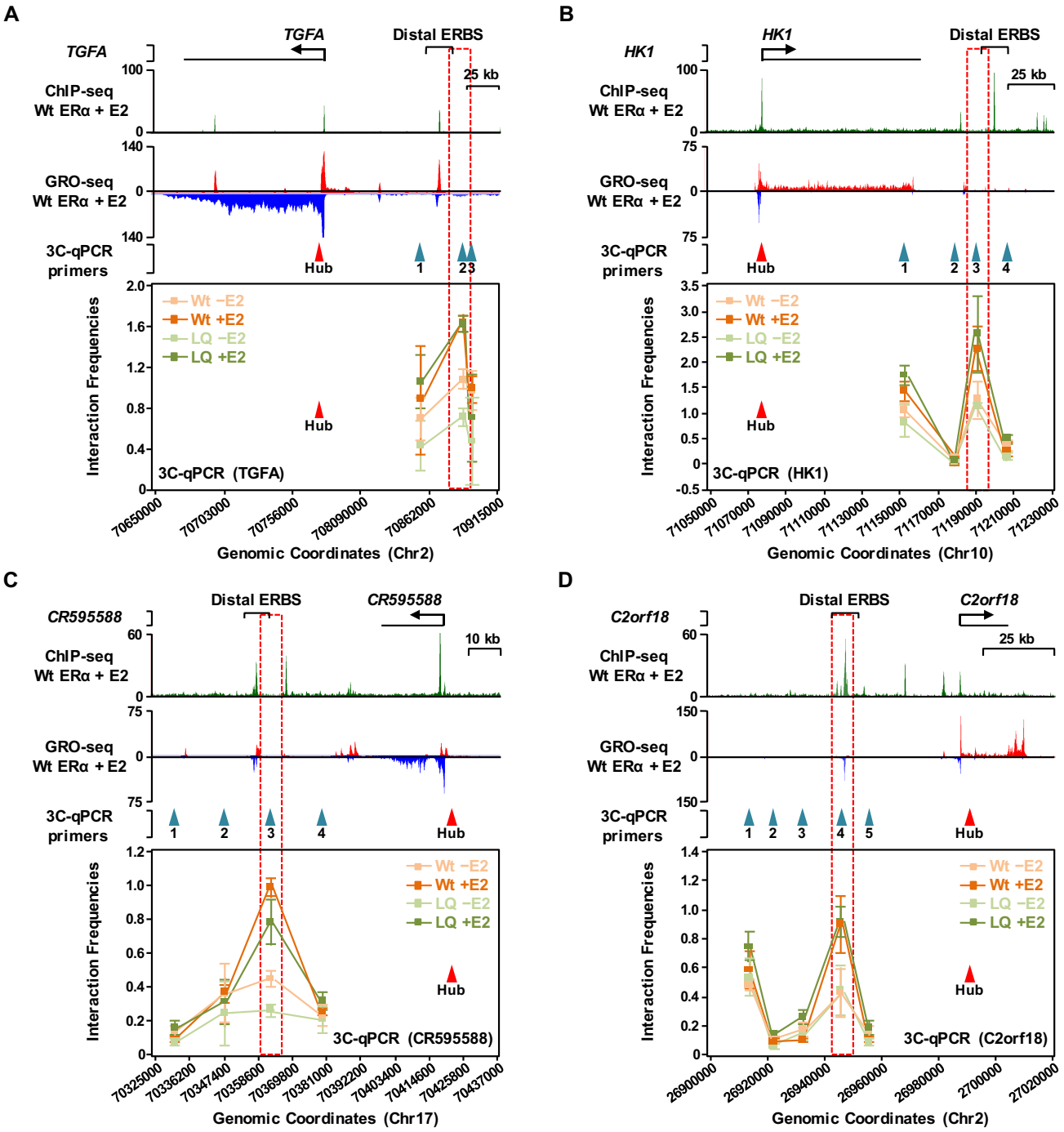
[Appendix S8 is on the next page]


Pol II ChIP-qPCR
A
OTUB2 Promoter
TGFA Promoter
HK1 Promoter
CR595588 Promoter
C2orf18 Promoter
Percent of Input
E2
Wt
LQ
B
TSS
Gene Body
Pausing Index
RPKM
C
– E2
+ E2
Wt
LQ
Norm. Read Counts
Distance from TSS (kb)

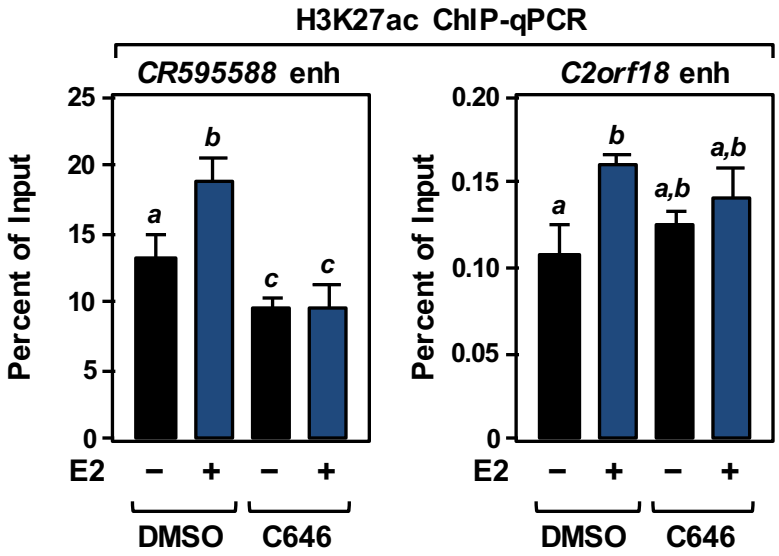
APPENDIX S9
Supplemental Figures Related To Chapter 3

Appendix S9. E2-dependent enhancer-promoter chromatin loop formation is maintained with ERα L540Q, in spite of impaired SRC recruitment.
(Top) Browser tracks for ERα ChIP-seq and GRO-seq in 231/ERαWt cells after 45 min. of E2 treatment, shown with the location of the 3C primers. *(Bottom)* 3C-qPCR assays showing chromatin looping from a distal ERα binding site to the (A) *TGFA*, (B) *HK1*, (C) *CR595588*, or (D) *C2orf18* promoters in 231/ERαWt and 231/ERαLQ cells ± 45 min. of E2 treatment. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to Wt at eac h genomic location in each condition ($p < 0.05$; two-way ANOVA).

