SIGNAL REGULATED GENE EXPRESSION: DEFINING THE EFFECTS OF ESTROGEN SIGNALING THROUGH GENOMIC AND PROTEOMIC ANALYSES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Nasun Hah

January 2011

SIGNAL REGULATED GENE EXPRESSION: DEFINING THE EFFECTS OF ESTROGEN SIGNALING THROUGH GENOMIC AND PROTEOMIC ANALYSES

Nasun Hah, Ph.D.

Cornell University 2011

Estrogens play crucial roles in regulating gene expression in physiological and disease states. Estrogens acts through estrogen receptors (ERs) and their binding sites in genomic DNA to modulate transcription by RNA polymerase II. Although recent gene-specific and genomic analyses have provided considerable information about of estrogen-dependent transcription, many aspects of the estrogen signaling network have not yet been elucidated. The goal of my studies was to uncover new information about the immediate and direct effects of estrogen signaling at the cell membrane, in the cytoplasm, and in the nucleus to elucidate the underlying regulatory networks.

First, I investigated an ER transcriptional coregulators, SWI/SNF, an ATP-dependent chromatin remodeling complex. I explored the molecular functions of the BAF57 and BAF180 subunits of SWI/SNF using a quantitative proteomic approach called SILAC (Stable Isotope Labeling by Amino Acids in Cell Culture). I found that depletion of BAF57 results in a significant depletion of BAF180 from the SWI/SNF complex without decreasing the total cellular BAF180 levels, resulting in an accumulation of cells in the G2/M phase. Knockdown of BAF57 also causes transcriptional misregulation of cell cycle-related genes involved in the late G2 checkpoint. Collectively, these studies have elucidated the role of BAF57 and BAF180 in the transcriptional control of cell proliferation.

Second, I have used GRO-Seq (Global Nuclear Run-On and Massively Parallel

Sequencing) to explore the immediate effects of estrogen signaling on the transcriptome of breast cancer cells. I found that estrogen directly regulates a strikingly large fraction of the transcriptome in a rapid, robust, and unexpectedly transient manner. In addition to protein coding genes, estrogen regulates the distribution and activity of all three RNA polymerases, and virtually every class of non-coding RNA that has been described to date. I also identified a large number of previously undetected estrogen-regulated intergenic transcripts, many of which are found proximal to ERα binding sites. These results provide the most comprehensive measurement of the primary and immediate estrogen effects to date.

I expect that genome-wide inferences based on the direct estrogen-regulated transcriptome in combination with estrogen-regulated signaling pathway will be useful for understanding estrogen biology.

BIOGRAPHICAL SKETCH

Nasun Hah was born in Seoul, Korea. She spent her most of her youth in Amsterdam, the Netherlands, as well as Osaka and Tokyo, Japan, where she graduated high school. In 1995, she returned to Seoul and in 1997, she matriculated at Yonsei University, Seoul, Korea, where she majored Biochemistry and developed a strong interest in basic science and research. In 1999, she went to University of California, Irvine, where she particpated in an exchange student program in Biological Studies. In 2001, she completed her undergraduate studies at Yonsei University and received a B.Sc in Biochemisty. In the same year, she started a graduate program at Yonsei University, continuing to major in Biochemisty. During this time, she studied the roles of matrix metalloproteinases-9 in hepatocellular carcinoma, as well as the functions of endostatin, an antiangiogenic peptide in angiogenesis. In 2004, she joined the Graduate Field of Biochemistry, Molecular and Cell Biology at Cornell University, and in 2005, she joined the laboratory of Dr. W. Lee Kraus. In her graduate studies, Nasun studied the mechanisms of estrogen signaling, including the functions of an estrogen receptor co-regulator, the SWI/SNF chromatin remodeling complex, in cell cycle regulation – work that involved a three month stay in the lab of Dr. Marc Timmers in Utrecht, The Netherlands in the fall of 2006. She also investigated the estrogen-regulatedtranscriptome using a novel genomic approache called GRO-seq, as well as the estrogen-regulated the phosphoproteome using SILAC and IMAC technologies. In 2010, she won the Endocrone Society's Presidential Poster Award. In 2010, she successfully defended her Ph.D. thesis and in 2011, she started her postdoctoral studies with Dr. Ron Evans at The Salk Institute in San Diego.

To my parents, for their unconditional support and love

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to my advisor, Dr. W. Lee Kraus, for giving me such a great opportunity to learn about science and life. He has taught me the great value of dedication, the meaning of hard work, and the necessity of patience for the long journeys of science and life. I am also grateful that he was patient in putting up with my strong personality all this time, and for encouraging and supporting me to finish this long journey with a great joy and success.

I would also like to thank my committee members: Dr. John Lis for following my research with interest and for his expert advice throughout my graduate program, and Dr. Ruth Collins for providing scientific insights and encouraging me to be more confident and a better scientist. I also would like to thank Dr. Marc Timmers for his support and assistance with my first research project.

I would like to thank all the past and current members of Kraus lab for their great support, stimulating scientific interactions, and for making the Kraus lab environment fun and enjoyable. I would like to extend a special thanks to Dr. Raga Krishnakumar, who is the first person I met at Cornell University. Starting out as a housemate, we became classmates, then lab mates. She gave me much needed support and was a great friend for me throughout my entire time in graduate school. I am very fortunate and lucky to have her as one of my closest friends in my life! I also would like to thank Dr. Don Ruhl for being such a great and fun bay-mate for the entire time I was in graduate school. I also would like to thank "amazing" Dr. Charles Danko, not only for the great collaborations, but also for his positive attitude, generosity, and encouragement for the past two years.

I would like to thank all my collaborators: Dr. Leighton Core for being a great teacher for GRO-seq and for putting up with a million questions, Dr. Annemieke

Kolkman and Dr. Albert Heck for the proteomic analyses in Utrecht, Dr. Pim Pijnappel for teaching me step by step about SILAC. I also would like to thank all the members of the Timmers lab for welcoming me during my three month visit to Utrecht for the SILAC experiments.

A special thanks to all my dear friends: Jiwon Hwang, Yeonhee Kim, Songeun Lee, Damien Garbett, and my gang: Raga Krishnakumar, Xin Li, Shamoni Maheshwari, Ram Viswanatha and Srich Murugesan. I am lucky to have you all as my friends and grateful for your support and friendship.

I would like to thank my sisters and their families for taking care of me and supporting me so well all the time. Having two sisters is like having two best friends!

Lastly, I would like to express my greatest gratitude and love to my parents, who stand by me, love me, support me, and believe in me unconditionally in every possible way. I would not have been here without you! Thank you for being there for me. I love you!

TABLE OF CONTENTS

APPENDIX 2: Mapping of the Estrogen-Regulated Phosphoproteome Network

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Mechanisms of Transcriptional Regulation by Estrogen Receptors

1.1. Summary

Estrogens are a class of endogenous hormones plays critical roles in a wide range of human physiology, including sexual development, reproduction, cardiovascular and neuronal activity, as well as liver, fat, and bone metabolism. Therefore, dysregulation of estrogen biology is involved in development of a variety of human diseases such breast and uterine cancers, osteoporosis, cardiovascular and neurodegenerative diseases, and insulin resistance. In many cases, estrogen acts through estrogen receptor (ER) proteins to regulate target gene expression, which eventually results in alterations to cellular and organismal physiology. Many studies over the past three decades have elucidated the basic mechanisms of transcriptional regulation by ERs. Also, recent genome-wide studies, which analyzed gene expression and factor binding to genomic DNA on a global scale, have provided a tremendous amount of information about ER-regulated transcription. However, there are still unanswered basic questions. Foremost among these is the question about which genes comprise the set of immediate, direct or primary estrogen-regulated targets, which comprise the set of indirect or secondary targets and what are the molecular mechanisms that control the expression of these direct target genes. Here, I provide an overview of how estrogen signaling controls gene expression.

1.2. Introduction

Over the past three decades, gene specific studies using in vitro and in vivo tools as well as genome-wide studies of transcription factor binding, histone modification, and gene expressions has provided a wealth of information in nuclear receptor mediated transcriptional regulation. Estrogen receptors are one of the most

extensively studied and characterized nuclear receptor families and known to involved in a variety of fundamental developmental and physiological processes as well as human diseases (Deroo and Korach, 2006). Many studies have shown the molecular mechanisms how estrogen and ERs regulate transcription and the emerging picture indicates that ERs bind at promoter-proximal and distal enhancers directly, or indirectly through other transcription factors such as activator protein-1 (AP1), Sp1, NF-κB, etc (Kininis and Kraus, 2008). The binding of ERs promotes the recruitment of coregulators that mediate posttranslational modifications of histone or other transcription factors (Kininis and Kraus, 2008, Dilworth and Chambon, 2001). Liganded ERs and coregulators connect the estrogen signaling to RNA polymerase II (Pol II) and basal transcriptional machinery, and induce the changes in Pol II occupancy or activity of pre-occupied Pol II (Dilworth and Chambon, 2001). This ultimately changes in transcriptome of cells. However, there are still basic questions to be answered about which transcripts comprise the set of direct or primary estrogen-regulated targets and what are the molecular mechanisms that control the expression of these direct targets.

1.3. Estrogen and human physiology

Estrogens are a class of endogenous steroid hormones that plays critical roles in a wide range of human physiology and patho-physiology in both men and women. Estrogen is involved in normal sexual and reproductive system development as well as cardiovascular and neuronal activity and metabolism of liver, fat, and bone. Therefore, dysregulation of estrogen biology is involved in development of a variety of human diseases such breast and uterine cancers, osteoporosis, cardiovascular and neurodegenerative diseases, and insulin resistance (Deroo and Korach, 2006).

Notably, estrogen and ERs have been shown to play a mitogenic role in breast, uterine, and ovarian cancers (Mor et al., 1998; Rochefort, 1998). The etiology of these cancers has shown that estrogen stimulates the unregulated proliferation of some target tissues, which can interfere is normal physiological processes in cells (Platet et al., 2004; Prall et al., 1998).

Currently, ER positive cancers are treated with selective ER modulators (SERMs) and aromatase inhibitors to change estrogen signaling in cancer cells (Osborne et al., 2005), which ultimately change the outcome of ER mediate transcription. It has also been shown that ERα regulates bone mass in both males and females (Sims et al., 2003) and vascular functions (Cid et al., 2002). Some clinical studies have demonstrated that estrogen hormone replacement therapy reduces the incidence of osteoporosis (Sims et al., 2003) and heart attack in women after menopause, indicating that estrogen plays key role in maintaining normal physiology (Cid et al., 2002; Sandberg, 2002). Given the fact that many of targeted therapies focus on changing in expression of ERs and ligands, it is important to understand what are the critical set of direct targets of estrogen signaling and how they are regulated in target cells, which will provide more insights for developing therapeutics with minimal side effects.

1.4. Estrogen receptors: Two isoforms - ERα and ERβ

ERs are members of the nuclear receptor (NR) superfamily, which function as a ligand-regulated transcription factors. The actions of estrogens are mediated by two main ER isoforms, ERα and ERβ. Although these two isoforms are encoded from two separate genes in two different chromosomal locations, they share about 96% homology in their DNA binding domain (DBD), whereas there is only 56% homology

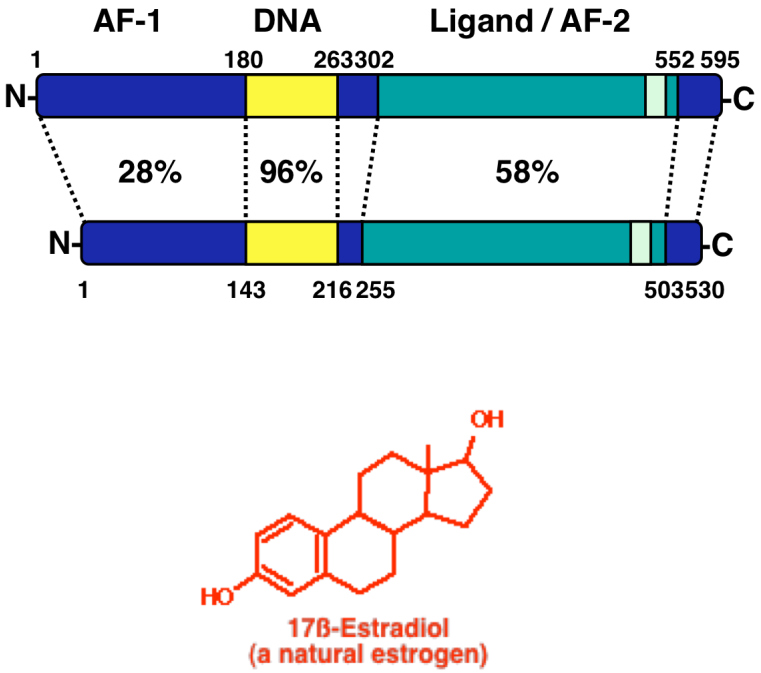


Figure 1.1. Estrogen receptors and estrogen: ligand-regulated, DNA-binding transcription factors

in ligand binding domain (LBD) and 28% in N-terminal activation function 1 (AF-1) (Saville et al., 2000) (Fig 1.1). These two isoforms can also homo- or hetero-dimerize for molecular actions (Bai and Giguere, 2003), indicating that the two isoforms can act together or separately for differentiating their functions. There are overlapping functions between two isoforms due, in part, to a significant homology of their DNA binding domains. However, ERα and ERβ have different expression patterns in tissues, distinguishable differences in structure, and distinct biological functions (Harrington et al., 2003).

1.5. Molecular actions of estrogen receptors

Upon stimulation of cells with estrogen, such as the naturally occurring 17β-estradiol (E2), the hormone can diffuse into cells and bind ERs to promote global changes in gene expression. In the classical or "direct" genomic pathway, once E2 binds to ERs, the receptors dimerize and bind directly to estrogen responsive elements (EREs), which consists of two AGGTCA half sites arranged palindromically around a three basepair spacer (Xiao et al., 2010). In the non-classical genomic actions of estrogen, ERs bind indirectly to cis-acting DNA regulatory elements by "tethering" through other transcription factors, such as AP-1, Sp1, and NF-κB. In either case, chromatin-bound ERs recruit various coregulators to target gene promoters to facilitate transcription by the RNA polymerase II (Pol II) machinery (Cheung and Kraus, 2010; Stender et al., 2010). The distinct classes of coregulators are recruited by ERs, which include (1) histone modifying enzymes (HMEs), (2) bridging or scaffolding factors, (3) mediators and (4) chromatin remodeling complexes (CRCs) (Acevedo and Kraus, 2004; Katzenellenbogen and Katzenellenbogen, 2000) (Fig 1.2). Recent studies have also shown that ERs, however, are also found in the cytoplasm

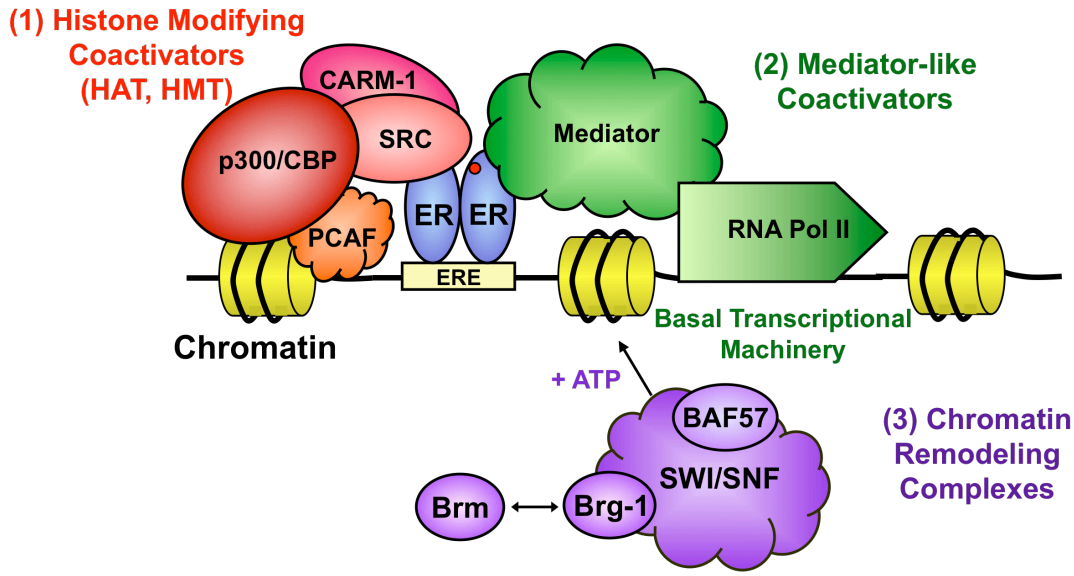


Figure 1.2. Schematic representation of the transcriptional regulation by estrogen receptors. Details are described in the text.

and at the plasma membrane, where they can mediate "membrane initiated" or "extranuclear" cellular signaling responses through kinase cascades (Levin, 2005; Watson et al., 2005). Other proteins associated with the plasma membrane may also bind estrogens and mediate estrogen responses (e.g., the orphan G protein-coupled receptor, GPR30) (Mizukami, 2010), although these are more controversial (Fig. 1.3).

In all cases, estrogen signaling ultimately affects transcriptional outcomes either directly at promoters and enhancers or indirectly through other signaling pathways.

1.6. Estrogen receptor coregulators

As described above, transcriptional regulation by ERs are also largely dependent on recruitments of coregulators that can modulate the activity of Pol II transcriptional machinery. Coregulators usually do not have DNA binding domains and are recruited to the target promoter through protein-protein interactions via nuclear receptor interacting domain, LXXLL motif. Coregulators can act with ERs as coactivators to enhance the transcription such as the p160/steroid receptor coactivators (SRC) family (Dutertre and Smith, 2003) or as corepressors to repress the transcriptional activities such as nuclear receptor corepressor (NCoR) and silencing mediator for retinoic and thyroid receptor (SMRT) (Huang et al., 2002).