[Appendix S9 is on the next page]


A
TGFA
Distal ERBS
ChIP-seq
Wt ERα + E2
GRO-seq
Wt ERα + E2
3C-qPCR
primers
Hub
25 kb
Wt −E2
Wt +E2
LQ −E2
LQ +E2
Interaction Frequencies
3C-qPCR (TGFA)
Genomic Coordinates (Chr2)
B
HK1
Distal ERBS
3C-qPCR (HK1)
Genomic Coordinates (Chr10)
C
CR595588
Distal ERBS
10 kb
3C-qPCR (CR595588)
Genomic Coordinates (Chr17)
D
C2orf18
Distal ERBS
3C-qPCR (C2orf18)
Genomic Coordinates (Chr2)

APPENDIX S10
Supplemental Figures Related To Chapter 3

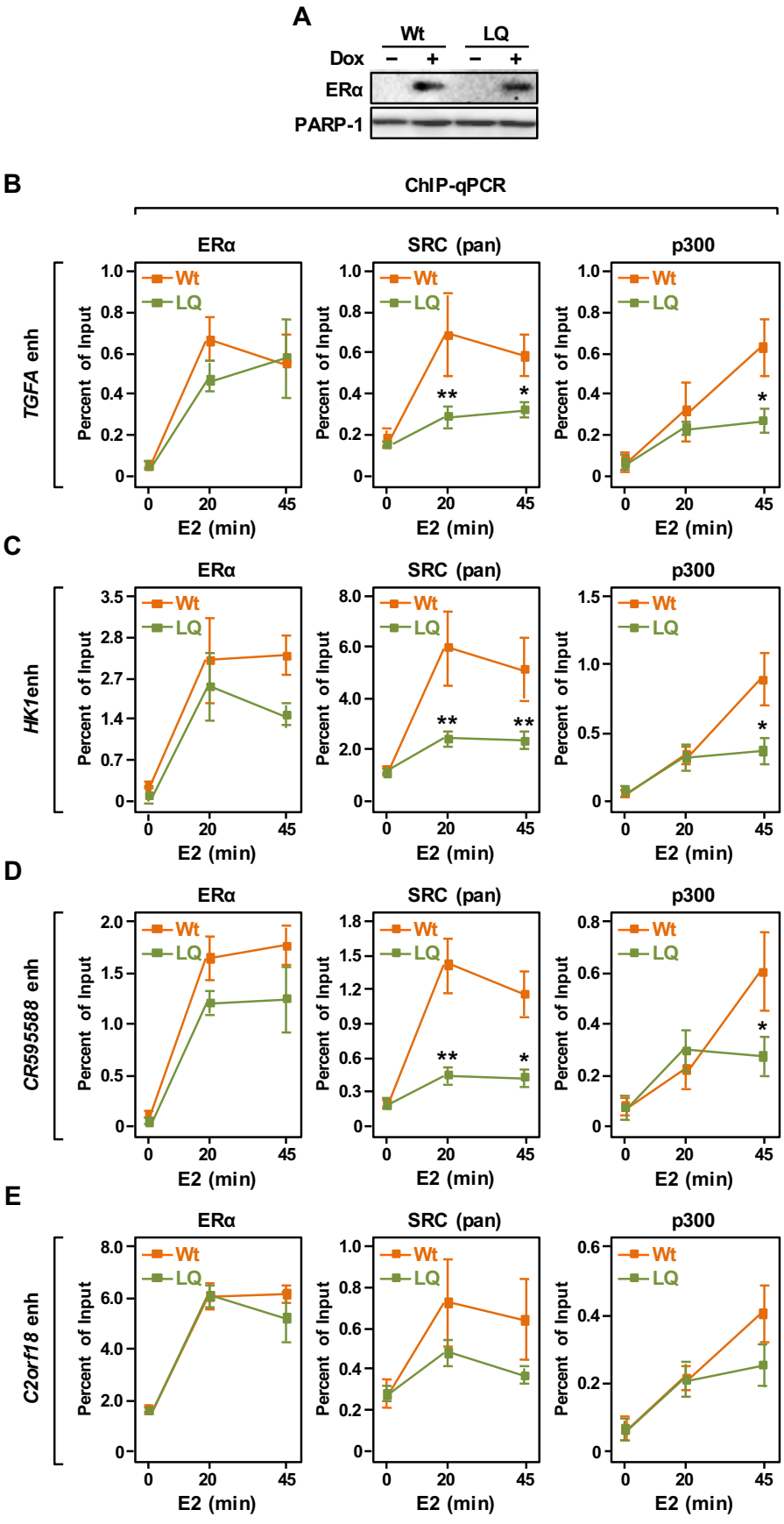


Appendix S10. p300 HAT activity is required for E2-stimulated increases in H3K27ac levels.
ChIP-qPCR assays for H3K27ac in 231/ERαWt cells treated ± E2 for 45 min. in the presence of the p300/CBP HAT inhibitor C646. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Bars marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 0.05$; two-way ANOVA).