The p160/SRC family is one of the most well characterized groups of ER coactivators. Especially, SRC1 and SRC3 have shown to have histone acetyltransferase (HAT) activities, which is required for the formation of open chromatin structure at target promoters. SRC has also shown to recruit p300/CBP (Acevedo and Kraus, 2003; Kraus and Kadonaga, 1998; Kraus et al., 1999) as well as coactivator-associated arginine methyltransferase 1 (CARM1) at the promoter regions

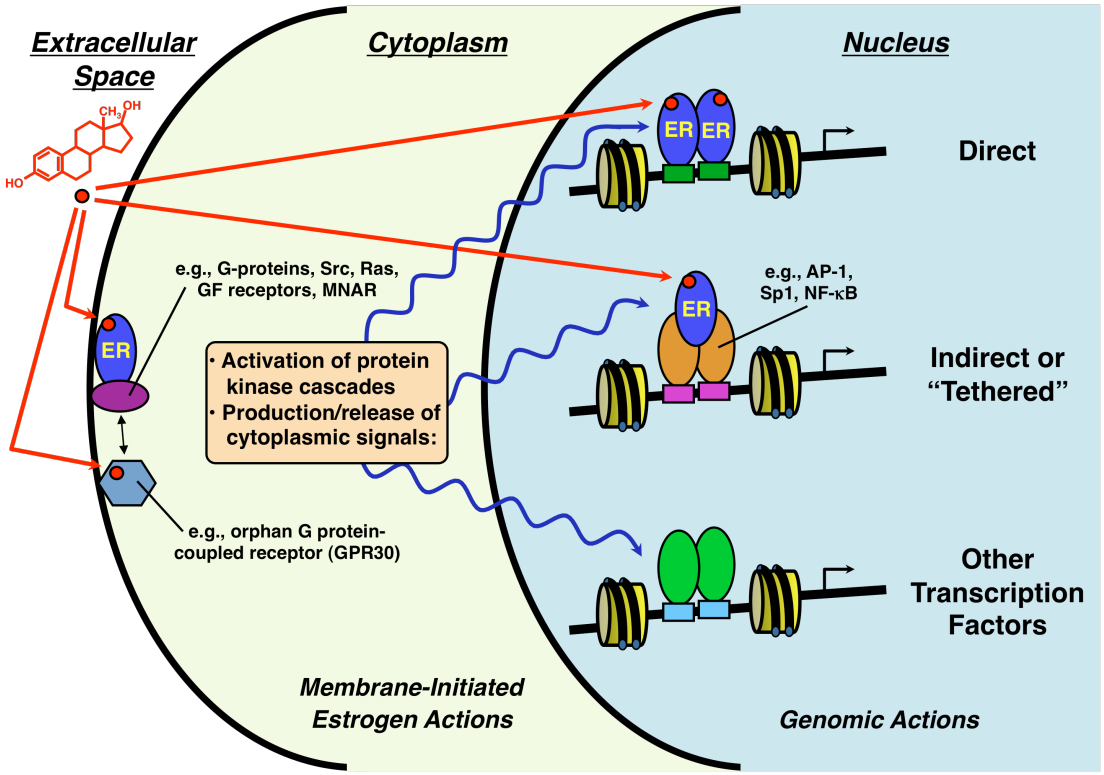


Figure 1.3. Schematic representation of estrogen signaling. Details are described in the text.

to further assist transcription initiation by ERs (Frietze et al., 2008). On the otherhand, the most well-studies corepressor, NCoR/SMRT, has shown to be preferentially recruited to the target promoters in response to antagonist to repress transcription (Cohen et al., 1998).

Chromatin remodeling complexes such as a multi-subunit Swi/Snf ATPase complex are also recruited through the interaction with ERs to alter the structure of nucleosomes, which facilitates other transcriptional factors to access to chromatin (Garcia-Pedrero et al., 2006; Inoue et al., 2002; Zhang et al., 2007). Another class of coregulator, the Mediator complex (e.g., TRAP/DRIP) can interact with both ERs and Pol II transcriptional machinery at target promoters (Wada-Hiraike et al., 2005; Warnmark et al., 2001). Recent studies have identified and characterized more coregulators of ERs that can modulate estrogen-dependent transcription in context specific manners (McKenna et al., 1998; McKenna et al., 1999). Together, the collective actions of all the coregulators that are recruited to estrogen target gene promoters can modulate transcriptional outcomes of estrogen signaling (Fig 1.2).

1.7. Estrogen-dependent transcriptional activation: RNA polymerase II recruitment vs. pausing

How ERs bind to their cognate binding sites upon estrogen treatment and how the DNA sequences at those binding sites affect ER binding on a genomic scale have been studied extensively. However, in order to understand how ERs regulate transcription, it is essential to understand the underlying mechanism how Pol II is regulated at estrogen target promoters. In a widely accepted view of estrogen-mediated transcriptional regulation, ERs and coactivators are recruited to target genes upon estrogen stimulation, which in turn modulate the recruitment of RNA

polymerase II transcriptional machinery to their promoters (Carroll and Brown, 2006; Huh et al., 1999). Under this classical view of the transcriptional regulation, the recruitment of pre-initiation complex is a rate-limiting step (Huh et al., 1999). Interestingly, recent gene specific- and genome-wide studies have shown an alternative view of gene regulation by Pol II, where Pol II is widely distributed at the promoters of target genes prior to estrogen treatment (Kininis et al., 2009; Welboren et al., 2009a; Welboren et al., 2009b). This indicates that there is another level of regulation of Pol II before the productive elongation step.

Recently, promoter proximal-pausing of Pol II has been suggested to play critical roles in assembly of promoter-proximal nucleosome and gene expression (Core and Lis, 2008, 2009). Although the predominant mechanism of the regulation of Pol II remains largely unknown, a number of genome-genome wide studies using a variety of model systems, such as *Drosophila*, yeast, and mammalian cells, showed significant enrichment of Pol II localization at 5' end of genes compare to gene bodies or 3' end of unexpressed genes prior to signal induction, implying post-recruitment regulation of Pol II. The mechanisms of paused Pol II are still largely unknown, however, it has shown to be enriched at the promoters of developmental control or environmental response genes in Drosophila. Especially, the genes that are expressed synchronously during development tend to have more paused Pol II (Chopra et al., 2009; Hendrix et al., 2008). Additionally, the active histone mark such as tri-methylated histone H3K4 and paused Pol II were observed at the promoters of non-transcribing genes in human embryonic stem cells as well as in differentiated cells. Therefore, it has been speculated that promoter proximal pausing of Pol II might be the critical gene regulatory mechanism where cells need to respond to developmental or environmental signals in more rapid, coordinated, and synchronous manners.

In estrogen-induced gene expression, transcription is regulated by both coregulator-dependent recruitment as well as post-recruitment of Pol II, however, it has not been clearly understood what is the determining factors for a subset of genes to prefer either regulatory mechanisms (Core and Lis, 2008, 2009; Seila et al., 2009). There are still remaining questions regarding what other proteins factors are involved and how they are regulated as well as how paused Pol II can be released into productive elongation in response to specific signal such as estrogen (Brookes and Pombo, 2009; Lis et al., 2000; Marshall and Price, 1992).

1.8. In vitro studies: Molecular mechanisms of ER-dependent transcription

Despite of a wealth of information provided by gene specific and genome-wide studies, the detailed molecular mechanisms of nuclear receptor-mediated transcription have largely been elucidated by a variety of in vitro studies using "cell-free" systems over the past three decades (Ruhl and Kraus, 2009). In vitro chromatin assembly and transcription assays allowed dissection of the molecular mechanisms and elucidation of the specific roles of ligands, basal transcriptional machinery, coregulators and nucleosome binding proteins in estrogen-dependent transcriptional regulation.

Using in vitro systems, it was found that agonist ligands are required for NR-mediate transcriptional activation through overcoming repressive characteristics of chromatin structure (Kraus and Kadonaga, 1998). Upon estrogen binding to its receptor, basal transcription from in vitro assembled chromatin was significantly elevated. Also, as the basal transcription machinery was characterized, in vitro systems allowed exploration of the functional interactions between nuclear receptors and the basal transcriptional machinery (Klein-Hitpass et al., 1998; McKenna et al., 1999). Reconstitutions of transcription in vitro allowed the elucidation of how ERα

and the Pol II machinery act together to regulate transcriptional activation. This finding further led to the identification of a variety of NR-coregulators that are required for transcriptional activation.

The p160/SRC family was one of the initially identified NR-coregulators, which functions as a bridging factor between NRs and histone modifying enzymes (Chen et al., 1997; Spencer et al., 1997). Additionally, the protein acetyltransferase p300/CBP was originally found based on cell-based assays; however, its specific roles in transcriptional regulation was elucidated using in vitro chromatin assembly and transcription system in combination with a variety of biochemical analysis (Chakravarti et al., 1996; Kamei et al., 1996). In vitro studies have revealed that the acetyltransferase activity of p300/CBP is required for maximal transcriptional activities of NRs by modifying nucleosome histones, as well as nuclear receptors. In addition to the p160/SRC proteins and p300/CBP, various histone-modifying enzymes, such as CARM1 and PRMT1, have also been shown to facilitate ERα-mediated transcription in vitro (Barrero and Malik, 2006; Xu et al., 2001). In vitro chromatin assembly and transcription systems have also allowed elucidation of the roles of chromatin remodeling complexes, such as ISWI and SWI/SNF (Dilworth et al., 2000), as well as Mediator (TRAP/DRIP) complexes, in NR-mediated transcription (Malik and Roeder, 2000; Rachez and Freedman, 2001).

Recent in vitro studies have also shown the roles of nucleosome binding proteins in ER-dependent transcription. For example, the linker histone H1 and the nucleosome-binding protein PARP-1 bind competitively at overlapping sites on nucleosomes, resulting in the compaction of chromatin and the repression of ERα-dependent transcription (Kim et al., 2004).

In addition to the characterization of all the components that are required for transcriptional regulations, in vitro systems have also helped to elucidate the temporal

order of transcriptional events, as well as the kinetics of ER binding to chromatin template. Although biochemical and in vitro system approaches lack the in vivo context, these methods will continue to contribute to our understanding of the underlying molecular mechanisms of transcriptional regulation. They will also help to validate new concepts and hypotheses that can be drawn from high-throughput genome-wide studies. In particular, in post-genomic era, a wealth of information needs to be supported by mechanistic details of molecular mechanism by in vitro and biochemical studies.

1.9. Genome-wide studies

1.9.1. Gene expression microarrays

Recently, numerous studies using gene expression microarrays have been performed to identify estrogen-regulated target genes. However, the results of these studies have revealed many discrepancies in the numbers of estrogen-regulated genes determined, ranging from 100 to 1,500 in the ERα-positive MCF-7 human breast cancer cell line (Carroll et al., 2006; Kininis and Kraus, 2008; Lin et al., 2004; Welboren et al., 2009a). The variation in the number of binding sites is likely due to many different factors, such as differences in growth conditions, E2-treatment conditions, the coverage provided by the genomic platforms, and different bioinformatic approaches. In the end, these analyses have failed to produce a clear picture of the number and identity of the direct estrogen-regulated gene set.

Preliminary attempts to further define the direct estrogen-regulated transcriptome using the translation inhibitor cycloheximide indicated that only 20 to 30% of the genes showing changes in expression are primary targets (Lin et al., 2004), suggesting that the majority of genes that change expression in response to estrogen

treatment are regulated by secondary or tertiary responses rather than by primary responses. However, determining direct estrogen target genes from cyclohexamide experiments can be problematic in two respects: (1) cyclohexamide is an extremely toxic compound, which has well-characterized non-specific effects on target cells and (2) cyclohexamide does not inhibit the effects of non-coding regulatory RNAs on gene expression, which is becoming widely recognized as an important gene regulatory mechanism underlying the regulation of many genes (Sidhu et al., 1998). Furthermore, technical issues with expression microarrays, such as high levels of noise and poor coverage of non-coding, regulatory RNAs, suggest that they are not the best approach for defining direct target genes.

1.9.2. Chromatin Immunoprecipitation (ChIP)-based analysis

Genomic localization analyses, in which genomic ERα binding sites are mapped using ChIP-based technologies, have also been used in to define direct estrogen-regulated target genes (Welboren et al., 2009b). Studies using ChIP-chip or ChIP-seq have found that ERα binds to between 5,000 and 10,000 locations across the genome. ERα genomic localization studies provided some consistent view of the general patterns of ERα binding across the genome. Surprisingly, only < 5% of ERα binding sites defined by these studies are located within 1 kb of the annotated transcription start sites, with the majority of sites located distally to gene promoters, many by as much as 100 kb (Welboren et al., 2009a). Although these intergenic ERα binding sites are functional and can mediate estrogen-dependent gene expression, their larger number and distal locations make it nearly impossible to use them as a means to define direct estrogen-regulated target promoters. Similar attempts to use ChIP-chip and ChIP-seq to monitor the estrogen-dependent changes in Pol II localization as a surrogate of defining direct estrogen-regulated target promoters provided a better

insights (Welboren et al., 2009a). However, the technical aspects of ChIP was unable to distinguish productive Pol II from preinitiation complex or pre-loaded Pol II at estrogen-regulated gene promoters, and a large number of Pol II localization were detected beyond annotated transcription start sites.

1.9.3. Global nuclear Run-On and massively parallel Sequencing (GRO-seq)

It is clear that new techniques and approaches are required to conclusively identify direct estrogen target genes and to establish the hierarchy of network interactions that characterize secondary changes in gene expression (Core and Lis, 2008, 2009; Core et al., 2008). In my studies described herein, I used Global nuclear Run-On and massively parallel Sequencing (GRO-seq), which is a direct sequencing method that provides a "map" of the position and orientation of all engaged RNA polymerases across the genome at extremely high resolution.

Several aspects of GRO-seq make it uniquely suited to serve as a general method for assessing changes in transcription caused by estrogen stimulation. First, GRO-seq is able to resolve instantaneous changes in the recruitment of RNA polymerases. This rapid temporal resolution can be used to identify the most immediate transcriptional effects of estrogen signaling. Second, GRO-seq is a direct measure of transcription. In contrast, methods that use RNA abundance as a measure of transcription, based either on microarray or deep sequencing technologies, measure a complex function of both RNA stability and processing. Third, unlike polymerase ChIP-seq, GRO-seq identifies the orientation of transcription, allowing the detection of antisense and divergent transcripts, which are clearly a significant fraction of the transcriptome and may have important regulatory roles. Fourth, GRO-seq allows the identification and characterization of transcripts lacking existing annotations in an unbiased way, which is also significant for detecting novel transcripts.

1.10. New facets of estrogen-regulated transcriptome by GRO-Seq analyses

My study using the GRO-Seq method reveals new facets of the estrogen signaling pathway in transcriptional regulation. Using GRO-seq in combination with a novel bioinformatic approach based on Hidden Markov Models (HMMs), I determined all (i.e., both annotated and unannotated) genomic regions in MCF-7 cells that are transcribed by Pols I, II, and III. In addition, I identified the primary transcriptional targets of E2 signaling by focusing on short treatments, prior to the activation of secondary targets (i.e., 0, 10, and 40 min.). This unique approach has revealed a picture of estrogen-regulated gene expression that differs considerably from previous views based on expression microarray studies.

I found that E2 signaling directly regulates a strikingly large fraction (~26%) of the MCF-7 transcriptome in a rapid, robust, and unexpectedly transient manner. In addition to protein coding transcripts, nearly every other class of transcript that has been described to date is also regulated by E2, including annotated microRNAs and non-coding RNAs (e.g., tRNAs, rRNAs), as well as a large number and variety of other unannotated, non-coding transcripts. Increased transcription of rRNA and tRNA genes by Pol I and Pol III, respectively, provided a means to accelerate the translation of newly synthesized protein-coding transcripts.

Finally, comparisons of E2-regulated transcripts defined by GRO-seq to ERα binding sites defined by ChIP-seq suggests that many of the primary targets may be regulated by the direct actions of ERα at its binding sites, allowing us to categorize and understand a much larger fraction of the intergenic ERα binding sites. Collectively, the results presented here reveal many unexpected features of E2-regulation, providing the most comprehensive measurement of the primary and

immediate effects of E2 signaling to date. My results provide a model and resource for understanding rapid signal-dependent transcription in other systems.

1.11. Conclusions and future directions

For the past decade, genome-wide studies have provided tremendous amount of new information in hormone-regulated transcription. Microarray and sequencing-based genomic analyses not only revealed expression changes in mRNA upon hormone stimulation, but also allowed understand transcription factor including nuclear receptor bindings as well as localization of RNA polymerases. The extensive genomic analyses yield some common features of NR-mediated transcription such that nuclear receptor binding sites are more likely to be located distally from 5' and 3' end of annotated genes, the hormone stimulations not only recruit NRs to their binding sites, but also recruit coregulators which can modulate transcriptional outcome, distal NR binding sites are often occupied by lineage- and sequence- specific transcription factors prior to hormone stimulation and distal enhancer can communicate with promoter-proximal region through long range interactions. Although genomic data generates a wealth of information, most of genomic conclusion that are deduced from the extensive analyses is correlative rather than causative. Therefore, future tasks in the post-genomic era are not only to develop more powerful bioinformatics tools, but also to advance and apply new techniques that can provide more mechanistic details in genome-wide scale.

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., and Kraus, W.L. (2004). Transcriptional activation by nuclear receptors. Essays Biochem *40*, 73-88.

Bai, Y., and Giguere, V. (2003). Isoform-selective interactions between estrogen receptors and steroid receptor coactivators promoted by estradiol and ErbB-2 signaling in living cells. Mol Endocrinol *17*, 589-599.

Barrero, M.J., and Malik, S. (2006). Two functional modes of a nuclear receptor-recruited arginine methyltransferase in transcriptional activation. Mol Cell *24*, 233-243.

Brookes, E., and Pombo, A. (2009). Modifications of RNA polymerase II are pivotal in regulating gene expression states. EMBO Rep *10*, 1213-1219.

Carroll, J.S., and Brown, M. (2006). Estrogen receptor target gene: an evolving concept. Mol Endocrinol *20*, 1707-1714.

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.

Chakravarti, D., LaMorte, V.J., Nelson, M.C., Nakajima, T., Schulman, I.G., Juguilon, H., Montminy, M., and Evans, R.M. (1996). Role of CBP/P300 in nuclear receptor signalling. Nature *383*, 99-103.

Chen, H., Lin, R.J., Schiltz, R.L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M.L., Nakatani, Y., and Evans, R.M. (1997). Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. Cell *90*, 569-580.

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

Chopra, V.S., Hong, J.W., and Levine, M. (2009). Regulation of Hox gene activity by transcriptional elongation in Drosophila. Curr Biol *19*, 688-693.

Cid, M.C., Schnaper, H.W., and Kleinman, H.K. (2002). Estrogens and the vascular endothelium. Ann N Y Acad Sci *966*, 143-157.

Cohen, R.N., Wondisford, F.E., and Hollenberg, A.N. (1998). Two separate NCoR (nuclear receptor corepressor) interaction domains mediate corepressor action on thyroid hormone response elements. Mol Endocrinol *12*, 1567-1581.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791-1792.

Core, L.J., and Lis, J.T. (2009). Paused Pol II captures enhancer activity and acts as a potent insulator. Genes Dev *23*, 1606-1612.

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.

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

Dilworth, F.J., and Chambon, P. (2001). Nuclear receptors coordinate the activities of chromatin remodeling complexes and coactivators to facilitate initiation of transcription. Oncogene *20*, 3047-3054.