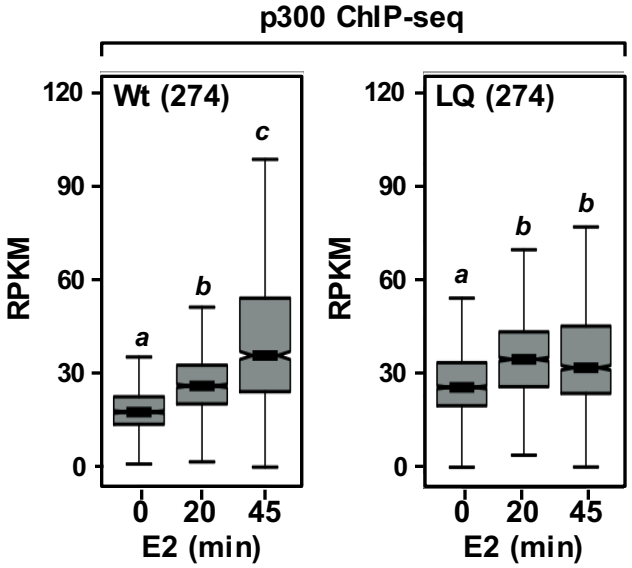
APPENDIX S11
Supplemental Figures Related To Chapter 3

Appendix S11. Impaired SRC recruitment with ERα L540Q causes abortive enhancer formation and target gene transcription *(continued)*.
(A) Western blot showing doxycycline (Dox)-dependent expression of ERα wild-type (Wt) and ERα L540Q (LQ) in MDA-MB-231 cells (231/ERαWt and 231/ERαLQ cells, respectively).
(B - E) p300 is recruited in an SRC-independent manner during the initial phase of enhancer formation ("enhancer priming"). ChIP-qPCR assays for ERα, SRC (pan), and p300 in 231/ERαWt and 231/ERαLQ cells treated with a time course of E2. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to Wt at each time point (*, $p < 0.05$; **, $p < 0.0005$; two-way ANOVA).

[Appendix S11 is on the next page]


A
Wt
LQ
Dox
– + – +
ERα
PARP-1
B
ChIP-qPCR
ERα
SRC (pan)
p300
TGFA enh
HK1enh
CR595588 enh
C2orf18 enh
Wt
LQ
Percent of Input
E2 (min)
0
20
45
**
*

APPENDIX S12
Supplemental Figures Related To Chapter 3

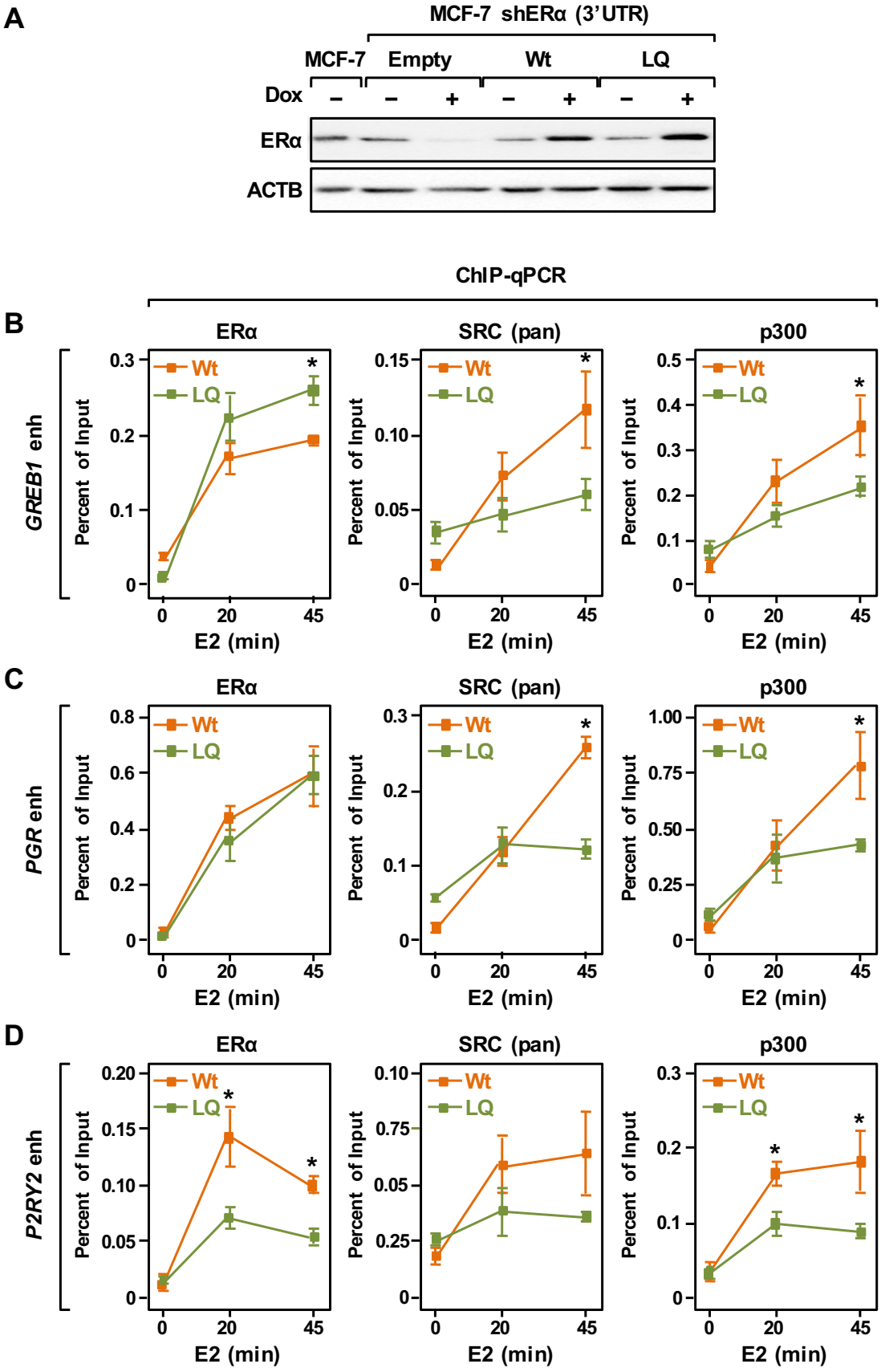


Appendix S12. p300 is recruited to ERαLQ binding sites similarly to ERαWt binding sites during enhancer priming, but not during enhancer maturation.
Box plots of p300 ChIP-seq read counts for a time course of E2 treatment at 274 ERα binding sites common to ERαWt and ERαLQ with >2-fold E2-dependent induction of p300 recruitment at 20 min. Box plots marked with different letters (*a*, *b*, *c*) are significantly different from each other ($p < 2.2 \times 10^{-16}$; Wilcoxon rank-sum test).

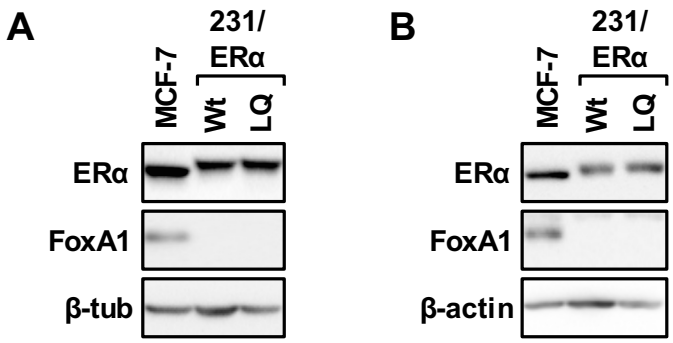
APPENDIX S13
Supplemental Figures Related To Chapter 3

Appendix S13. Impaired SRC recruitment by ERα L540Q results in abortive p300 recruitment in MCF-7 cells.
(A) Western blotting for ERα in MCF-7 cells with Dox-dependent shRNA-mediated knockdown of endogenous ERα in combination with Dox-dependent re-expression of ERαWt or ERαLQ. The control sample with an empty re-expression vector shows the efficiency of knockdown.
(B-D) p300 recruitment at ERαLQ binding sites is induced during enhancer priming, but is attenuated during enhancer maturation in MCF-7 cells. ChIP-qPCR assays for ERα (*left*), SRC (pan) (*middle*), and p300 (*right*) at (B) *GREB1* enhancer, (C) *PGR* enhancer, and (D) *P2RY2* enhancer in the cell lines described in (A) treated with a time course of E2. The percent of input for SRC and p300 enrichment is normalized to the level of ERα binding in the corresponding conditions. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to Wt at each time point (*, $p < 0.05$; two-way ANOVA).

[Appendix S13 is on the next page]


A
MCF-7 shERα (3'UTR)
MCF-7
Empty
Wt
LQ
Dox
– – + – + – +
ERα
ACTB
ChIP-qPCR
B
C
D
ERα
SRC (pan)
p300
GREB1 enh
PGR enh
P2RY2 enh
Percent of Input
E2 (min)
Wt
LQ
0
20
45
*