Dilworth, F.J., Fromental-Ramain, C., Yamamoto, K., and Chambon, P. (2000). ATP-driven chromatin remodeling activity and histone acetyltransferases act sequentially during transactivation by RAR/RXR In vitro. Mol Cell *6*, 1049-1058.

Dutertre, M., and Smith, C.L. (2003). Ligand-independent interactions of p160/steroid receptor coactivators and CREB-binding protein (CBP) with estrogen receptor-alpha: regulation by phosphorylation sites in the A/B region depends on other receptor domains. Mol Endocrinol *17*, 1296-1314.

Frietze, S., Lupien, M., Silver, P.A., and Brown, M. (2008). CARM1 regulates estrogen-stimulated breast cancer growth through up-regulation of E2F1. Cancer Res *68*, 301-306.

Garcia-Pedrero, J.M., Kiskinis, E., Parker, M.G., and Belandia, B. (2006). The SWI/SNF chromatin remodeling subunit BAF57 is a critical regulator of estrogen receptor function in breast cancer cells. J Biol Chem *281*, 22656-22664.

Harrington, W.R., Sheng, S., Barnett, D.H., Petz, L.N., Katzenellenbogen, J.A., and Katzenellenbogen, B.S. (2003). Activities of estrogen receptor alpha- and beta-

selective ligands at diverse estrogen responsive gene sites mediating transactivation or transrepression. Mol Cell Endocrinol *206*, 13-22.

Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., and Levine, M.S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc Natl Acad Sci U S A *105*, 7762-7767.

Huang, H.J., Norris, J.D., and McDonnell, D.P. (2002). Identification of a negative regulatory surface within estrogen receptor alpha provides evidence in support of a role for corepressors in regulating cellular responses to agonists and antagonists. Mol Endocrinol *16*, 1778-1792.

Huh, J.R., Park, J.M., Kim, M., Carlson, B.A., Hatfield, D.L., and Lee, B.J. (1999). Recruitment of TBP or TFIIB to a promoter proximal position leads to stimulation of RNA polymerase II transcription without activator proteins both in vivo and in vitro. Biochem Biophys Res Commun *256*, 45-51.

Inoue, H., Furukawa, T., Giannakopoulos, S., Zhou, S., King, D.S., and Tanese, N. (2002). Largest subunits of the human SWI/SNF chromatin-remodeling complex promote transcriptional activation by steroid hormone receptors. J Biol Chem *277*, 41674-41685.

Kamei, Y., Xu, L., Heinzel, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S.C., Heyman, R.A., Rose, D.W., Glass, C.K.*, et al.* (1996). A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. Cell *85*, 403-414.

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.

Kim, M.Y., Mauro, S., Gevry, N., Lis, J.T., and Kraus, W.L. (2004). NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. Cell *119*, 803-814.

Kininis, M., Isaacs, G.D., Core, L.J., Hah, N., and Kraus, W.L. (2009). Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol Cell Biol *29*, 1123-1133.

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

Klein-Hitpass, L., Schwerk, C., Kahmann, S., and Vassen, L. (1998). Targets of activated steroid hormone receptors: basal transcription factors and receptor interacting proteins. J Mol Med *76*, 490-496.

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.

Kraus, W.L., Manning, E.T., and Kadonaga, J.T. (1999). Biochemical analysis of distinct activation functions in p300 that enhance transcription initiation with chromatin templates. Mol Cell Biol *19*, 8123-8135.

Levin, E.R. (2005). Integration of the extranuclear and nuclear actions of estrogen. Mol Endocrinol *19*, 1951-1959.

Lin, C.Y., Strom, A., Vega, V.B., Kong, S.L., Yeo, A.L., Thomsen, J.S., Chan, W.C., Doray, B., Bangarusamy, D.K., Ramasamy, A.*, et al.* (2004). Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. Genome Biol *5*, R66.

Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. Genes Dev *14*, 792-803.

Malik, S., and Roeder, R.G. (2000). Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. Trends Biochem Sci *25*, 277-283.

Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol Cell Biol *12*, 2078-2090.

McKenna, N.J., Nawaz, Z., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (1998). Distinct steady-state nuclear receptor coregulator complexes exist in vivo. Proc Natl Acad Sci U S A *95*, 11697-11702.

McKenna, N.J., Xu, J., Nawaz, Z., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (1999). Nuclear receptor coactivators: multiple enzymes, multiple complexes, multiple functions. J Steroid Biochem Mol Biol *69*, 3-12.

Mizukami, Y. (2010). In vivo functions of GPR30/GPER-1, a membrane receptor for estrogen: from discovery to functions in vivo. Endocr J *57*, 101-107.

Mor, G., Yue, W., Santen, R.J., Gutierrez, L., Eliza, M., Berstein, L.M., Harada, N., Wang, J., Lysiak, J., Diano, S.*, et al.* (1998). Macrophages, estrogen and the microenvironment of breast cancer. J Steroid Biochem Mol Biol *67*, 403-411.

Osborne, C.K., Schiff, R., Arpino, G., Lee, A.S., and Hilsenbeck, V.G. (2005). Endocrine responsiveness: understanding how progesterone receptor can be used to select endocrine therapy. Breast *14*, 458-465.

Platet, N., Cathiard, A.M., Gleizes, M., and Garcia, M. (2004). Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. Crit Rev Oncol Hematol *51*, 55-67.

Prall, O.W., Rogan, E.M., and Sutherland, R.L. (1998). Estrogen regulation of cell cycle progression in breast cancer cells. J Steroid Biochem Mol Biol *65*, 169-174.

Rachez, C., and Freedman, L.P. (2001). Mediator complexes and transcription. Curr Opin Cell Biol *13*, 274-280.

Rochefort, H. (1998). [Estrogens, cathepsin D and metastasis in cancers of the breast and ovary: invasion or proliferation?]. C R Seances Soc Biol Fil *192*, 241-251.

Ruhl, D.D., and Kraus, W.L. (2009). Chapter 5 biochemical analyses of nuclear receptor-dependent transcription with chromatin templates. Prog Mol Biol Transl Sci *87*, 137-192.

Sandberg, K. (2002). HRT and SERMs: the good, the bad...and the lovely? Trends Endocrinol Metab *13*, 317-318.

Saville, B., Wormke, M., Wang, F., Nguyen, T., Enmark, E., Kuiper, G., Gustafsson, J.A., and Safe, S. (2000). Ligand-, cell-, and estrogen receptor subtype (alpha/beta)-dependent activation at GC-rich (Sp1) promoter elements. J Biol Chem *275*, 5379-5387.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell Cycle *8*, 2557-2564.

Sidhu, J.S., Marcus, C.B., Parkinson, A., and Omiecinski, C.J. (1998). Differential induction of cytochrome P450 gene expression by 4n-alkyl-methylenedioxybenzenes in primary rat hepatocyte cultures. J Biochem Mol Toxicol *12*, 253-262.

Sims, N.A., Clement-Lacroix, P., Minet, D., Fraslon-Vanhulle, C., Gaillard-Kelly, M., Resche-Rigon, M., and Baron, R. (2003). A functional androgen receptor is not sufficient to allow estradiol to protect bone after gonadectomy in estradiol receptor-deficient mice. J Clin Invest *111*, 1319-1327.

Spencer, T.E., Jenster, G., Burcin, M.M., Allis, C.D., Zhou, J., Mizzen, C.A., McKenna, N.J., Onate, S.A., Tsai, S.Y., Tsai, M.J*., et al.* (1997). Steroid receptor coactivator-1 is a histone acetyltransferase. Nature *389*, 194-198.

Stender, J.D., Kim, K., Charn, T.H., Komm, B., Chang, K.C., Kraus, W.L., Benner, C., Glass, C.K., and Katzenellenbogen, B.S. (2010). Genome-Wide Analysis of Estrogen Receptor-{alpha} DNA Binding and Tethering Mechanisms Identifies Runx1 as a Novel Tethering Factor in Receptor-Mediated Transcriptional Activation. Mol Cell Biol.

Wada-Hiraike, O., Yano, T., Nei, T., Matsumoto, Y., Nagasaka, K., Takizawa, S., Oishi, H., Arimoto, T., Nakagawa, S., Yasugi, T*., et al.* (2005). The DNA mismatch repair gene hMSH2 is a potent coactivator of oestrogen receptor alpha. Br J Cancer *92*, 2286-2291.

Warnmark, A., Almlof, T., Leers, J., Gustafsson, J.A., and Treuter, E. (2001). Differential recruitment of the mammalian mediator subunit TRAP220 by estrogen receptors ERalpha and ERbeta. J Biol Chem *276*, 23397-23404.

Watson, C.S., Bulayeva, N.N., Wozniak, A.L., and Finnerty, C.C. (2005). Signaling from the membrane via membrane estrogen receptor-alpha: estrogens, xenoestrogens, and phytoestrogens. Steroids *70*, 364-371.

Welboren, W.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009a). Genomic actions of estrogen receptor alpha: what are the targets and how are they regulated? Endocr Relat Cancer *16*, 1073-1089.

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. (2009b). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J *28*, 1418-1428.

Xiao, R., Roman-Sanchez, R., and Moore, D.D. (2010). DamIP: a novel method to identify DNA binding sites in vivo. Nucl Recept Signal *8*, e003.

Xu, W., Chen, H., Du, K., Asahara, H., Tini, M., Emerson, B.M., Montminy, M., and Evans, R.M. (2001). A transcriptional switch mediated by cofactor methylation. Science *294*, 2507-2511.

Zhang, B., Chambers, K.J., Faller, D.V., and Wang, S. (2007). Reprogramming of the SWI/SNF complex for co-activation or co-repression in prohibitin-mediated estrogen receptor regulation. Oncogene *26*, 7153-7157.

CHAPTER 2.

A Role for BAF57 in Cell Cycle-Dependent Transcriptional Regulation by the SWI/SNF Chromatin Remodeling Complex[§]

[§] This work was published as Nasun Hah, Annemieke Kolkman, Donald D. Ruhl, W. W. M. Pim Pijnappel, Albert J. R. Heck, H. Th. Marc Timmers, and W. Lee Kraus. A Role for BAF57 in Cell Cycle-Dependent Transcriptional Regulation by the SWI/SNF Chromatin Remodeling Complex. *Cancer Res*. (2010) 70:4402-4411. The text is reprinted here with permission from the publisher. Minor modifications have been made. The contributions to this work by the other authors were as follows: D.D.R., mononucleosome remodeling assay (Fig. 2.5); A.K. and W.M.P., assistance with the proteomic analysis.

2.1. Summary

The SWI/SNF complex is an ATP-dependent chromatin remodeling complex that plays pivotal roles in gene regulation and cell cycle control. In the present study, we explored the molecular functions of the BAF57 subunit of SWI/SNF in cell cycle control via transcription regulation of cell cycle-related genes. We affinity purified SWI/SNF from HeLa cells stably expressing FLAG-tagged BAF47/Ini1 with or without stable shRNA-mediated knockdown of BAF57. The subunit composition of the holo- and BAF57-depleted SWI/SNF complexes from these cells were determined using a quantitative SILAC (Stable Isotope Labeling by Amino Acids in Cell Culture)-based proteomic approach. Depletion of BAF57 resulted in a significant co-depletion of BAF180 from the SWI/SNF complex without decreasing total cellular BAF180 levels. In biochemical assays of SWI/SNF activity, the holo- and BAF57/BAF180-depleted SWI/SNF complexes exhibited similar activities. However, in cell proliferation assays using HeLa cells, knockdown of BAF57 resulted in an accumulation of cells in the G2/M phase, inhibition of colony formation, and impaired growth in soft agar. Knockdown of BAF57 also caused transcriptional misregulation of various cell cycle-related genes, especially genes involved in late G2. Collectively, our results have a identified a new role for BAF57 within the SWI/SNF complex that is required for (1) maintaining the proper subunit composition of the complex and (2) cell cycle progression through the transcriptional regulation of a subset of cell cycle-related genes.

2.2. Introduction

A variety of histone- and chromatin-modifying complexes have evolved to

facilitate the access of nuclear proteins to genomic DNA in chromatin. These include ATPase-containing chromatin remodeling complexes, which mobilize or alter the structure of nucleosomes, the fundamental repeating units of chromatin (Narlikar *et al.*, 2002). SWI/SNF is chromatin remodeling complex that contains either BRG1 or BRM as an ATPase catalytic core subunit, as well as a set of BRG1/BRM associated factors (BAFs) (Mohrmann and Verrijzer, 2005). SWI/SNF plays a key role in the regulation of gene expression and, as such, regulates diverse cellular processes, including the cell cycle (Moshkin *et al.*, 2007; Muchardt and Yaniv, 2001; Roberts and Orkin, 2004). Loss of SWI/SNF function is associated with cancer development, and several subunits of the complex (e.g., BRG1, BRM, BAF47/Ini1/SMARCB1) have been shown to function as tumor suppressors (Reisman *et al.*, 2009; Roberts and Orkin, 2004). Interestingly, many gene loci encoding SWI/SNF subunits are mutated or deleted in cancers (e.g., BAF47/SMARCB1/Ini1/SNF5, BAF57/SMARCE1, and BAF180/Polybromo/PB1) (Decristofaro *et al.*, 2001; Reisman et al., 2009; Roberts and Orkin, 2004). Therefore, it is not surprising that SWI/SNF is a major contributor to tumorigenesis and malignancy. A greater understanding of how individual SWI/SNF subunits contribute to the cellular function of the complex, however, is needed.

Although biochemical assays have shown that BRG1 or BRM and three BAFs (i.e., BAF170, BAF155 and BAF47) are sufficient to remodel nucleosomes (Phelan *et al.*, 1999), the roles of the other BAFs are less clear. Recent studies have suggested roles for these remaining BAFs as modulators of SWI/SNF nucleosome remodeling activity or as a determinant of SWI/SNF specificity (Mohrmann and Verrijzer, 2005). Among the BAFs, BAF57 is only present in higher eukaryotes. It has been shown to cooperate with transcriptional coregulators, such as p160/steroid receptor coactivator (SRC1) (Belandia *et al.*, 2002; Kiskinis *et al.*, 2006) and methyl-CpG binding protein

(MeCP2) (Harikrishnan *et al.*, 2005), as well as bind and regulate the activity of nuclear hormone receptors, such as estrogen receptor and androgen receptor (Belandia et al., 2002; Garcia-Pedrero *et al.*, 2006; Huang *et al.*, 2005; Link *et al.*, 2008; Link *et al.*, 2005). BAF57 facilitates the recruitment of SWI/SNF to nuclear receptors bound at target promoters or enhancers in response to hormone to regulate transcription (Chen *et al.*, 2006). In the studies described herein, we have addressed whether BAF57 functions simply an adapter molecule between SWI/SNF and transcriptional regulators, or if it might play a role in regulating the composition or activity of SWI/SNF at specific promoters.

Although BAF57 is unique to metazoan SWI/SNF (Wang *et al.*, 1998), most other subunits are well conserved from yeast to mammals (Mohrmann and Verrijzer, 2005). The existence of two functionally distinct human SWI/SNF complexes with distinct subunit compositions is also well conserved from yeast to mammals: BAF, which contains BAF250, but not BAF180, and PBAF, which contains BAF180, but not BAF250 (Mohrmann *et al.*, 2004; Mohrmann and Verrijzer, 2005; Nie *et al.*, 2000; Wang, 2003). Interestingly, the BAF180-containing PBAF complex and its homologs have been shown to play a role in cell cycle regulation (Cairns *et al.*, 1996; Cao *et al.*, 1997; Du *et al.*, 1998; Hsu *et al.*, 2003; Huang *et al.*, 2004; Moshkin et al., 2007; Muchardt and Yaniv, 2001; Xue *et al.*, 2000). Furthermore, recent studies have shown that the gene encoding BAF180 is often truncated in breast cancers, leading to disregulation of cell cycle-related genes (Sekine *et al.*, 2005; Xia *et al.*, 2008).

In the studies described herein, we have explored the physical and functional interactions between BAF57 and BAF180 in the control of cell proliferation using proteomic, biochemical, molecular, and cell-based assays. We have found that depletion of BAF57 results in a significant co-depletion of BAF180 from the SWI/SNF complex and the accumulation of cells in the G2/M phase of the cell cycle,

as well as transcriptional misregulation of cell cycle-related genes involved in late G2 phase. Collectively, our studies have shed new light on the role of BAF57 in the transcriptional control of cell proliferation in cancer cells.

2.3. Results

2.3.1. BAF57 promotes the stable association of BAF180 with the hSWI/SNF complex.

In this study, we used a proteomic approach in conjunction with a variety of cell-based assays to systematically address the functions of BAF57 in more detail. Specifically, we explored the role of BAF57 in determining the subunit composition of the hSWI/SNF complex. To this end, we used short hairpin RNA (shRNA) technology to stably knockdown BAF57 in HeLa-Ini1-11 cells (Fig. 2.1A), a HeLa-derived cell line stably expressing a FLAG epitope-tagged version of BAF47/Ini1 (Sif *et al.*, 1998). The FLAG-tagged BAF47/Ini1 subunit allows rapid, single-step immunoaffinity purification of essentially native hSWI/SNF complex from the cells (Fig. 2.1B). We then used SILAC coupled with mass spectrometry to quantitatively determine the subunit composition of the complex without or with BAF57 knockdown (holo- and BAF57-depleted SWI/SNF, respectively). As illustrated in Fig. 2.1C, control cells (i.e., HeLa-Ini1-11 cells expressing an shRNA against luciferase; HeLa-Ini1/LucKD cells) were grown in medium containing heavy isotope-labeled L-arginine and L-lysine, whereas BAF57 knockdown cells (HeLa-Ini1/BAF57KD) were grown in the same medium containing light L-arginine and L-lysine. After six doubling times, the SWI/SNF complexes purified from the two cell lines were combined in a 1:1 ratio based on the amount of the FLAG-tagged BAF47/Ini1 subunits and then subjected to mass spectrometry analysis.