APPENDIX S14
Supplemental Figures Related To Chapter 3

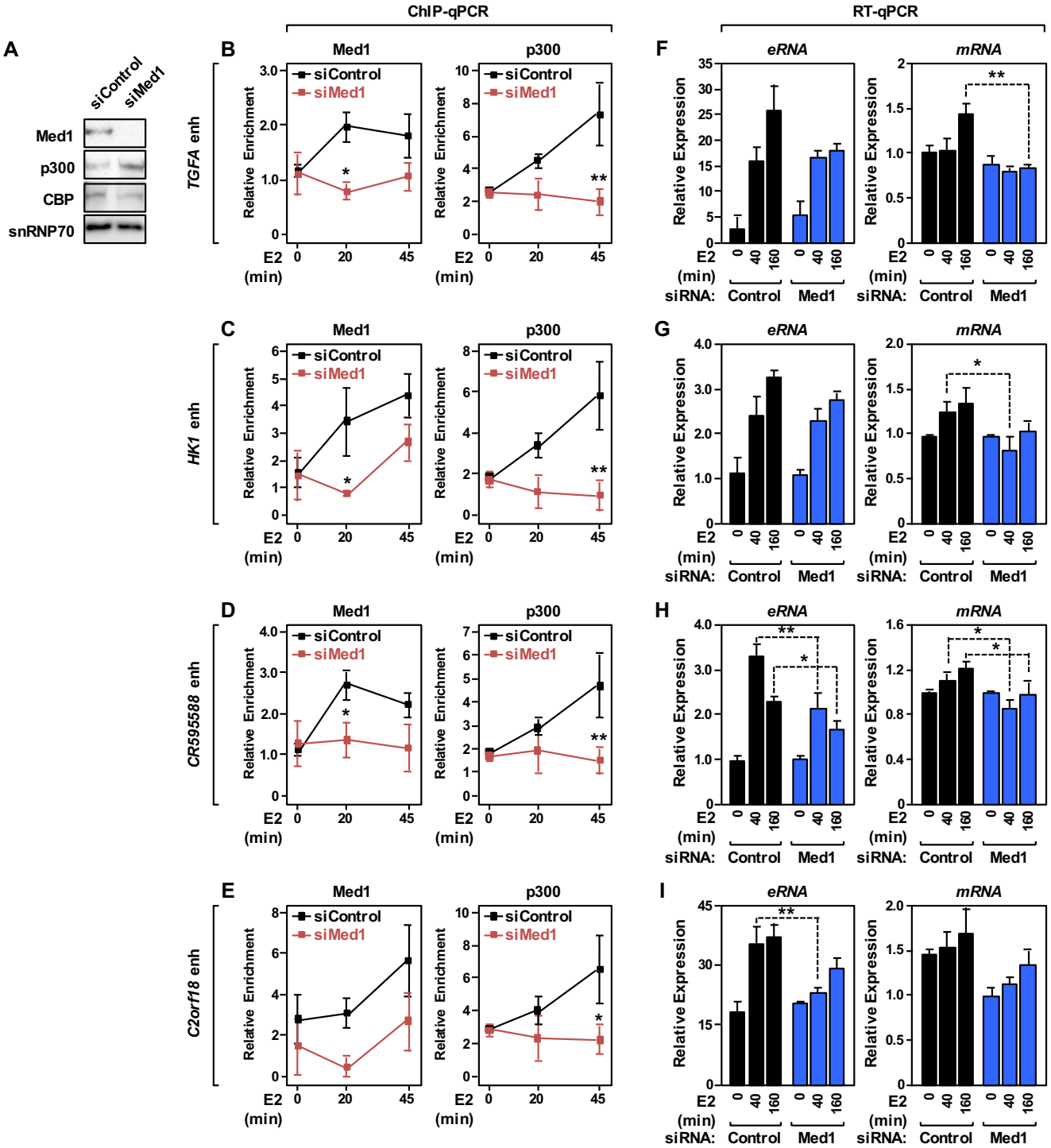


Appendix S14 Differential expression of the pioneer factor FoxA1 in MCF-7, 231/ERαWt, and 231/ERαLQ cells.
(A and B) Western blots showing the expression of ERα and FoxA1 in MCF-7, 231/ERαWt, and 231/ERαLQ cells. (A) 231/ERα cells with constitutive expression of ERαWt or ERαLQ. (B) 231/ERα cells with Dox-inducible expression of ERαWt or ERαLQ.

APPENDIX S15
Supplemental Figures Related To Chapter 3

Appendix S15. SRC-independent ERα enhancer priming requires Mediator *(continued)*.
(A) Western blots showing the expression levels of Med1, p300, and CBP in 231/ERαWt cells with or without siRNA-mediated knockdown of Med1.
(B - E) SRC-independent recruitment of p300 requires Mediator. ChIP-qPCR assays for Med1 and p300 in Med1-depleted doxycycline (Dox)-inducible 231/ERαWt cells treated with a time course of E2. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the siRNA control at each time point (*, $p < 0.005$: **, $p < 0.0001$; two-way ANOVA).
(F - I) Mediator is required for E2-induced gene expression. RT-qPCR assays in Med1-depleted doxycycline (Dox)-inducible 231/ERαWt cells treated with a time course of E2. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the siRNA control at each time point (*, $p < 0.05$; **, $p < 0.005$; two-way ANOVA).

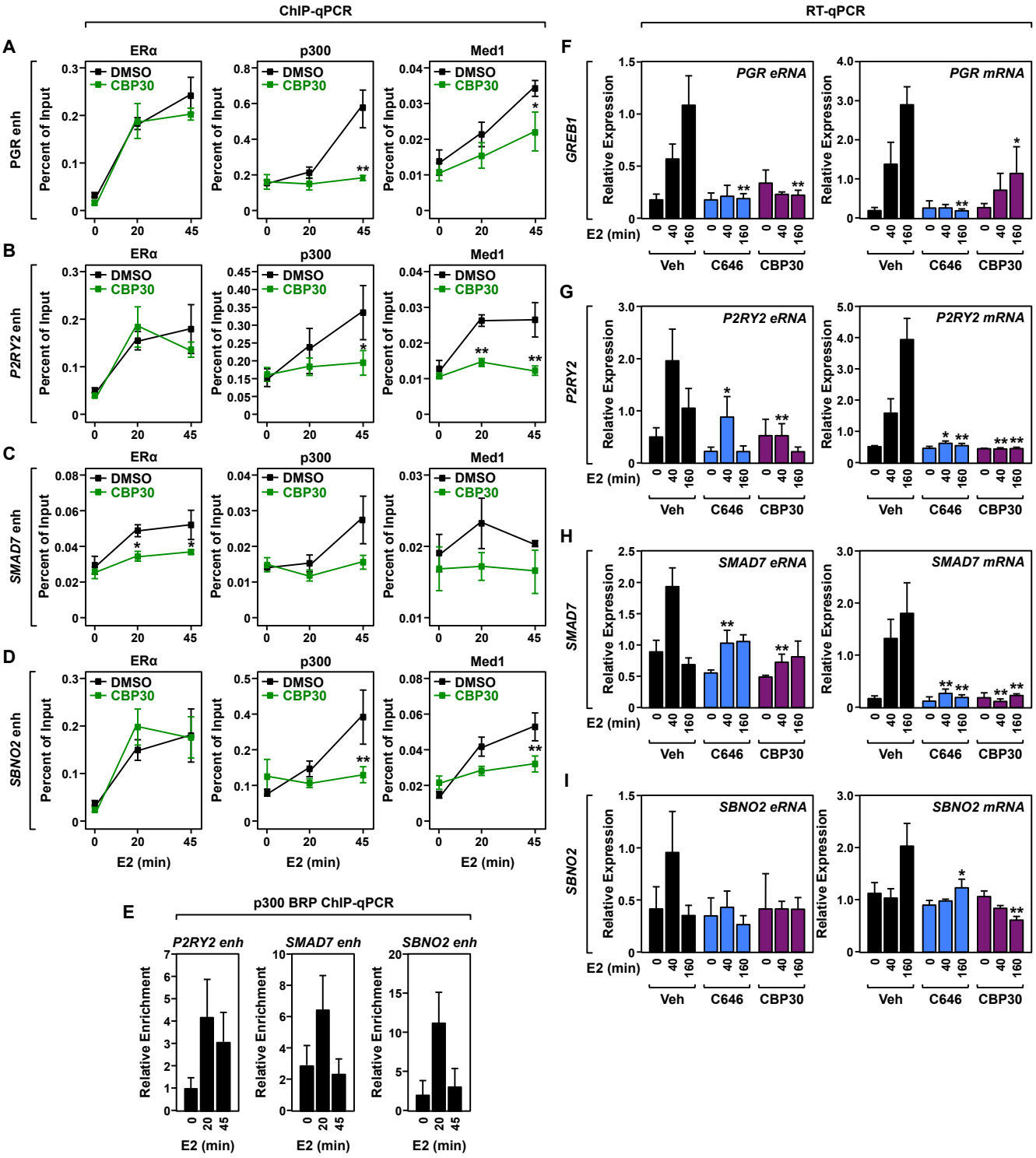
[Appendix S15 is on the next page]


A
siControl
siMed1
Med1
p300
CBP
snRNP70
ChIP-qPCR
RT-qPCR
B
C
D
E
F
G
H
I
Med1
p300
eRNA
mRNA
TGFA enh
HK1 enh
CR595588 enh
C2orf18 enh
siControl
siMed1
Relative Enrichment
Relative Expression
E2
(min)
siRNA:
Control
Med1

APPENDIX S16
Supplemental Figures Related To Chapter 3

Appendix S16. The p300 bromodomain is required for ERα enhancer priming *(continued)*.
(A - D) The p300 bromodomain is required for p300 recruitment during enhancer priming. ChIP-qPCR assays for ERα, p300, and Med1 in MCF-7 cells treated with a time course of E2 in the presence of the p300/CBP bromodomain inhibitor SGC-CBP-30 (CBP30). Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the DMSO control at each time point (*, $p < 0.05$; **, $p < 0.005$; two-way ANOVA).
(E) The p300 bromodomain is recruited to ERα binding sites during enhancer priming. ChIP-qPCR assays for the IgG Fc-fused p300-BRP cassette in MCF-7 cells treated with a time course of E2. Each bar represents the mean + S.E.M. for at least three independent biological replicates.
(F - I) The p300 bromodomain and p300 acetyltransferase activity are required for E2-induced gene expression. RT-qPCR assays in MCF-7 cells treated with a time course of E2 in the presence of the p300 HAT inhibitor C646 or the bromodomain inhibitor SGC-CBP-30 (CBP30). Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the DMSO control at each time point (*, $p < 0.05$; **, $p < 0.005$; two-way ANOVA).

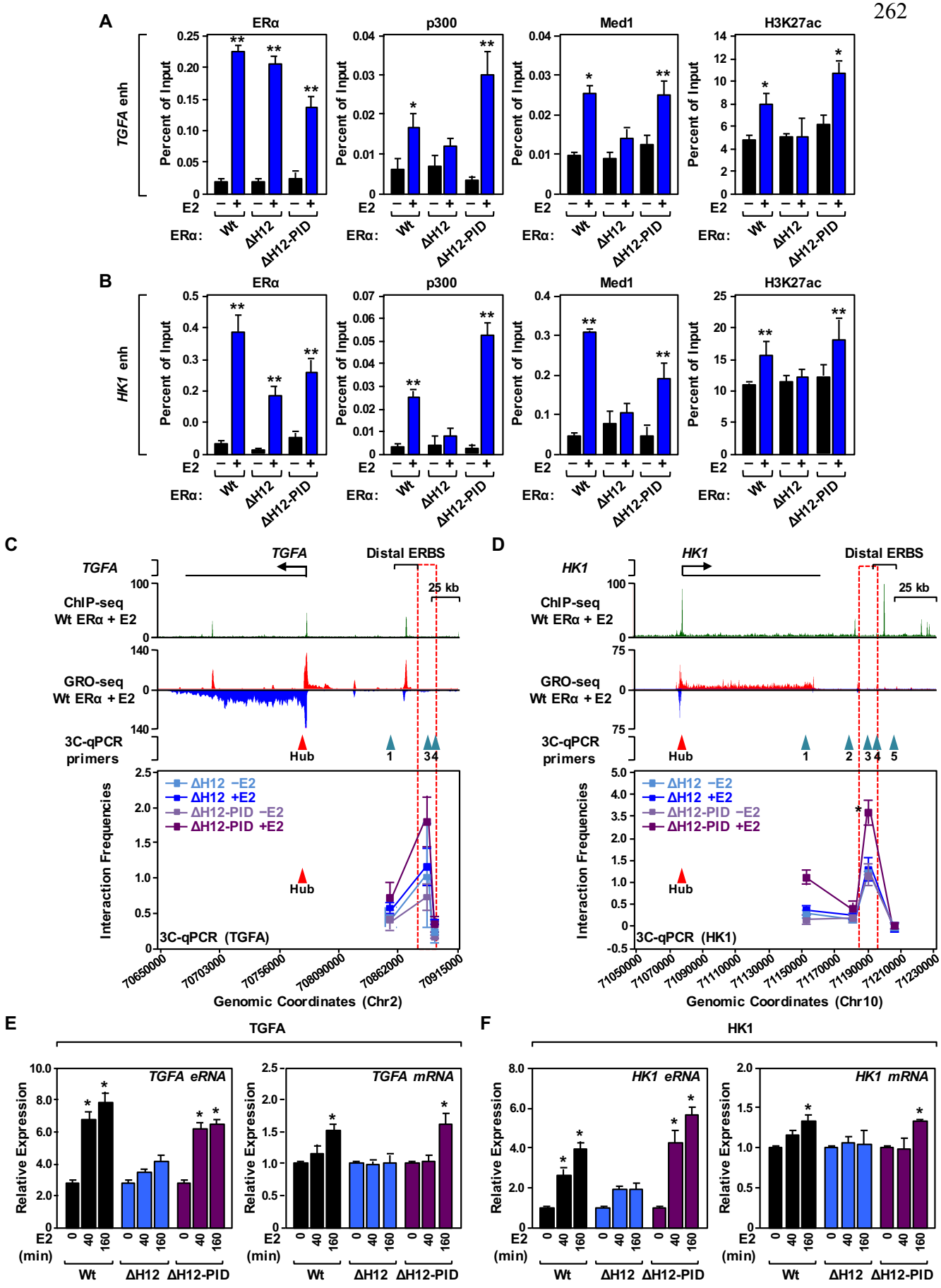
[Appendix S16 is on the next page]