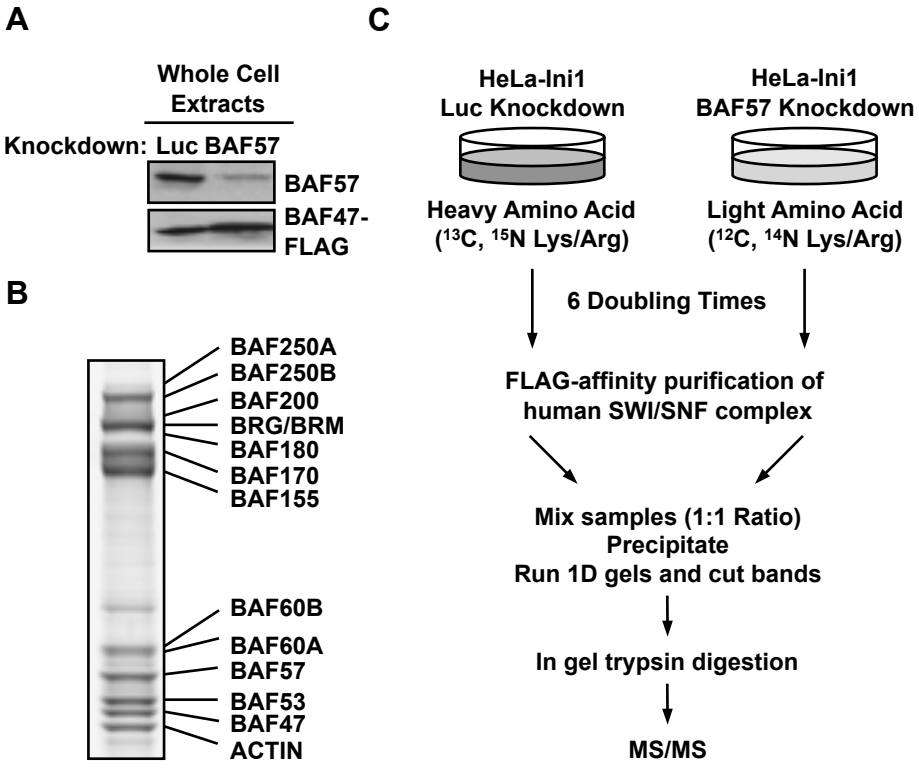
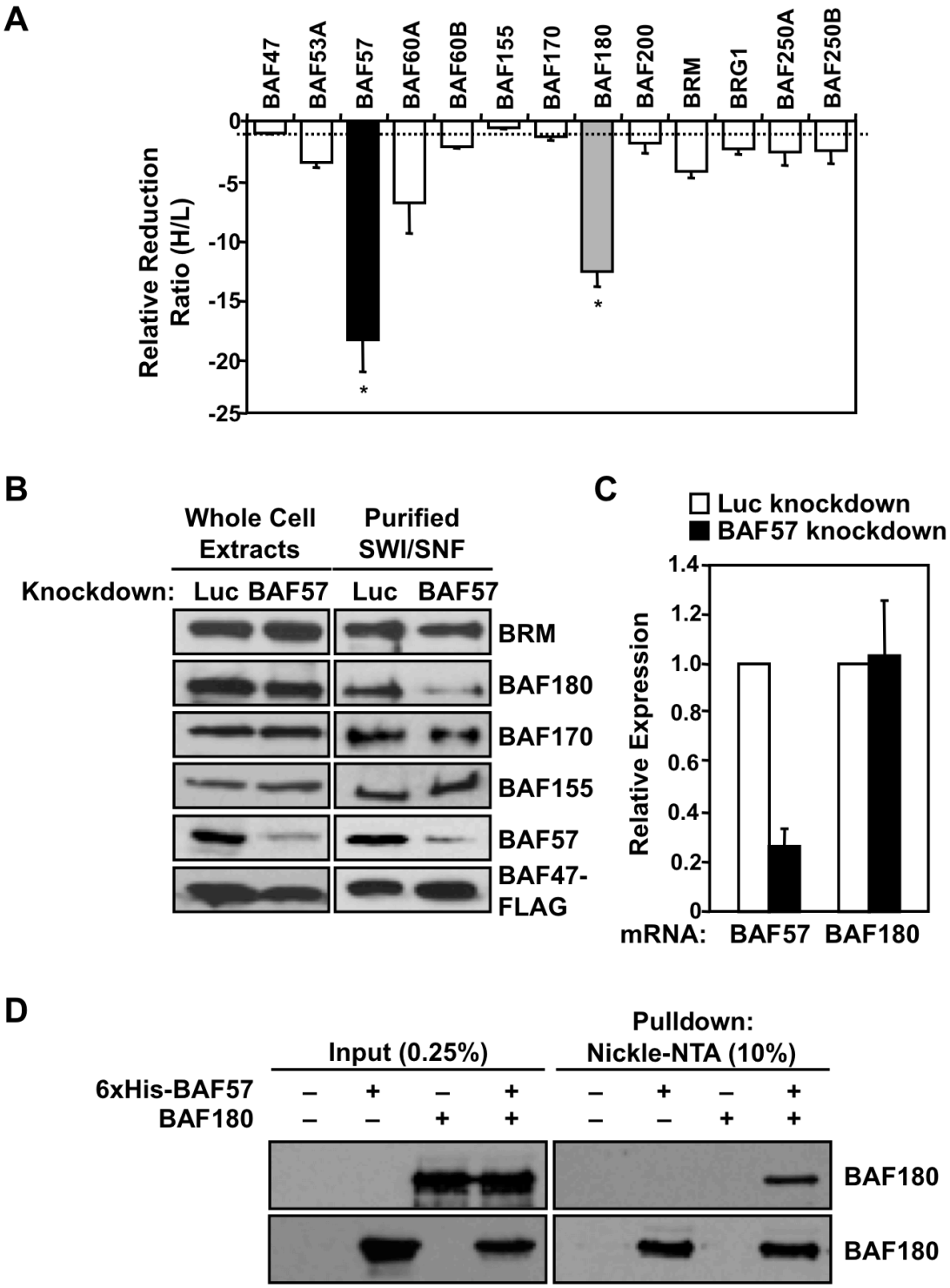


Figure 2.1. Overview of SILAC and immunoaffinity purification of hSWI/SNF complex. (A) Confirmation by Western blotting of BAF57 knockdown in whole cell extracts from HeLa-IniI/Luc cells and HeLa-Ini1/BAF57KD cells. (B) Schematic diagram of the SILAC-based proteomic analysis scheme. (C) SDS-PAGE analysis of FLAG-affinity purified holo- and BAF57-depleted SWI/SNF complexes prepared under SILAC conditions, mixed in a 1:1 ratio, and visualized using Coomassie Blue G-250 staining.

All SWI/SNF subunits were detected with high mass coverage and a large number of unique peptides from both complexes. Using this information, the "heavy:light" ratio of each peak pair was determined and an overall enrichment ratio for each subunit was calculated. In this study, a high ratio of heavy isotope to light isotope (>1) for a given polypeptide indicates a reduced level of the subunit in the BAF57-depleted complex, whereas a low ratio (<1) indicates an elevated level of the subunit in the BAF57-depleted complex. For example, BAF57 had a heavy:light ratio of 18.4, indicating considerably more BAF57 polypeptide in the holo-SWI/SNF complex (heavy) compared to the BAF57-depleted complex (light). Thus, as expected, BAF57 levels were reduced considerably upon knockdown (Fig. 2.2A). BAF57 knockdown also had a striking and statistically significant effect on the levels of BAF180, which had a heavy:light ratio of 12.7 (i.e., it was reduced by 12.7-fold in the complex upon BAF57 knockdown) (Fig. 2.2A). Although a number of other subunits also showed a modest reduction in levels upon BAF57 knockdown, these were not statistically significant (Fig. 2.2A).

To confirm the results of the proteomics analysis, we performed Western blot analyses with whole cell extracts and purified SWI/SNF from the HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cell lines (Fig. 2.2B). These results also showed a reduction in BAF180 levels within the complex upon BAF57 knockdown. In addition, BAF57 depletion did not change the total cellular levels of BAF180 (Fig. 2.2B; see whole cell extract) or affect BAF180 mRNA levels (Fig. 2.2C), indicating a direct effect of BAF57 on the association BAF180 with the SWI/SNF complex. We also found BAF180 to be depleted from SWI/SNF immunoprecipitated from BT549 human breast cancer cells, which do not express a functional BAF57 protein (Kiskinis et al., 2006) (Fig. 2.3). Furthermore, recombinant human BAF57 and human BAF180 co-

Figure 2.2. Knockdown of BAF57 alters the subunit composition of the hSWI/SNF complex by promoting co-depletion of the BAF180 subunit. (A) SILAC-based proteomic analysis of the composition of FLAG-affinity purified holo- and BAF57-depleted SWI/SNF complexes. The ratios of heavy to light isotope for each peptide peak pair for each SWI/SNF subunits were determined by mass spectrometry and are expressed as the relative reduction, which is assigned a negative value to emphasize the reduction. Each bar = mean ± the range from two biological replicates. * = significant at $p < 0.05$, Student's t-test. (B) Confirmation by Western blotting of changes in the subunit composition of SWI/SNF upon BAF57 knockdown. (C) Analysis of BAF57 and BAF180 mRNA levels in HeLa-IniI/Luc cells and HeLa-Ini1/BAF57KD cells by qRT-PCR. Each bar = mean + SEM, $n \geq 5$. (D) Interaction between recombinant BAF57 and BAF180 in cells in the absence of other hSWI/SNF subunits. 6xHis-BAF57 and untagged BAF180 were expressed either individually or together in Sf9 cells using recombinant baculoviruses. Interactions were assessed by nickel-NTA affinity chromatography and Western blotting with BAF57 and BAF180 antibodies.


A
BAF47
BAF53A
BAF57
BAF60A
BAF60B
BAF155
BAF170
BAF180
BAF200
BRM
BRG1
BAF250A
BAF250B
Relative Reduction Ratio (H/L)
0
-5
-10
-15
-20
-25
*
*
B
Whole Cell Extracts
Purified SWI/SNF
Knockdown: Luc BAF57 Luc BAF57
BRM
BAF180
BAF170
BAF155
BAF57
BAF47-FLAG
C
Luc knockdown
BAF57 knockdown
Relative Expression
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
mRNA: BAF57 BAF180
D
Input (0.25%)
Pulldown: Nickle-NTA (10%)
6xHis-BAF57 – + – + – + – +
BAF180 – – + + – – + +
BAF180
BAF180

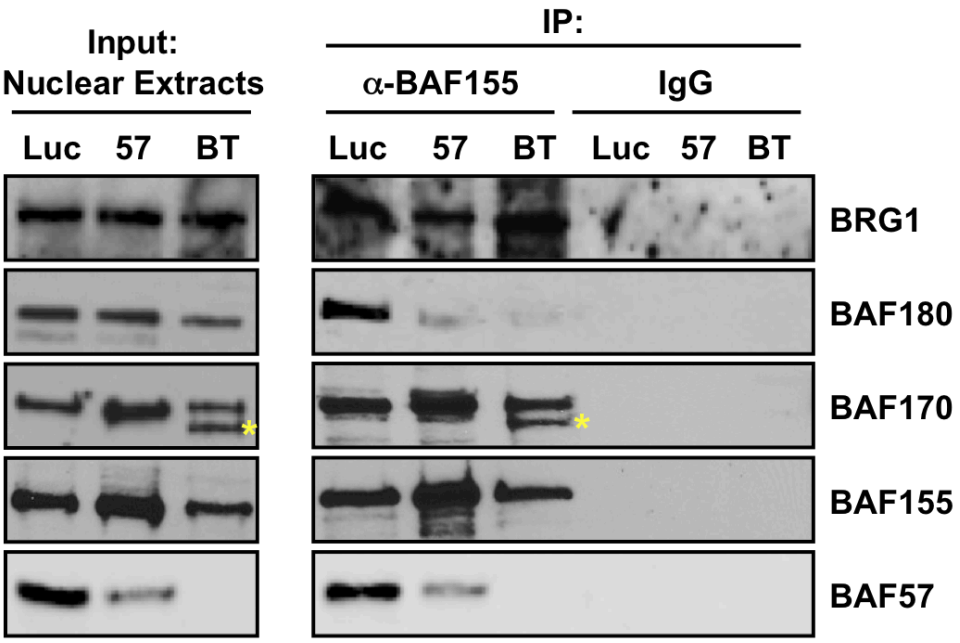


Figure 2.3. SWI/SNF immunoprecipitated from the BAF57-negative BT549 cell line is depleted of BAF180. SWI/SNF was immunoprecipitated from nuclear extracts prepared from HeLa-Ini1/LucKD, HeLa-Ini1/BAF57KD, and BT549 cells using an antibody to BAF155, as well as control IgG. The input and immunoprecipitated material was assayed by Western blotting for the SWI/SNF subunits indicated. BT549 cells express a non-functional truncated version of BAF57 . These results show that depletion of BAF180 from the SWI/SNF complex upon depletion or in the absence of BAF57 occurs in another cell type.

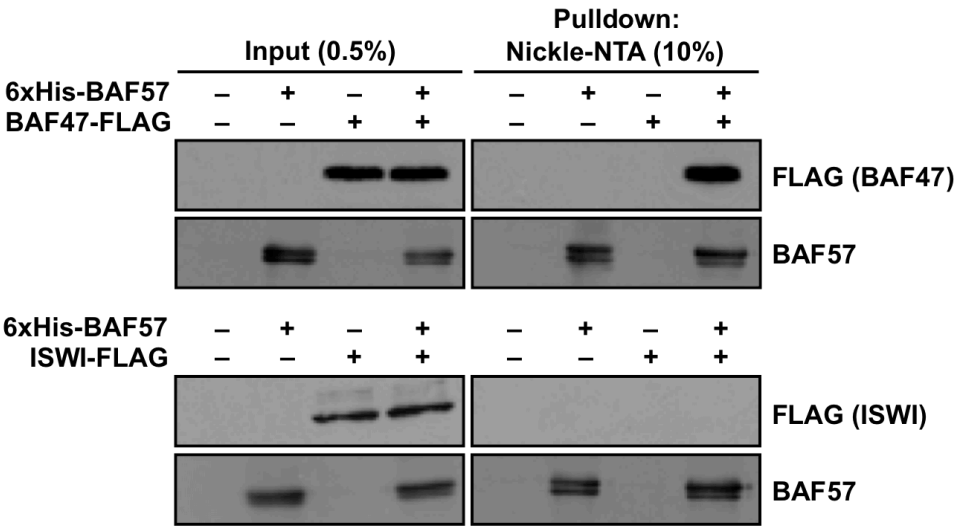


Figure 2.4. BAF57 binds BAF47, but not the chromatin remodeler ATPase subunit ISWI, in the absence of other hSWI/SNF subunits. 6xHis-BAF57 and FLAG-tagged BAF47 or ISWI were expressed either individually or together in Sf9 cells using recombinant baculoviruses. Interactions were assessed by nickel-NTA affinity chromatography and Western blotting with BAF57 and BAF180 antibodies. These results provide positive (BAF47) and negative (ISWI) controls for the interaction assays shown in Fig. 2D.

expressed in insect cells interacted specifically and robustly in the absence of other human SWI/SNF subunits (Fig. 2.2D and 4), suggesting a direct interaction between the two proteins. Together, our quantitative proteomic analysis and follow-up experiments indicate that BAF57 is required for the stable association of BAF180 with the hSWI/SNF complex. Given the interesting biology of BAF180, as well as its dramatic response to BAF57 knockdown, we decided to focus on the functional interplay between BAF57 and BAF180 in the remaining studies.

2.3.2. Knockdown of BAF57 and co-depletion of BAF180 from the SWI/SNF does not affect the nucleosome remodeling activity of SWI/SNF.

To determine if knockdown of BAF57 and the co-depletion of BAF180 affects the nucleosome remodeling activity of SWI/SNF, we performed in vitro nucleosome remodeling assays using SWI/SNF complex purified from the HeLa-Ini1 Luc and BAF57 knockdown cell lines. In this assay, mononucleosomes were assembled using a 571-bp ^{32}P-end-labeled linear DNA template containing the 601-nucleosome positioning element (Fig. 2.5A). Upon addition of purified active hSWI/SNF and ATP, nucleosome remodeling results in the exposure of an HhaI restriction enzyme site. The extent of cleavage by HhaI, which releases a 242 bp ^{32}P-labeled DNA fragment, is indicative of the extent of nucleosome remodeling. As shown in Figs. 5B and 5C, both the holo- and BAF57-depleted SWI/SNF complexes had similar nucleosome remodeling activities, suggesting that depletion of BAF57 and co-depletion of BAF180 does not affect the remodeling activity of the complex. This result fits well with the observation that a core SWI/SNF sub-complex comprising BRG1/BRM, BAF170, BAF155, and BAF47 retains nearly full remodeling activity (Phelan et al., 1999).

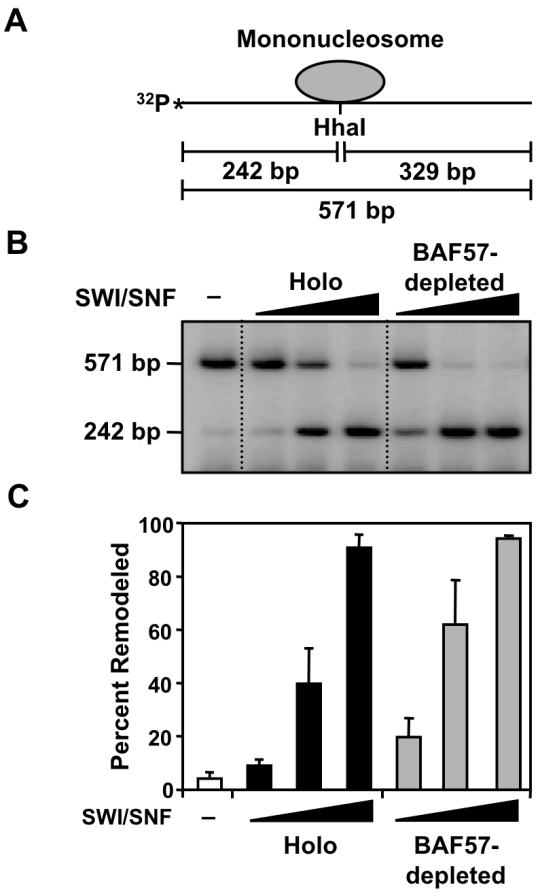


Figure 2.5. BAF57 depletion does not affect hSWI/SNF nucleosome remodeling activity. (A) Schematic of the 571 bp mononucleosome containing a 601-nucleosome positioning sequence. (B) Mononucleosome remodeling assays. A two-fold titration series of intact (i.e., from HeLa-Ini1/LucKD cells) and BAF57-depleted (i.e., from HeLa-Ini1/BAF57KD cells) hSWI/SNF complexes was assessed for remodeling activity. The extent of HhaI cutting was assessed and quantified using native PAGE followed by autoradiographic imaging and phosphoimaging analysis. (C) Quantification of the mononucleosome remodeling assays like those shown in panel B. Each bar = mean + SEM, n = 3.