ChIP-qPCR
RT-qPCR
A
B
C
D
ERα
p300
Med1
DMSO
CBP30
Percent of Input
PGR enh
P2RY2 enh
SMAD7 enh
SBNO2 enh
E2 (min)
E
p300 BRP ChIP-qPCR
P2RY2 enh
SMAD7 enh
SBNO2 enh
Relative Enrichment
F
G
H
I
GREB1
P2RY2
SMAD7
SBNO2
PGR eRNA
PGR mRNA
P2RY2 eRNA
P2RY2 mRNA
SMAD7 eRNA
SMAD7 mRNA
SBNO2 eRNA
SBNO2 mRNA
Relative Expression
Veh
C646
CBP30

APPENDIX S17
Supplemental Figures Related To Chapter 3

Appendix S17. Forced recruitment of p300 to ERα binding sites promotes Mediator recruitment, enhancer formation, and target gene expression *(continued)*.
(A and B) Forced recruitment of p300 to an inactive ERα binding site through the SRC2 PID restores Mediator recruitment and H3K27ac enrichment. ChIP-qPCR assays for ERα, p300, Med1, and H3K27ac in the MDA-MB-231 cell lines expressing the ERαs described in Figure 6A ± E2 treatment for 45 min. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the vehicle control in each cell line (*, $p < 0.05$; **, $p < 0.005$; two-way ANOVA).
(C and D) Forced recruitment of p300 to inactive ERα binding sites restores enhancer-promoter chromatin looping. (*Top*) Browser tracks for ERα ChIP-seq and GRO-seq in 231/ERαWt and after 45 min. of E2 treatment, shown with the location of the 3C primers. (*Bottom*) 3C-qPCR assays showing the chromatin looping from a distal ERα binding site to the *TGFA* and *HK1* promoters in 231/ERαΔH12 or 231/ERαΔH12-PID cells. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to 231/ERαΔH12 cells for each treatment (*, $p < 0.0001$; two-way ANOVA).
(E) Forced recruitment of p300 to an inactive ERα binding site restores E2-responsive eRNA production and gene expression. RT-qPCR assays in 231/ERαWt, 231/ERαΔH12 or 231/ERαΔH12-PID cells treated in a time course of E2. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Asterisks indicate significant differences compared to the vehicle control in each cell line (*, $p < 0.05$; two-way ANOVA).

[Appendix S17 is on the next page]


A
ERα
p300
Med1
H3K27ac
TGFA enh
Percent of Input
E2
ERα:
Wt
ΔH12
ΔH12-PID
B
HK1 enh
C
TGFA
Distal ERBS
25 kb
ChIP-seq
Wt ERα + E2
GRO-seq
Wt ERα + E2
3C-qPCR
primers
Hub
ΔH12 −E2
ΔH12 +E2
ΔH12-PID −E2
ΔH12-PID +E2
Interaction Frequencies
3C-qPCR (TGFA)
Genomic Coordinates (Chr2)
D
HK1
3C-qPCR (HK1)
Genomic Coordinates (Chr10)
E
TGFA
TGFA eRNA
TGFA mRNA
Relative Expression
(min)
F
HK1
HK1 eRNA
HK1 mRNA

APPENDIX S18
Supplemental Figures Related To Chapter 3

Appendix S18. p300, Mediator, and SRCs link ERα enhancer function to cell growth and clinical outcomes for ER-positive breast cancers *(continued)*.