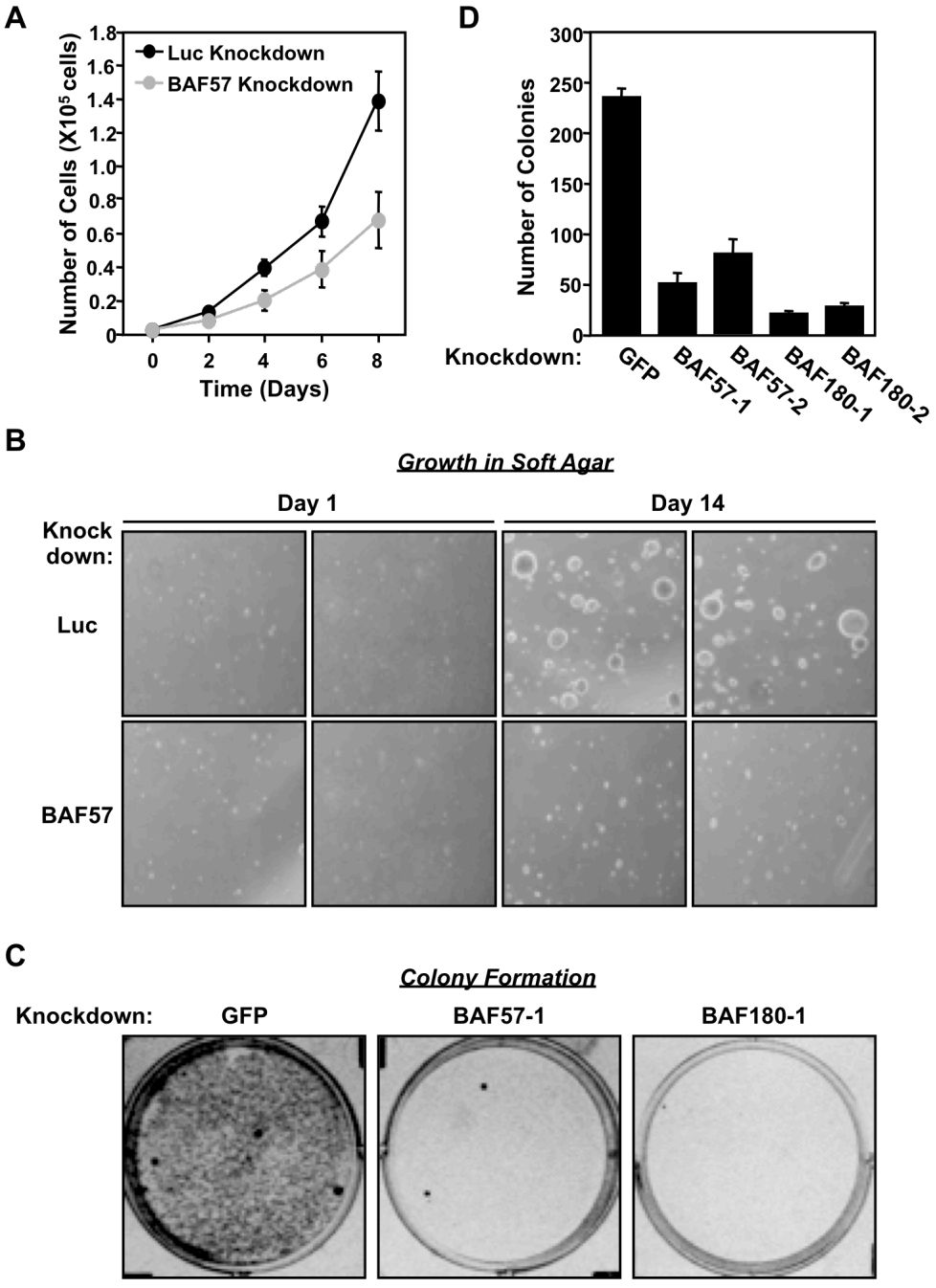
2.3.3. Knockdown of BAF57 reduces cell proliferation and growth in soft agar.

As a follow-up to the biochemical assays, we assessed the role of BAF57 in the control of cell growth and proliferation. First, we compared the rate of cell proliferation of HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells over an 8-day time course. The proliferation rate of the BAF57 knockdown cell line was reduced by approximately two-fold compared to the Luc knockdown control cell line (Fig. 2.6A). Next, we compared the growth of HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells in soft agar over a 14 day time course. This is a common method to monitor anchorage- independent growth, which is one of the hallmarks of cellular transformation. The size, but not the number, of foci was dramatically reduced in the BAF57 knockdown cell line compared to the Luc knockdown control cell line (Fig. 2.6B).

2.3.4. BAF180 knockdown is phenotypically similar to BAF57 knockdown in colony formation assays.

We were unable to determine the effect of BAF180 knockdown in the proliferation and soft agar growth assays described above since HeLa-Ini1 cells and other cell lines are not viable when subjected to stable knockdown of BAF180 (data not shown). Thus, to compare the effects of BAF57 or BAF180 knockdown on cell growth in side-by-side experiments, we performed colony formation assays in HeLa-Ini1-11 cells with shRNA-mediated knockdown of BAF57 or BAF180 under appropriate drug selection, but without stable propagation of the knockdown cells. We tested two different targeting sequences for each factor and both significantly reduced colony formation when compared to control cells expressing an shRNA

Figure 2.6. BAF57 knockdown decreases cell proliferation, growth in soft agar, and colony formation. (A) Effect of stable BAF57 knockdown on the proliferation of HeLa-Ini1 cells. The data shown are for HeLa-Ini1/LucKD cells versus HeLa-Ini1/BAF57KD cells. Each point = mean ± SEM, n = 6. (B) Effect of BAF57 knockdown on the growth of HeLa-Ini1-11 cells in soft agar. The data shown are for HeLa-Ini1/LucKD cells versus HeLa-Ini1/BAF57KD cells. Two panels for each condition are shown. n = 2. (C) Effect of BAF57 or BAF180 knockdown on colony formation in HeLa-Ini1-11 cells. The cells were plated and then infected with recombinant retroviruses expressing shRNAs directed against GFP (as a control), BAF57, or BAF180, followed by selection with puromycin for 7 days. (D) Quantification of the colony formation assays like those shown in panel B. Each bar = mean + SEM, n = 3.


A
Number of Cells (X10^5 cells)
Luc Knockdown
BAF57 Knockdown
Time (Days)
D
Number of Colonies
Knockdown:
GFP
BAF57-1
BAF57-2
BAF180-1
BAF180-2
B
Growth in Soft Agar
Day 1
Day 14
Knock down:
Luc
BAF57
C
Colony Formation
Knockdown:
GFP
BAF57-1
BAF180-1

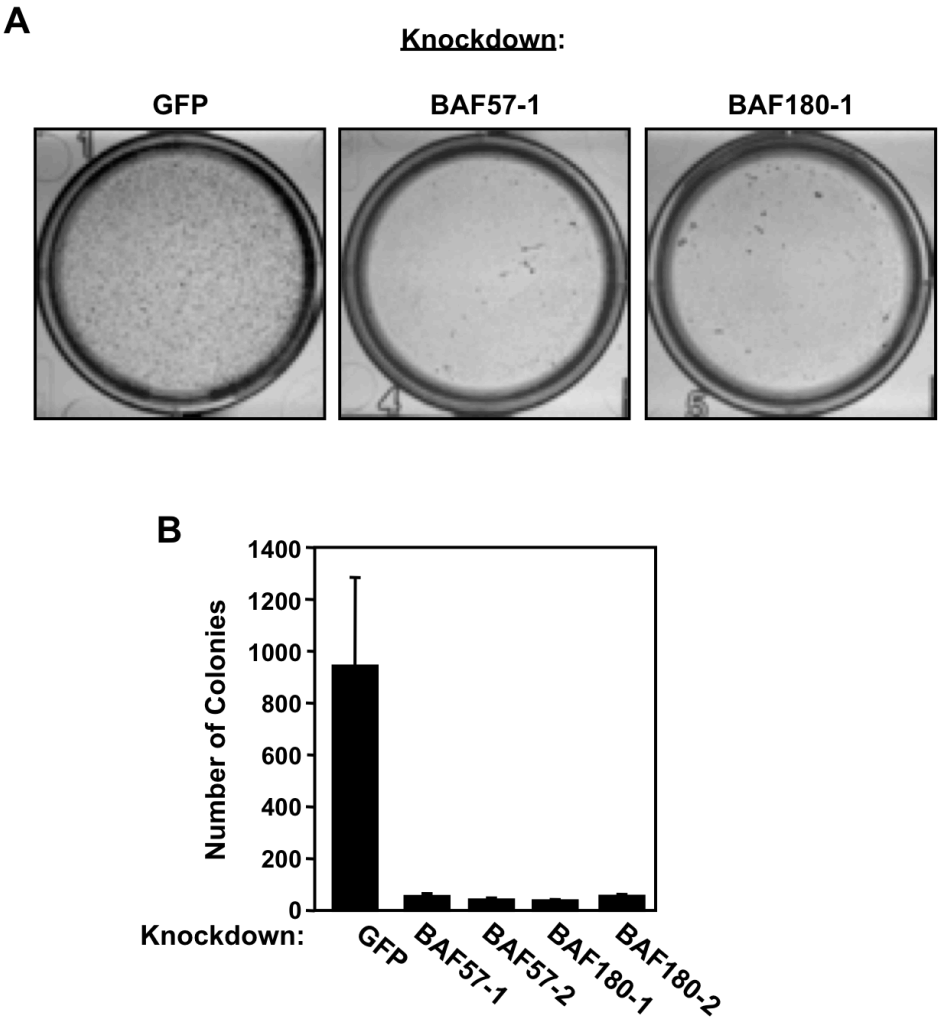


Figure 2.7. BAF57 knockdown decreases colony formation in MCF-7 human breast cancer cells. (A) Effect of BAF57 or BAF180 knockdown on colony formation in MCF-7 cells. The cells were plated and then infected with recombinant retroviruses expressing shRNAs directed against GFP (as a control), BAF57, or BAF180, followed by selection with puromycin for 7 days. (B) Quantification of the colony formation assays like those shown in panel A. Each bar = mean + SEM, n = 3.

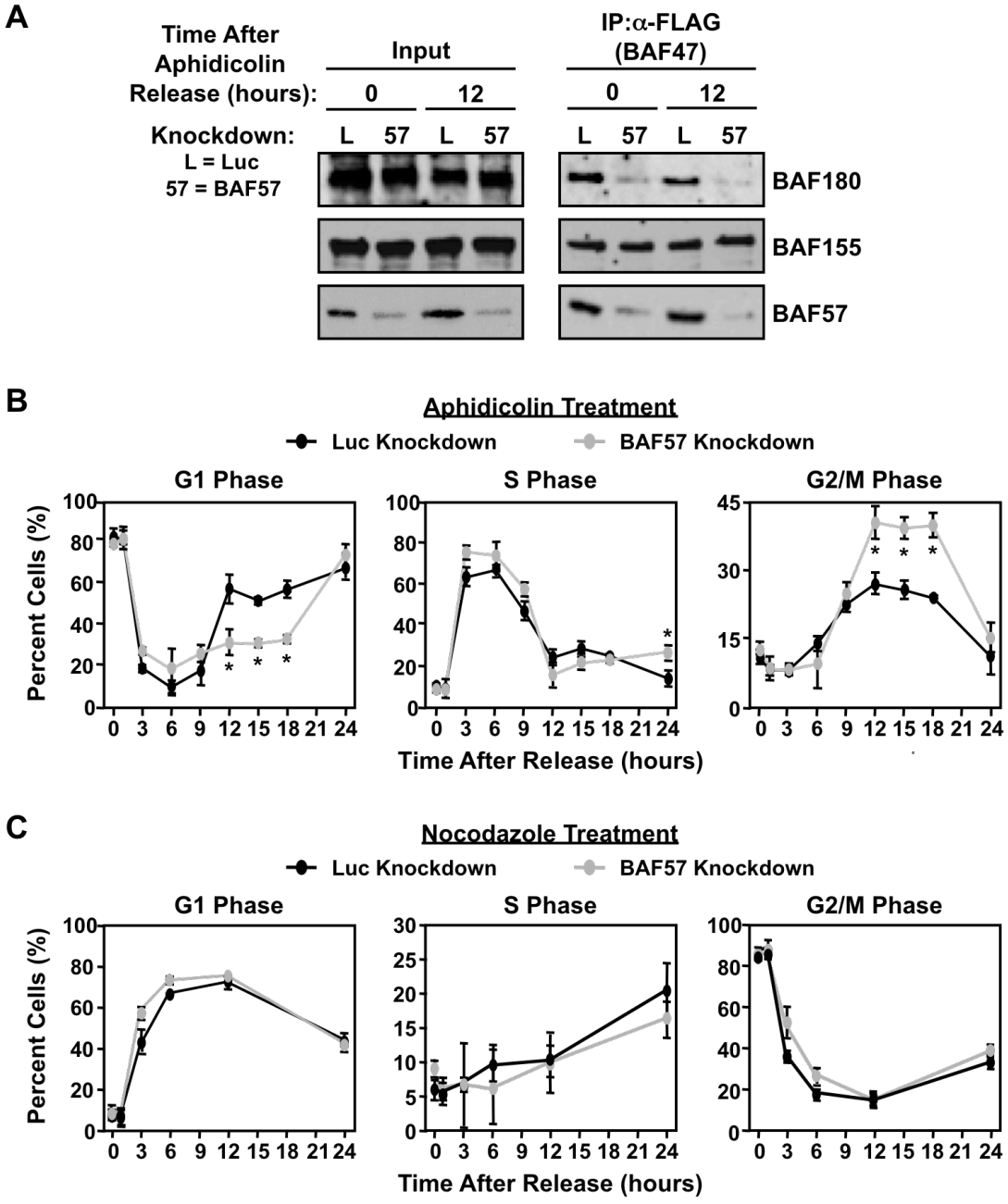
targeting GFP (Fig. 2.6C and 6D). Similar results were observed in MCF-7 human breast cancer cells (Fig. 2.7). Together, the results of these assays indicate an important role for both BAF57 and BAF180 in supporting cell proliferation.

2.3.5. Knockdown of BAF57 and co-deletion of BAF180 from SWI/SNF promotes the accumulation of cells in G2/M phase.

The cell growth and proliferation assays shown above (Fig. 2.6) suggest that depletion of BAF57 or BAF180 slows or inhibits progression through the cell cycle. To explore this possibility in more detail, we performed cell cycle analyses of HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells synchronized using aphidicholin or nocodazole, which arrest cells at the G1/S transition or G2/M transition (or early M phase), respectively (Burke, 2000; Lalande, 1990). Immunoprecipitation experiments from cells arrested with aphidicholin and released confirmed the depletion of BAF57 and BAF180 an 0 and 12 hours post-release (Fig. 2.8A).

The distribution of cells at different points in the cell cycle was analyzed by flow cytometry at 0, 1, 3, 6, 9, 12, 15, 18, and 24 hours after release from arrest. Cells synchronized with aphidicholin and released from G1 arrest showed significant differences in the cell cycle progression profile between the Luc and BAF57 knockdown cells at the 12, 15, and 18 hour time points (Fig. 2.8B). Specifically, the BAF57 knockdown cells showed a significant reduction in G1 phase (~50% of control cells) and a significant accumulation in G2/M phase (~1.7-fold compared to the control cells). In contrast, cells synchronized using nocodazole and released from G2/M arrest did not show significant differences between the Luc and BAF57 knockdown cells (Fig. 2.8C). This may be due to the fact that nocodazole blocks polymerization of microtubules in early M phase (Burke, 2000). If BAF57 knockdown promotes accumulation in G2 phase, as suggested by Fig. 2.5B, then

Figure 2.8. BAF57 knockdown increases the accumulation of cells in G2/M. HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells were arrested either at G1/S phase using aphidicolin (5 μg/ml) (panels A and B) or at G2/M phase using nocodazole (100 nM) (panel C). (A) Knockdown of BAF57 causes the depletion of BAF180 from SWI/SNF at multiple points in the cell cycle. After release arrest at the time points indicated, SWI/SNF was immunoprecipitated from nuclear extracts prepared from the cells and the levels of different subunits were analyzed by Western blotting. HeLa-Ini1/LucKDHeLa-Ini1/BAF57KD(B and C) After release from arrest, the cells were fixed, stained with propidium iodide followed, and analyzed by flow cytometry at the time points indicated. Each point = mean ± SEM, n = 3. * = significant at $p < 0.01$, Student's t-test.


A
Time After Aphidicolin Release (hours):
Input
IP:α-FLAG (BAF47)
0
12
0
12
Knockdown:
L 57 L 57 L 57 L 57
L = Luc
57 = BAF57
BAF180
BAF155
BAF57
B
Aphidicolin Treatment
Luc Knockdown
BAF57 Knockdown
G1 Phase
S Phase
G2/M Phase
Percent Cells (%)
Time After Release (hours)
C
Nocodazole Treatment
Luc Knockdown
BAF57 Knockdown
G1 Phase
S Phase
G2/M Phase
Percent Cells (%)
Time After Release (hours)

nocodazole acts after BAF57, which would allow the BAF57 cells an opportunity to "catch up" to the control cells upon arrest and abrogate any obvious effect. Together, the results of these assays show that BAF57 has a specific effect on cell cycle progression at late G2 phase.

Collectively, the results from our proteomic analyses, cell proliferation assays, and colony formation assays suggest that BAF57 and BAF180 function together within the SWI/SNF complex to support cell growth, proliferation, and progression through the cell cycle.