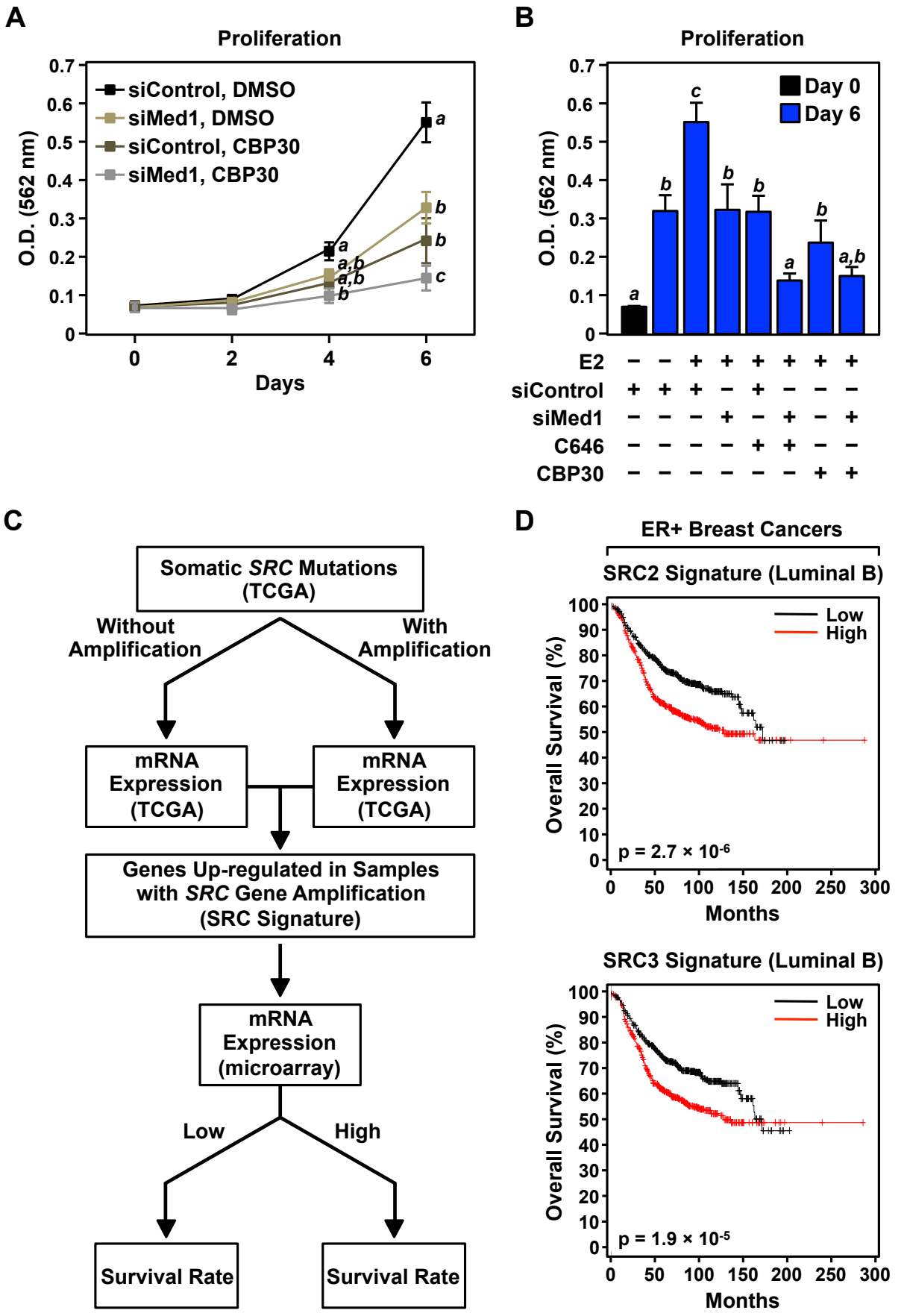
(A) Growth curve showing the combinatorial effects of p300 inhibition and Med1 depletion on E2-dependent MCF-7 cell proliferation. Proliferation was measured after siRNA-mediated Med1-depletion ± the p300/CBP bromodomain inhibitor SGC-CBP-30 (CBP30) in the presence of E2. Each point represents the mean ± S.E.M. for at least three independent biological replicates. Points marked with different letters (*a*, *b, c*) are significantly different from each other ($p < 0.05$; two-way ANOVA).

(B) Proliferation assays showing the combinatorial effects of p300 inhibition and Med1 depletion on MCF-7 cell proliferation. The assays were performed as described in Fig. 7A and panel (A) above. The results are shown for 0 or 6 days of proliferation. Each bar represents the mean + S.E.M. for at least three independent biological replicates. Bars marked with different letters (*a*, *b, c*) are significantly different from each other ($p < 0.05$; one-way ANOVA).

(C) The pipeline of Kaplan-Meier analysis using a gene signature as collective genes up-regulated in tumor samples with *SRC2* or *SRC3* amplification compared to the tumor samples without amplification. Tumor samples on data from TCGA were divided in two groups depending on *SRC* gene amplification. Genes up-regulated in samples with *SRC* gene amplification relative to samples without amplification were denoted as signature genes. The expression of each signature gene (high or low) determined in curated microarray datasets were used to stratify ER-positive breast cancer patients into two groups. Clinical outcomes associated with the microarray datasets were plotted for patients in each group as in Fig. 7D.

(D) Kaplan-Meier plots for Luminal B ER-positive breast cancer patients using a set of genes (signature genes) whose expression is up-regulated in samples with *SRC2* or *SRC3* amplification compared to samples without amplification based on data from TCGA. The expression levels (high or low) of the signature genes determined in curated microarray datasets stratify patients into two groups. Their overall survival rates are shown in the plots.

[Appendix S18 is on the next page]


A
Proliferation
siControl, DMSO
siMed1, DMSO
siControl, CBP30
siMed1, CBP30
O.D. (562 nm)
Days
B
Proliferation
Day 0
Day 6
O.D. (562 nm)
E2
siControl
siMed1
C646
CBP30
C
Somatic SRC Mutations (TCGA)
Without Amplification
With Amplification
mRNA Expression (TCGA)
mRNA Expression (TCGA)
Genes Up-regulated in Samples with SRC Gene Amplification (SRC Signature)
mRNA Expression (microarray)
Low
High
Survival Rate
Survival Rate
D
ER+ Breast Cancers
SRC2 Signature (Luminal B)
Low
High
Overall Survival (%)
p = 2.7 × 10-6
Months
SRC3 Signature (Luminal B)
Low
High
Overall Survival (%)
p = 1.9 × 10-5
Months