2.3.6. Knockdown of BAF57 and co-depletion of BAF180 from SWI/SNF alters the expression cell cycle-related genes.

A major function of SWI/SNF is to regulate the transcriptional activation or repression of a subset of target genes (Mohrmann and Verrijzer, 2005; Moshkin et al., 2007; Narlikar et al., 2002). Based on the results of our cell cycle analyses shown in Fig. 2.8, we hypothesized that BAF57 might be required for the expression of genes involved in cell cycle progression. We compared the expression of a panel of more than 30 cell cycle-related genes between the HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells at various time points after release from aphidicholin arrest (Fig. 2.9A and data not shown). Interestingly, we found that BAF57 knockdown delays or inhibits the expression of a subset of genes required for progression through G2/M phase, including CCNB1, CDC2, CENPE, CENPF, CCNF, CCNG1, and CENPA (Fig. 2.9A and data not shown). These genes encode factors required at late G2 phase for cell cycle progression by regulating spindle checkpoint activation. MYC, a gene that encodes the oncogenic transcription factor c-Myc, is also down-regulated in BAF57 knockdown cells. c-Myc has positive effect on cell growth and tumorigenic

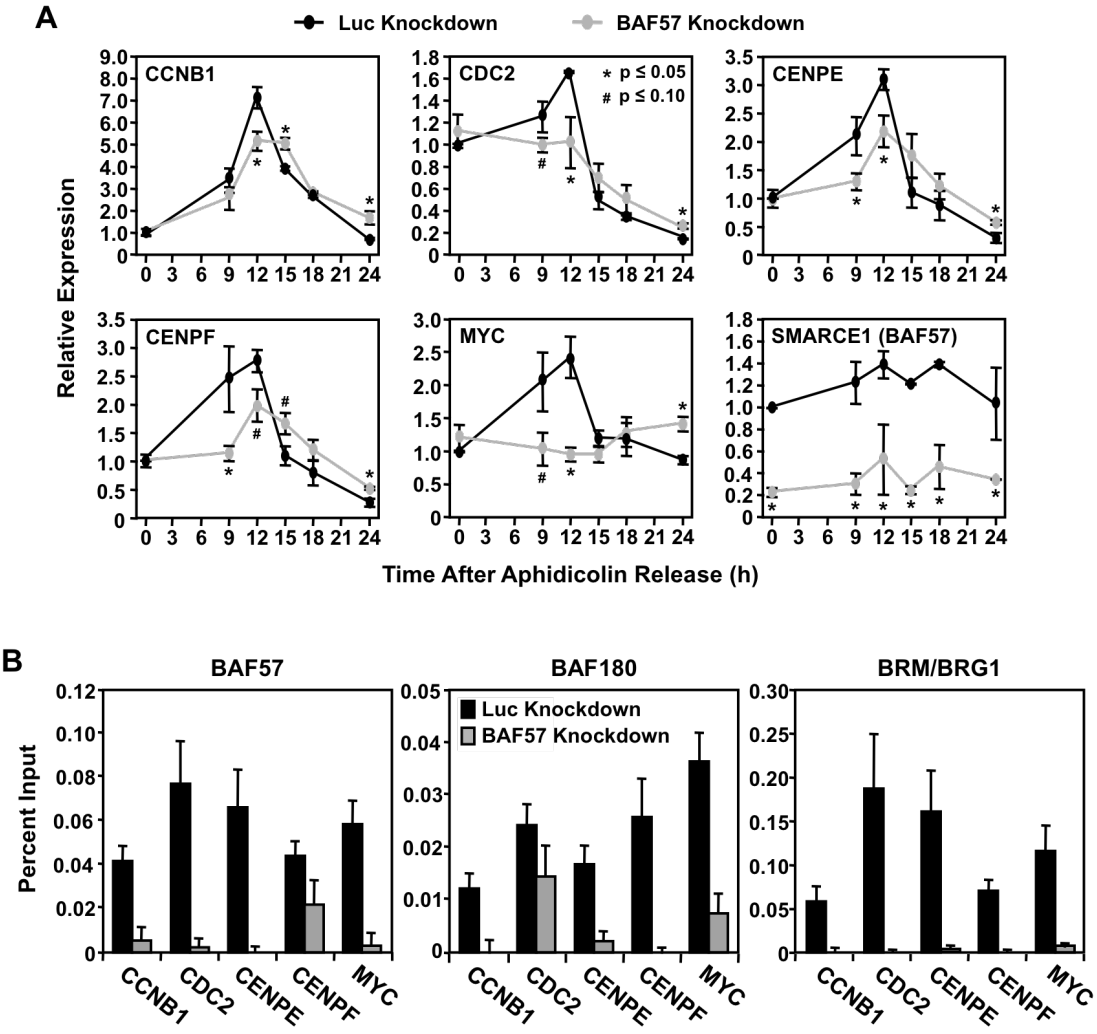
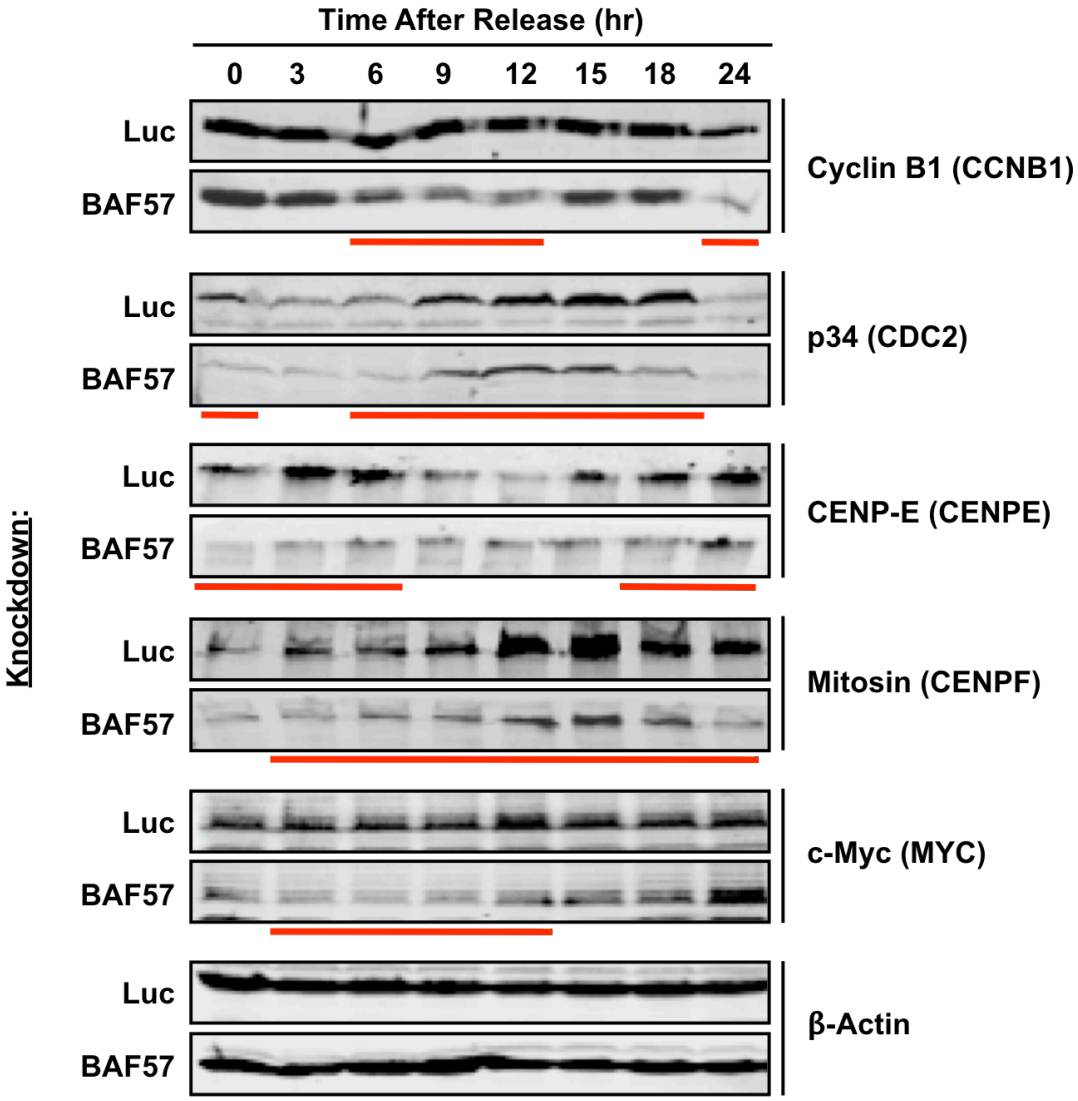


Figure 2.9. BAF57 knockdown alters the expression of genes required for G2/M progression, as well as the recruitment of SWI/SNF to target gene promoters. (A) The expression of cell cycle-related genes in HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells was analyzed using qRT-PCR at the time points indicated post-release from aphidicolin arrest. The results are expressed relative to the 0 hour time point in the HeLa-Ini1/LucKD cells. Each bar = mean + SEM, $n \geq 5$. * and # = significant at $p \leq 0.05$ and $p \leq 0.10$, respectively, by Student's t-test. (B) ChIP-qPCR analysis of the recruitment of SWI/SNF subunits at the promoters of cell cycle-regulated genes. Each bar = mean + SEM, $n \geq 3$.

Figure 2.10. BAF57 knockdown alters the levels of protein products encoded by genes required for G2/M progression. The levels of cell cycle-related proteins in HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells were analyzed by Western blotting at various times post-release from aphidicolin arrest, as indicated. β-actin was used as an internal control. The red bars below the blots indicate time points where obvious differences between the two cell lines was observed. A representative result from three similar experiments is shown. These results show that levels of the cognate protein products encoded by the cell cycle-related genes are also altered during the cell cycle upon BAF57 knockdown in a manner that, for the most part, reflects the expression of the genes (Supplementary Fig. S4), although such an analysis does not account for alternate regulatory mechanims (e.g., MicroRNAs, intrinsic RNA stability).


Time After Release (hr)
0
3
6
9
12
15
18
24
Knockdown:
Luc
BAF57
Cyclin B1 (CCNB1)
p34 (CDC2)
CENP-E (CENPE)
Mitosin (CENPF)
c-Myc (MYC)
β-Actin

transformation (Mateyak *et al.*, 1997). The levels of the cognate protein products encoded by these genes are also altered during the cell cycle upon BAF57 knockdown in a manner that, for the most part, reflects the expression of the genes (Fig. 2.10), although such an analysis does not account for alternate regulatory mechanisms (e.g., microRNAs, intrinsic RNA stability). In contrast, many of the G1/S phase-regulating genes that we screened (e.g., CCND1, CCNE1) were slightly up-regulated or unchanged in the BAF57 knockdown cells compared to the control cells (data not shown), indicating that cell cycle progression is not compromised at G1/S phase, as expected from the cell cycle analyses.

A potential role of the BAF subunits of SWI/SNF is to promote gene-specific recruitment of the complex to a subset of genes. To determine how BAF57 knockdown and concomitant loss of BAF180 from the SWI/SNF complex can affect SWI/SNF recruitment to promoters to alter gene expression, we performed chromatin immunoprecipitation (ChIP) assays with antibodies against BRM/BRG1, BAF57, and BAF180 in HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells, focusing on the promoters of the G2/M-related genes noted above. As expected, knockdown of BAF57 abolished the BAF57 signal at the promoters of these genes (Fig. 2.9B). Also as expected, knockdown of BAF57 inhibited the recruitment of BAF180 to the same promoters (Fig. 2.9B). Moreover, knockdown of BAF57 blocked the recruitment of BRM/BRG1 (Fig. 2.9B). These results indicate that although a BRM/BRG1-containing complex remains intact in the cells upon BAF57 knockdown and co-depletion of BAF180 (Fig. 2.2B), it loses its ability to be properly recruited to the promoters of key SWI/SNF target genes (Fig. 2.9B). Together, the results of our gene regulation assays suggest that BAF57 knockdown and co-depletion of BAF180 from the SWI/SNF complex affects the expression of a subset of genes that are required for progression through G2 phase.

2. 4. Discussion

In this study, we have examined the role of the BAF57 subunit in transcriptional regulation and cell proliferation using a variety of biochemical, proteomic, molecular, and cell-based assays. We have found that RNAi-mediated depletion of BAF57 from cells: (i) alters the composition of the SWI/SNF complex by promoting the dissociation (or preventing the association) of BAF180, (ii) decreases the rate of cell proliferation by promoting the accumulation of cells in the late G2 phase, (iii) alters the composition or prevents the association of SWI/SNF at target gene promoters, and (iv) promotes the down-regulation of a subset of gene that are required for cell cycle progression from G2 to M phase. These findings highlight the functional interplay between SWI/SNF subunits within the complex that regulate biological outcomes, such as cell proliferation.

A role for BAF57 in the regulation of gene expression and cell proliferation by SWI/SNF.

SWI/SNF regulates the transcription of subsets of genes controlling key biological processes, including the cell cycle and tumorigenesis (Moshkin et al., 2007; Muchardt and Yaniv, 2001; Reisman et al., 2009). The ability of SWI/SNF to be (i) targeted to specific promoters by sequence-specific DNA binding transcriptional regulators and (ii) mobilize or structurally alter promoter nucleosomes to make it ideally suited for its role in transcriptional regulation (Mohrmann and Verrijzer, 2005; Narlikar et al., 2002). Although a core SWI/SNF complex containing BRG1 or BRM and three BAFs (i.e., BAF170, BAF155 and BAF47) retains its ability to remodel nucleosomes (Phelan et al., 1999), the remaining BAFs likely function as modulators

of SWI/SNF nucleosome remodeling activity or as a determinant of SWI/SNF specificity (Mohrmann and Verrijzer, 2005).

BAF57, a SWI/SNF subunit unique to higher eukaryotes (Wang et al., 1998), plays a role in transcriptional regulation by binding to (i) DNA through its HMG-like domain and (ii) transcriptional regulators through protein-protein interactions (Belandia et al., 2002; Pal *et al.*, 2003; Wang et al., 1998). Our results from HeLa cells indicate that BAF57-dependent gene expression is required for normal progression through the cell cycle. Likewise, previous studies have shown that inhibition of BAF57 with an amino-terminally-deleted BAF57 dominant negative mutant or an inhibitory polypeptide reduces androgen receptor-dependent cell proliferation (Link et al., 2008; Link et al., 2005). In contrast to these studies showing a positive role for BAF57 in cell cycle progression, other studies have suggested a negative role. For example, re-expression of BAF57 in a BAF57-deficient cell line induces cell cycle arrest and restores contact inhibition, which is accompanied by the induction of a wide variety of genes (Wang *et al.*, 2005b). These contrasting roles of BAF57 - both as a promoter and inhibitor of cell proliferation - suggest that BAF57 function may be context-specific or may require careful regulation of BAF57 levels within a narrowly defined range.

A role for BAF57 in determining the composition of SWI/SNF: Effects on BAF180.

Our results showing that BAF57 (i) interacts with BAF180 and (ii) is required for the stable association of BAF180 with SWI/SNF in multiple cell types provides new insights about the assembly and function of the complex. The loss of BAF57 results in the loss of BAF180 from the complex and a loss of SWI/SNF recruitment to target gene promoters. The link that we have uncovered between BAF57 and BAF180 is

particularly interesting because the latter is the signature subunit of the PBAF complex, one of two functionally distinct mammalian SWI/SNF complexes that have distinct subunit compositions (Mohrmann and Verrijzer, 2005). Previous studies of BAF180 within PBAF have demonstrated a role in cell cycle regulation at G2/M phase, possibly through a mechanism involving the localization of PBAF to the kinetochores of mitotic chromosomes (Xue et al., 2000). BAF180 has also been implicated in the transcriptional regulation of cell cycle-related genes, such as p21 (Xia et al., 2008). These results fit well with our observations that knockdown of BAF57 and co-depletion of BAF180 from the complex leads to the misregulation of a subset of genes that are required for cell cycle progression from G2 to M phase.

Knockdown of BAF57 and co-depletion of BAF180 from SWI/SNF alters the expression cell cycle-related genes.

The expression of the five cell cycle-related genes that we tested is altered during the cell cycle upon BAF57 knockdown, as shown in Fig. 2.9. The levels of the protein products encoded by these genes are also altered during the cell cycle upon BAF57 knockdown in a manner that, for the most part, reflects the expression of the genes (Fig. 2.10). Here we describe the biology of these genes. CCNB1, also known as G2/mitotic-specific cyclin (Innocente *et al.*, 1999; Ren *et al.*, 2005), is abundantly expressed at G2/M phase. CDC2, a catalytic subunit of M-phase promoting factor (MPF), is essential for G1/S and G2/M transitions and normally promotes entry into mitosis (Stark and Taylor, 2004). CENPE and CENPF, which encode centromere proteins (Saffery *et al.*, 1999), are required for recruitment and assembly of kinetochore proteins, proper mitotic progression, and chromosome segregation (Rao *et al.*, 2009; Testa *et al.*, 1994; Wang *et al.*, 2005a). Reduced expression of CENPF and CENPE promotes mitotic defects due to improper chromosome segregation (Laoukili

et al., 2005). MYC is a proto-oncogene that is involved in cell cycle progression, apoptosis, and cellular transformation. Overexpression of MYC has been shown to promote excessive proliferation (Mateyak et al., 1997).

Collectively, our results have a identified a new role for BAF57 within the SWI/SNF complex that is required for the proper regulation of the cell cycle through the transcriptional regulation of a subset of cell cycle-related genes. In addition, our results suggest that the effects of BAF57 are mediated, at least in part, through the association of BAF180 with the complex, which might play a role in directing the assembly of the PBAF versus BAF complexes. Interestingly, the cellular levels of BAF57 protein are precisely regulated by other subunits in the complex (Chen and Archer, 2005), which suggests a complex regulatory network controlling the composition and activity of SWI/SNF.

2.5. Materials and Methods

Antibodies. The antibodies used for Western blot analysis, immunoprecipitation, and chromatin immunoprecipitation (ChIP) are as follows: (1) FLAG (Sigma-Aldrich, Inc.), (2) BAF57, BRG1, and BRM [custom rabbit polyclonal antisera generated by Pocono Rabbit Farm and Laboratory, Inc. (PRF&L); note that the BRM antiserum recognizes both BRM and BRG1, with a preference for BRM], (3) BAF155 and BAF170 (Santa Cruz Biotechnology, Inc. and Bethyl Laboratories, Inc.), (4) BAF180 (Bethyl Laboratories, Inc.), (5) CENP-F/mitosin (Bethyl Laboratories, Inc.), (6) CENP-E and c-Myc (Santa Cruz Biotechnology, Inc.), and (7) cyclin B1, CDC2, and β-Actin (Cell Signaling Technology, Inc.).

Cell culture. HeLa-Ini1-11 cells (Sif et al., 1998), which express FLAG-tagged BAF47, were purchased from the National Cell Culture Center and maintained in Joklik-modified MEM medium (Sigma-Aldrich, Inc.) supplemented with 10% newborn calf serum. For the SILAC analyses, the HeLa-Ini1-11 cells were maintained in custom MEM medium (Cambrex Corporation) supplemented with 10% dialyzed fetal bovine serum (Cambrex Corporation), non-essential amino acids, L-glutamine, L-leucine, penicillin/streptomycin, and either (1) "light" L-lysine and L-arginine or (2) heavy isotope-labeled amino acids, L-lysine-$^{13}C_6$, $^{15}N_2$-HCl (Sigma-Aldrich, Inc./Isotec) and L-arginine-$^{13}C_6$, $^{15}N_4$-HCl (Sigma-Aldrich, Inc./Isotec).

Phoenix-Ampho retrovirus producer cells (ATCC) were cultured in Dulbecco's modified Eagle's medium (Sigma-Aldrich, Inc.) with 10% fetal bovine serum. MCF7 cells were kindly provided by Dr. Benita Katzenellenbogen, University of Illinois, Urbana-Champaign, and maintained in MEM medium (Sigma-Aldrich, Inc.) supplemented with 5% calf serum. BT549 cells were purchased from the American Type Culture Collection and maintained in RPMI1640 medium (Sigma-Aldrich, Inc.) supplemented with 10% fetal bovine serum, 1 mM sodium pyruvate, and 0.023 IU/ml insulin.

RNAi-mediated knockdown of BAF57 and BAF180. RNAi-mediated knockdown of BAF57 or BAF180 in HeLa-Ini1-11 cells was accomplished by retrovirus-mediated gene transfer of short hairpin RNA (shRNA) sequences targeting either BAF57 or BAF180 using the pSuper.Retro system (Oligoengine, Inc.) including control shRNAs targeting luciferase or GFP. Retroviruses were prepared following a standard transfection protocol using the pSuper.Retro shRNA vectors and the Phoenix-Ampho cell line. The recombinant retroviruses were used to transduce HeLa-Ini1-11 cells, which were then selected with 1.0 μg/ml puromycin (Sigma-Aldrich, Inc.). All

knockdown experiments also included control cells transduced with a retroviral vector expressing an shRNA targeting luciferase or green fluorescent protein. Stable knockdown and outgrowth of cells under drug selection resulted in the HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cell lines.

The target short hairpin RNA sequences used in this study were as follows: BAF57-1:, 5'-TACGTGGTTTCTGTATTAA-3', BAF57-2: 5'-AAGGAGAACCGTACATGAGCA-3' (Chen and Archer, 2005), BAF180-1: 5'-AATCCATCAGGACGTCTCATT-3', BAF180-2: 5'-AAGTACAAGGAGGTCGTTTAT-3' (Yan *et al.*, 2005), Luciferase: 5'-GATATGGGCTGAATACAAA-3' (Reynolds *et al.*, 2004), and green fluorescent protein: 5'- GAAGCTGACCCTGAAGTTCATC-3'.

Large-scale purification of FLAG-tagged hSWI/SNF complexes. Purification of human SWI/SNF complex from HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells was performed as described previously (Sif et al., 1998), with modifications. The cells were grown in 3 L to 5 L suspension cultures in a controlled environment bioreactors and used to prepare nuclear extracts as described previously (Dignam *et al.*, 1983), with modifications. Briefly, the cells were collected, washed three times with ice-cold PBS, resuspended in buffer A [10 mM Tris-HCl (pH 7.9), 20% (v/v) glycerol, 10 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 0.2 mM phenylmethylsulfonylfluoride (PMSF), 1 mM benzamidine, 2 μg/ml leupeptin, 2 μg/ml aprotinin, 2 μg/ml pepstatin] and incubated on ice for 30 minutes. The cells were the pelleted, resuspended in buffer A again, and disrupted by Dounce homogenization ~70 stokes with a tight pestle to obtain nuclei, which were collected by centrifugation. The isolated nuclei were resuspended in buffer C [20 mM HEPES (pH 7.9), 25% (v/v) glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1 mM

DTT, 0.2 mM PMSF, 1 mM benzamidine, 2 μg/ml leupeptin, 2 μg/ml aprotinin, 2 μg/ml pepstatin], incubated with constant mixing for 30 minutes at 4°C, and centrifuged for 30 minutes at 17000 rpm in an RC5C centrifuge (Sorvall). The nuclear extract supernatants were dialyzed in BC-100 buffer [20 mM HEPES (pH 7.9), 20% (v/v) glycerol, 100 mM NaCl, 2 mM EDTA, 1 mM DTT, 0.2 mM PMSF] for 4 hours and collected.

For immunoaffinity purification of hSWI/SNF, approximately 20 mg of dialyzed nuclear extract was incubated in batch with FLAG-M2 agarose (Sigma-Aldrich, Inc.) for 8 to 12 hours. The resin was washed in batch once with BC-100, twice with BC-150 (i.e., BC-100 with 150 mM NaCl), twice with BC-300 (i.e., BC100 with 300 mM NaCl), and once with BC-150. The immunoprecipitates were eluted in batch with elution buffer [BC-150 that containing 0.2 μg/μl of FLAG peptide (Sigma-Aldrich, Inc.)], collected, and flash frozen in liq N2, and stored in aliquots at -80°C.

Proteomic analysis of hSWI/SNF complexes.

In-gel tryptic digestion. Approximately 30 μg of purified hSWI/SNF mixed at 1:1 ratio was loaded per 5 mm wide lane and run on a NuPAGE® Novex Bis-Tris Gel (Invitrogen, Inc.). The gel was stained with Coomassie Blue G-250. A gel lane was cut into twelve slices, and each slice was cut into small (approximately 1 mm^3) pieces. The gel pieces were washed, reduced in-gel with dithiothreitol, alkylated with iodoacetamide, and digested with trypsin at a concentration of 10 ng/ml overnight at 37°C as described previously (Wilm and Mann, 1996).

On-line nanoflow liquid chromatography FT-ICR-MS/MS. Peptides generated by in-gel digestion were analyzed by nanoflow liquid chromatography using an Agilent 1100 HPLC system (Agilent Technologies) comprising a solvent degasser, a binary pump, and a thermostated wellplate autosampler, coupled online to a 7-Tesla

LTQ-FT mass spectrometer (Thermo Electron Corporation). The system was operated in a set-up essentially as described previously (Meiring et al., 2002). Aqua C18, 5 μm resin (Phenomenex, Inc.) was used for the trap column and ReproSil-Pur C18-AQ, 3 μm resin (Dr. Maisch GmbH, Ammerbuch, Germany) was used for the analytical column. Peptides were trapped at 5 μl/min in 100% solvent A (0.1 M acetic acid in water) on a 2 cm trap column (100 μm internal diameter, packed in house) and eluted to a 20 cm analytical column (50 μm internal diameter, packed in house) at about 150 nl/min in a 50 min. gradient from 0 to 40% solvent B [0.1 M acetic acid in 8/2 (v/v) acetonitrile/water]. The column eluent was sprayed directly into the ESI source of the mass spectrometer via a butt-connected nano-ESI emitter (New Objective, Inc., Woburn, MA). The mass spectrometer was operated in data dependent mode, automatically switching between MS and MS/MS acquisition. Full scan MS spectra (from m/z 300 to 1500) were acquired in FT-ICR with a resolution of 100,000 at m/z 400 after accumulation to a target value of 1,000,000. The two most intense ions at a threshold above 5000 were selected for collision-induced fragmentation in the linear ion trap at normalized collision energy of 35% after accumulation to a target value of 10,000.

Protein identification. In post analysis processing, raw data were converted to peak lists using Bioworks Browser 3.1 software (Thermo Finnigan, San Jose, CA). For protein identification, MS/MS data were submitted to the UniProtKB/Swiss-Prot 50.8 database using Mascot Version 2.1 (Matrix Science) with the following settings: (1) taxonomy was set to human, (2) precursor and fragment masses were set to 15 ppm and 0.80-Da deviation, (3) trypsin was specified as the proteolytic enzyme, and up to two missed cleavages were allowed, (4) carbamidomethyl cysteine was set as fixed modification, (5) oxidized methionine, $^{13}C_6$-$^{15}N_2$ lysine, and $^{13}C_6$-$^{15}N_4$ arginine were

set as variable modifications, and (6) the peptide ion score cutoff was set to 20 and the protein ion score cutoff to 60.

Protein quantification. Relative quantification ratios of identified proteins were derived by MSQuant, which is open source software (http://msquant.sourceforge.net/). Briefly, peptide ratios between the monoisotopic peaks of "normal" and "heavy" forms of the peptide were calculated and averaged over consecutive MS cycles for the duration of their respective LC-MS peaks in the total ion chromatogram using FT-survey. Peptide ratios of the same protein were averaged to give protein abundance ratios as well as the respective standard deviation. Peptide ratios obtained by using the MSQuant software were all inspected manually. Our experiments, in agreement with data from other groups (Bendall *et al.*, 2008; Park *et al.*, 2009; Van Hoof *et al.*, 2007), showed that HeLa cells convert $^{13}C_6$-$^{15}N_4$-arginine to $^{13}C_5$-$^{15}N_1$-proline. In these experiments the conversion was estimated as 22.5 percent. We corrected the peptide ratio for this conversion as described (Mousson et al., 2008; Park et al., 2009; Van Hoof et al., 2007). The raw mass spectrometric data associated with this manuscript can be downloaded from ProteomeCommons.org Tranche, https://proteomecommons.org/tranche/, using the following hash:
7bYHsY6V9WuVrLKYEs1kabqklEoGyGG/BJ9qoTjaot4Ap9CmXpgt+TX51nA+Aw
yZ7xBjWN3i6FVmG1EotZlRKFg8g8AAAAAAAADqA==.

Standard co-immunoprecipitation of SWI/SNF subunits. Small-scale co-immunoprecipitation of SWI/SNF subunits from HeLa-Ini1/LucKD, HeLa-Ini1/BAF57KD, and BT549 nuclear extracts using an anti-BAF155 antibody was performed in a manner similar to the FLAG affinity purifications described above, with modifications. Approximately 3 μg of anti-BAF155 antibody or bulk IgG (as a control) were incubated with 1 mg of nuclear extract for 8 hours at 4°C. Pre-blocked

protein A beads were then added to the nuclear extract/antibody mixes and incubated for 2 hours at 4°C. The beads were washed in batch as described above for the FLAG affinity purifications, resuspended in SDS sample buffer, and heated at 95°C for 5 min. The inputs and immunoprecipitates were subjected to SDS-PAGE and Western blotting with antibodies to SWI/SNF subunits, as indicated.

BAF57-BAF180 interaction assays. A sequence encoding a 6xHis tag was added to the 5' end of the human BAF57 cDNA in a modified version of the pAcUW51 baculovirus transfer vector (BD Biosciences) by using a PCR-mediated approach. The untagged human BAF180 cDNA, provided by Dr. Ramon Parsons, Columbia University, was cloned into the pFastBac baculovirus vector (Invitrogen, Inc.). Recombinant baculoviruses for the expression of 6xHis-BAF57 and untagged BAF180 were generated using the BaculoGold (BD Biosciences) and Bac-to-Bac (Invitrogen, Inc.) systems, respectively. Recombinant baculoviruses for the expression of FLAG-tagged human BAF47 and FLAG-tagged Drosophila ISWI were provided by Dr. Robert Kingston. MGH/Harvard Medical School and Dr. James Kadonaga, UC San Diego, respectively.

6xHis-BAF57, untagged BAF180, FLAG-tagged BAF47, and FLAG-tagged ISWI were expressed either individually or in combinantion in Sf9 insect cells using the cognate recombinant baculoviruses. The cells were collected and resuspended in 1% NP-40 lysis buffer [20 mM Tris-HCl (pH 7.9), 10% (v/v) glycerol, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.2 mM PMSF, 1 mM benzamidine, 2 μg/ml leupeptin, 2 μg/ml aprotinin, 2 μg/ml pepstatin] and incubated on ice for 15 minutes. The cells were disrupted by Dounce homogenization ~15 strokes with a tight pestle and lysates were collected by centrifugation. Four ml of each extract was incubated with 100 ul of a 50% (v/v) slurry of nickel-NTA resin (Qiagen, Inc.) at 4°C with mixing for 4

hours. The beads were then collected and washed three times using 4 mL of wash buffer [20 mM Tris-HCl (pH 7.9), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.2 mM PMSF, 1 mM benzamidine, 2 μg/ml leupeptin, 2 μg/ml aprotinin, 2 μg/ml pepstatin]. The specifically bound proteins were eluted with elution buffer [20 mM Tris-HCl (pH 7.9), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 250 mM imidazole]. The eluates and input material were subjected to SDS-PAGE and Western blotting with BAF57, BAF180, and FLAG antibodies.

Mononucleosome remodeling assays. Mononucleosomes were assembled on a PCR-generated 571 bp double-stranded DNA fragment containing the 601 nucleosome positioning element (Lowary and Widom, 1998) as previously described (Lee and Narlikar, 2001). Mononucleosome remodeling reactions were assembled in 15 μl (final volume) reactions containing 1.5 nM of ^{32}P-labeled mononucleosomes, 3 mM ATP, 2 mM DTT, 1 U HhaI, 1x remodeling buffer [20 mM Tris (pH 7.5), 25 mM NaCl, 1.25 mM $MgCl_2$, 0.1 mg/ml BSA], as well as purified hSWI/SNF or BAF57-depleted hSWI/SNF as indicated (0.6 to 2.4 nM). The reactions were incubated for 40 min. at 30°C, followed by the addition of 1 μL of a 2.5 mg/ml proteinase K solution with subsequent incubation for 20 min. at 37°C. The DNA was recovered by phenol-chloroform extraction and ethanol precipitation and analyzed on a native 4% polyacrylamide gel run in 1x TBE followed by autoradiographic detection and phosphorimaging analysis (Molecular Dynamics, Inc.).

Cell proliferation and soft agar growth assays. For proliferation assays, HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells were plated at a density of 1 x 10^5 cells per well in a 6 cm dish (28.6 cm^2) and maintained by changing the medium every two days. At the indicated two-day intervals, the cells from duplicate wells were collected

individually by trypsinization, stained with trypan blue, and counted by using a hemocytometer.

For the soft agar growth assays, HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD were resuspended in a 0.3% soft agar matrix (Sigma Type VII) containing MEM medium with 5% calf serum. They were plated at a density of 4 x 10^3 cells per well in a 6-well dish (9.6 cm^2 per well) on a pre-solidified 0.7% soft agar matrix containing basal layer. Colony formation was observed under a microscope after 14 days.

Colony formation assays. For colony formation assays, HeLa-Ini1-11 were plated at a density of 1 x 10^5 cells per well in a 6-well dish (9.6 cm^2 per well) and maintained by changing the medium every two days. At day 0, the cells were infected with recombinant retroviruses expressing shRNAs against GFP (control), BAF57, or BAF180 (described above) and selected with 0.5 μg/ml puromycin for 7 days. Under these conditions, uninfected cells were all killed by day 7. Colonies were fixed with methanol and stained with Giemsa stain. Each well was visualized using a GelDoc system (BioRad, Inc.) under visible light and the number of colonies was counted using Quantity One Software (BioRad, Inc.).

Cell synchronization and cell cycle analyses. HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells were plated at a density of 1 x 10^6 cells per well in a 10 cm dish (78.5 cm^2 dish) and treated with either 100 nM of nocodazole or 5 μg/ml of aphidicolin for 16 hours. Nocodazole-arrested "rounded-up" cells were released from arrest by mechanical detachment (i.e, gentle shaking) and washing in PBS, followed by replating at a density of 1 x 10^6 cells per well in a 10 cm dish (78.5 cm^2 dish). Aphidicholin treated cells were released from arrest by washing with PBS three times and adding fresh medium. Cells were harvested by trypsinization at various time

points, washed with ice-cold PBS twice, and fixed with ice-cold 70% ethanol for 1 h at −20°C. The ethanol-fixed cells were washed with cold PBS and incubated with propidium iodide staining solution [0.1% Triton X-100, 40 μg/ml propidium iodide, 200 μg/ml RNase A] for 30 min at 37°C. Stained cells were analyzed by flow cytometry using BD FACSAria™ (BD-Biosciences, Inc.).

Gene expression analyses by reverse transcription-quantitative PCR (RT-qPCR). Changes in the expression of cell cycle-related genes were analyzed as previously described (Kininis et al., 2007), with a few modifications. HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells were plated at a density of 1 x 10^5 cells per well in a 6-well dish (9.6 cm^2 per well) and then arrested at G1/S phase using aphidicolin for 16 hours, with subsequent release from arrest, as described above. The cells were harvested at the time points indicated and total RNA was isolated using TRIzol® reagent (Invitrogen, Inc.) according to the manufacturer's specifications. Two μg of total RNA were reverse-transcribed by oligo(dT) priming using 600 units of MMLV reverse transcriptase (Promega, Inc.) per reaction according to the manufacturer's specifications. The synthesized cDNA was treated with 3 units of RNase H (Ambion) for 30 min at 37 °C and analyzed by qPCR using a 384-well Prism 770 real time PCR thermocycler (ABI, Inc.) for 45 cycles. The expression was normalized to TATA-binding protein (TBP) mRNA using the primer sets listed at the end of method section.

Chromatin immunoprecipitation (ChIP) assays. ChIP-qPCR assays were performed using buffers and solutions described previously (Kininis et al., 2007), with a few modifications. HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells were grown to ~85% confluence in 15 cm diameter dishes (177 cm^2) crosslinked with 10 mM

dimethyl suberimidate-HCl (DMS) in PBS for 10 min at room temperature, crosslinked with 1% formaldehyde in PBS for 10 min at 37°C, and quenching with 125 mM glycine for 5 min at 4°C. The crosslinked cells were washed with PBS, harvested, lysed with lysis buffer, and subjected to three 15-second bursts of sonication using a Digital Sonifier (Branson, Inc.) to obtain ~500 bp genomic DNA fragments. The chromatin-containing lysate was incubated with antibodies against BRM/BRG1, BAF57, or BAF180, as well as IgG or "no antibody" controls, for 12 hours at 4°C with mixing, after a small aliquot of input material was taken. The immunoprecipitates were collected by incubation with protein G-agarose beads for 1.5 hours at 4°C with mixing. The immunoprecipitates were washed three times with wash buffer and eluted by incubation overnight at 65°C in elution buffer, which also served to reverse the crosslinks. The immunoprecipitated DNA was then digested with proteinase K, extracted with phenol-chloroform, precipitated with ethanol, and analyzed by qPCR using a 384-well Prism 770 real time PCR thermocycler (ABI, Inc.) for 40 cycles using the primer sets listed at the end of method section.

Analyzing the levels of cell cycle-related protein by Western blotting. HeLa-Ini1/LucKD and HeLa-Ini1/BAF57KD cells were plated at a density of 1 x 10^6 cells per well in a 10 cm dish (78.5 cm^2 dish) and treated with 5 μg/ml of aphidicolin for 16 hours. Cells were released from arrest by washing with PBS three times and adding fresh medium. The cells were harvested at the time points indicated, washed with ice-cold PBS three times, and lysed in RIPA buffer. A total of 30 μg of protein per sample was subjected to SDS-PAGE and Western blotting with antibodies to cyclin B1, CDC2, CENP-E, CENP-F/mitosin, c-Myc, and β-Actin.

Primer sets for expression analysis.

CCNB1-Fwd: 5’-CAAGCCCAATGGAAACATCT-3’

CCNB1-Rev: 5’-GGATCAGCTCCATCTTCTGC-3’

CDC2-Fwd: 5’-TTTTCAGAGCTTTGGGCACT-3’

CDC2-Rev: 5’-AGGCTTCCTGGTTTCCATTT-3’

CENPE-Fwd: 5’-GTTGATCTTGCAGGCAGTGA-3’

CENPE-Rev: 5’-TGAAACCACCAACTTGTCCA-3’

CENPF-Fwd: 5’-GCTGCAGAGTTTGGAAAAGG-3’

CENPF-Rev: 5’-GCAGGCTTTCAGATTCCTTG-3’

MYC-Fwd: 5’-GGATTTTTTTCGGGTAGTGGAA-3’

MYC-Rev: 5’-TCCTGTTGGTGAAGCTAACGTT-3’

TBP-Fwd: 5’-AGACCATTGCACTTCGTG-3’

TBP-Rev: 5’-AAATCAGTGCCGTGGTTC-3’

Primer sets for ChIP analysis

CCNB1-Fwd: 5’-GCCTCTGTCACCTTCCAAAG-3’

CCNB1-Rev: 5’-CCACCTGGAGAGCAGTGAA-3’

CDC2-Fwd: 5’-TGGCACACGCAGGTACTACT-3’

CDC2-Rev: 5’-TCTTCCAATCAAAATAACCCTCA-3’

CENPE-Fwd: 5’-AGGAAGCCACAGCTAATGGA-3’

CENPE-Rev: 5’-AACGTGACAAGCCTGTTGTG-3’

CENPF-Fwd: 5’-GCATGACAAACACAGCACCT-3’

CENPF-Rev: 5’-TGATTAGGAAGGCGTGGTCT-3’

MYC-Fwd: 5’-GAGCAGCAGAGAAAGGGAGA-3’

MYC-Rev:I 5’-CAGCCGAGCACTCTAGCTCT-3’

REFERENCES

Belandia, B., Orford, R.L., Hurst, H.C., and Parker, M.G. (2002). Targeting of SWI/SNF chromatin remodelling complexes to estrogen-responsive genes. Embo J *21*, 4094-4103.

Bendall, S.C., Hughes, C., Stewart, M.H., Doble, B., Bhatia, M., and Lajoie, G.A. (2008). Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. Mol Cell Proteomics *7*, 1587-1597.

Burke, D.J. (2000). Complexity in the spindle checkpoint. Curr Opin Genet Dev *10*, 26-31.

Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R.D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell *87*, 1249-1260.

Cao, Y., Cairns, B.R., Kornberg, R.D., and Laurent, B.C. (1997). Sfh1p, a component of a novel chromatin-remodeling complex, is required for cell cycle progression. Mol Cell Biol *17*, 3323-3334.

Chen, J., and Archer, T.K. (2005). Regulating SWI/SNF subunit levels via protein-protein interactions and proteasomal degradation: BAF155 and BAF170 limit expression of BAF57. Mol Cell Biol *25*, 9016-9027.

Chen, J., Kinyamu, H.K., and Archer, T.K. (2006). Changes in attitude, changes in latitude: nuclear receptors remodeling chromatin to regulate transcription. Mol Endocrinol *20*, 1-13.

Decristofaro, M.F., Betz, B.L., Rorie, C.J., Reisman, D.N., Wang, W., and Weissman, B.E. (2001). Characterization of SWI/SNF protein expression in human breast cancer cell lines and other malignancies. J Cell Physiol *186*, 136-145.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Du, J., Nasir, I., Benton, B.K., Kladde, M.P., and Laurent, B.C. (1998). Sth1p, a Saccharomyces cerevisiae Snf2p/Swi2p homolog, is an essential ATPase in RSC and differs from Snf/Swi in its interactions with histones and chromatin-associated proteins. Genetics *150*, 987-1005.

Garcia-Pedrero, J.M., Kiskinis, E., Parker, M.G., and Belandia, B. (2006). The SWI/SNF chromatin remodeling subunit BAF57 is a critical regulator of estrogen receptor function in breast cancer cells. J Biol Chem *281*, 22656-22664.

Harikrishnan, K.N., Chow, M.Z., Baker, E.K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J.M., Jones, P.L., Sif, S.*, et al.* (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat Genet *37*, 254-264.

Hsu, J.M., Huang, J., Meluh, P.B., and Laurent, B.C. (2003). The yeast RSC chromatin-remodeling complex is required for kinetochore function in chromosome segregation. Mol Cell Biol *23*, 3202-3215.

Huang, C.Y., Beliakoff, J., Li, X., Lee, J., Li, X., Sharma, M., Lim, B., and Sun, Z. (2005). hZimp7, a novel PIAS-like protein, enhances androgen receptor-mediated transcription and interacts with SWI/SNF-like BAF complexes. Mol Endocrinol *19*, 2915-2929.

Huang, J., Hsu, J.M., and Laurent, B.C. (2004). The RSC nucleosome-remodeling complex is required for Cohesin's association with chromosome arms. Mol Cell *13*, 739-750.

Innocente, S.A., Abrahamson, J.L., Cogswell, J.P., and Lee, J.M. (1999). p53 regulates a G2 checkpoint through cyclin B1. Proc Natl Acad Sci U S A *96*, 2147-2152.

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.

Kiskinis, E., Garcia-Pedrero, J.M., Villaronga, M.A., Parker, M.G., and Belandia, B. (2006). Identification of BAF57 mutations in human breast cancer cell lines. Breast Cancer Res Treat *98*, 191-198.

Lalande, M. (1990). A reversible arrest point in the late G1 phase of the mammalian cell cycle. Exp Cell Res *186*, 332-339.

Laoukili, J., Kooistra, M.R., Bras, A., Kauw, J., Kerkhoven, R.M., Morrison, A., Clevers, H., and Medema, R.H. (2005). FoxM1 is required for execution of the mitotic programme and chromosome stability. Nat Cell Biol *7*, 126-136.

Lee, K.M., and Narlikar, G. (2001). Assembly of nucleosomal templates by salt dialysis. Curr Protoc Mol Biol *Chapter 21*, Unit 21 26.

Link, K.A., Balasubramaniam, S., Sharma, A., Comstock, C.E., Godoy-Tundidor, S., Powers, N., Cao, K.H., Haelens, A., Claessens, F., Revelo, M.P., *et al.* (2008). Targeting the BAF57 SWI/SNF subunit in prostate cancer: a novel platform to control androgen receptor activity. Cancer Res *68*, 4551-4558.

Link, K.A., Burd, C.J., Williams, E., Marshall, T., Rosson, G., Henry, E., Weissman, B., and Knudsen, K.E. (2005). BAF57 governs androgen receptor action and androgen-dependent proliferation through SWI/SNF. Mol Cell Biol *25*, 2200-2215.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol *276*, 19-42.

Mateyak, M.K., Obaya, A.J., Adachi, S., and Sedivy, J.M. (1997). Phenotypes of c-Myc-deficient rat fibroblasts isolated by targeted homologous recombination. Cell Growth Differ *8*, 1039-1048.

Meiring, H., E. Van der Heeft, G.J.t.H., and Jong, A.P.J.M.d. (2002). Nanoscale LC–MS(n): Technical design and applications to peptide and protein analysis,. Journal of Separation Science *25*, 557–568.

Mohrmann, L., Langenberg, K., Krijgsveld, J., Kal, A.J., Heck, A.J., and Verrijzer, C.P. (2004). Differential targeting of two distinct SWI/SNF-related Drosophila chromatin-remodeling complexes. Mol Cell Biol *24*, 3077-3088.

Mohrmann, L., and Verrijzer, C.P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochim Biophys Acta *1681*, 59-73.

Moshkin, Y.M., Mohrmann, L., van Ijcken, W.F., and Verrijzer, C.P. (2007). Functional differentiation of SWI/SNF remodelers in transcription and cell cycle control. Mol Cell Biol *27*, 651-661.

Mousson, F., Kolkman, A., Pijnappel, W.W., Timmers, H.T., and Heck, A.J. (2008). Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. Mol Cell Proteomics *7*, 845-852.

Muchardt, C., and Yaniv, M. (2001). When the SWI/SNF complex remodels...the cell cycle. Oncogene *20*, 3067-3075.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Nie, Z., Xue, Y., Yang, D., Zhou, S., Deroo, B.J., Archer, T.K., and Wang, W. (2000). A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. Mol Cell Biol *20*, 8879-8888.

Pal, S., Yun, R., Datta, A., Lacomis, L., Erdjument-Bromage, H., Kumar, J., Tempst, P., and Sif, S. (2003). mSin3A/histone deacetylase 2- and PRMT5-containing Brg1 complex is involved in transcriptional repression of the Myc target gene cad. Mol Cell Biol *23*, 7475-7487.

Park, S.K., Liao, L., Kim, J.Y., and Yates, J.R., 3rd (2009). A computational approach to correct arginine-to-proline conversion in quantitative proteomics. Nat Methods *6*, 184-185.

Phelan, M.L., Sif, S., Narlikar, G.J., and Kingston, R.E. (1999). Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. Mol Cell *3*, 247-253.

Rao, C.V., Yamada, H.Y., Yao, Y., and Dai, W. (2009). Enhanced genomic instabilities caused by deregulated microtubule dynamics and chromosome segregation: a perspective from genetic studies in mice. Carcinogenesis.

Reisman, D., Glaros, S., and Thompson, E.A. (2009). The SWI/SNF complex and cancer. Oncogene *28*, 1653-1668.

Ren, L., Feoktistova, A., McDonald, W.H., Haese, G.D., Morrell, J.L., and Gould, K.L. (2005). Analysis of the role of phosphorylation in fission yeast Cdc13p/cyclinB function. J Biol Chem *280*, 14591-14596.

Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S., and Khvorova, A. (2004). Rational siRNA design for RNA interference. Nat Biotechnol *22*, 326-330.

Roberts, C.W., and Orkin, S.H. (2004). The SWI/SNF complex--chromatin and cancer. Nat Rev Cancer *4*, 133-142.

Saffery, R., Earle, E., Irvine, D.V., Kalitsis, P., and Choo, K.H. (1999). Conservation of centromere protein in vertebrates. Chromosome Res *7*, 261-265.

Sekine, I., Sato, M., Sunaga, N., Toyooka, S., Peyton, M., Parsons, R., Wang, W., Gazdar, A.F., and Minna, J.D. (2005). The 3p21 candidate tumor suppressor gene BAF180 is normally expressed in human lung cancer. Oncogene *24*, 2735-2738.

Sif, S., Stukenberg, P.T., Kirschner, M.W., and Kingston, R.E. (1998). Mitotic inactivation of a human SWI/SNF chromatin remodeling complex. Genes Dev *12*, 2842-2851.

Stark, G.R., and Taylor, W.R. (2004). Analyzing the G2/M checkpoint. Methods Mol Biol *280*, 51-82.

Testa, J.R., Zhou, J.Y., Bell, D.W., and Yen, T.J. (1994). Chromosomal localization of the genes encoding the kinetochore proteins CENPE and CENPF to human chromosomes 4q24-->q25 and 1q32-->q41, respectively, by fluorescence in situ hybridization. Genomics *23*, 691-693.

Van Hoof, D., Pinkse, M.W., Oostwaard, D.W., Mummery, C.L., Heck, A.J., and Krijgsveld, J. (2007). An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. Nat Methods *4*, 677-678.

Wang, I.C., Chen, Y.J., Hughes, D., Petrovic, V., Major, M.L., Park, H.J., Tan, Y., Ackerson, T., and Costa, R.H. (2005a). Forkhead box M1 regulates the transcriptional network of genes essential for mitotic progression and genes encoding the SCF (Skp2-Cks1) ubiquitin ligase. Mol Cell Biol *25*, 10875-10894.

Wang, L., Baiocchi, R.A., Pal, S., Mosialos, G., Caligiuri, M., and Sif, S. (2005b). The BRG1- and hBRM-associated factor BAF57 induces apoptosis by stimulating expression of the cylindromatosis tumor suppressor gene. Mol Cell Biol *25*, 7953-7965.

Wang, W. (2003). The SWI/SNF family of ATP-dependent chromatin remodelers: similar mechanisms for diverse functions. Curr Top Microbiol Immunol *274*, 143-169.

Wang, W., Chi, T., Xue, Y., Zhou, S., Kuo, A., and Crabtree, G.R. (1998). Architectural DNA binding by a high-mobility-group/kinesin-like subunit in mammalian SWI/SNF-related complexes. Proc Natl Acad Sci U S A *95*, 492-498.

Wilm, M., and Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. Anal Chem *68*, 1-8.

Xia, W., Nagase, S., Montia, A.G., Kalachikov, S.M., Keniry, M., Su, T., Memeo, L., Hibshoosh, H., and Parsons, R. (2008). BAF180 is a critical regulator of p21 induction and a tumor suppressor mutated in breast cancer. Cancer Res *68*, 1667-1674.

Xue, Y., Canman, J.C., Lee, C.S., Nie, Z., Yang, D., Moreno, G.T., Young, M.K., Salmon, E.D., and Wang, W. (2000). The human SWI/SNF-B chromatin-remodeling complex is related to yeast rsc and localizes at kinetochores of mitotic chromosomes. Proc Natl Acad Sci U S A *97*, 13015-13020.

Yan, Z., Cui, K., Murray, D.M., Ling, C., Xue, Y., Gerstein, A., Parsons, R., Zhao, K., and Wang, W. (2005). PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes. Genes Dev *19*, 1662-1667.

CHAPTER 3.

Global Analysis of the Immediate Transcriptional Effects of Estrogen Signaling Reveals a Rapid, Extensive, and Transient Response

The contributions to this work by the other authors were as follows: C.G.D developed bioinformatics approach based on Hidden Markov Models (HMMs) and performed computational analyses, L.J.C. assisted with generation of GRO-seq libraries, and J.J.W. assisted with bioinformatics and computational analyses.

3. 1. Summary

We report the immediate effects of estrogen signaling on the transcriptome of breast cancer cells using Global Run-On and sequencing (GRO-seq). The data were analyzed using a new bioinformatic approach that allowed us to identify transcripts directly from the GRO-seq data. We found that estrogen signaling directly regulates a strikingly large fraction of the transcriptome in a rapid, robust, and unexpectedly transient manner. In addition to protein coding genes, estrogen regulates the distribution and activity of all three RNA polymerases, and virtually every class of non-coding RNA that has been described to date. We also identified a large number of previously undetected E2 regulated intergenic transcripts, many of which are found proximal to ERα binding sites. Collectively, our results provide the most comprehensive measurement of the primary and immediate estrogen effects to date and a resource for understanding rapid signal-dependent transcription in other systems.

3. 2. Introduction

Over the past three decades, studies using in vitro and in vivo tools to assess DNA binding, coregulator interactions, histone modifications, and transcription output have provided a wealth of information on nuclear receptor-mediated gene regulation. The steroidal hormone estrogen, acting through estrogen receptors (ERs) - one of the most extensively studied members of the nuclear receptor superfamily - plays key roles in a variety of fundamental developmental and physiological processes, as well as many disease states (Deroo and Korach, 2006). Mammals express two ER isoforms, ERα and ERβ, which exhibit distinct tissue-specific expression patterns and biological roles (Deroo and Korach, 2006; Warner et al., 1999). Notably, estrogen-

dependent ERα activity plays an important role in the etiology of about 70 percent of human breast cancers, and is widely used as both a prognostic indicator and a major therapeutic target (Platet et al., 2004). A better understanding of the estrogen-dependent actions of ERα at the molecular level will aid in finding better ways to target ERs for the prevention, diagnosis, and treatment of human diseases, such as breast cancers, as well as enhance our understanding of fundamental signal-regulated transcriptional responses that underlie mammalian development.

For these reasons, considerable interest exists in determining the molecular mechanisms by which liganded ERα regulates transcription. Previous studies have revealed that ERα functions primarily as a nuclear transcription factor, which dimerizes upon binding of its natural ligand, 17β-estradiol (E2), and acts as a potent regulator of gene expression. The actions of ERα on gene expression can be attributed to the direct or indirect ("tethered") binding of ERα at >10,000 sites in the genome, which may be located distally or proximally to the promoters of target genes (Carroll et al., 2005; Carroll et al., 2006; Welboren et al., 2009). The binding of ERα to the genome (1) promotes the recruitment of coregulators that mediate post-translational modification of histones or other transcription factors and (2) regulates the binding or activity of the RNA polymerase II (Pol II) transcriptional machinery, ultimately altering the transcriptome in estrogen-responsive cells (Acevedo and Kraus, 2004; Ruhl and Kraus, 2009).

To characterize the set of estrogen target genes, previous studies have used a variety of genomic technologies to determine steady-state gene expression patterns in the presence and absence of E2. These studies have failed to reveal a consistent view of the estrogen-regulated gene set. In particular, the use of expression microarrays has produced discrepancies in the numbers of estrogen-regulated genes in the widely used ERα-positive MCF-7 human breast cancer cell line, ranging from 100 to 1,500

(Cheung and Kraus, 2010; Kininis and Kraus, 2008). In addition, genomic ChIP analyses of ERα and Pol II have not produced a clear picture of the estrogen-regulated gene set either. This is due, in part, to the difficulty in assigning ERα binding events to specific gene regulatory outcomes (Carroll et al., 2006; Welboren et al., 2009). Another limitation of these analyses is that they have focused on the effects of estrogen signaling on Poll II transcription, without considering potential effects on Pol I and Pol III.

A fundamental weakness inherent in monitoring estrogen-dependent gene expression by assessing changes in mature mRNA is that longer treatments are required to allow time for mRNA accumulation (~3 to 24 hours). This time allows the accumulation of transcripts from primary ERα target genes, but also leads to a host of secondary transcriptional effects that are not directly mediated by ERα. In this regard, previous studies have found a wide array of genes encoding transcription factors are enriched in the estrogen-regulated gene set (Chang et al., 2006; Creighton et al., 2006; Frasor et al., 2003). To address these concerns, preliminary attempts to define the immediate transcriptional effects of estrogen signaling using the translation inhibitor cycloheximide indicated that only 20 to 30% of the genes showing changes in expression are primary targets (Lin et al., 2004), suggesting that the majority of genes that change expression in response to estrogen treatment are regulated indirectly. Inferring primary estrogen target genes based on cyclohexamide experiments, however, is problematic in three respects: (1) cyclohexamide is an extremely toxic compound, which has well-characterized non-specific effects on cells (Sidhu and Omiecinski, 1998), (2) cyclohexamide does not inhibit the effects of non-coding regulatory RNAs on gene expression, which is becoming widely recognized as an important mechanism underlying the regulation of many genes (Baek et al., 2008), and (3) the measure of steady-state mRNA depends not only on transcriptional regulation

by E2, but also on the rates of elongation, pre-mRNA processing, and mRNA degradation, which have been shown to play an important role in the regulation of certain mRNAs (Widelitz et al., 1987). Due to these factors, it is clear that a new approach is required to conclusively identify direct estrogen target genes. A data set defining the immediate effects of estrogen signaling on transcription by all three RNA polymerases would address many of these shortcomings and provide a model for rapid signal-dependent transcription in other systems.

Here, we used Global nuclear Run-On and Sequencing (GRO-seq) (Core et al., 2008) (Fig. 3.1) to identify the immediate effects of estrogen signaling on the entire transcriptome in MCF-7 cells. GRO-seq is a direct sequencing method that provides a "map" of the position and orientation of all engaged RNA polymerases across the genome at extremely high resolution. Several aspects of this method make it uniquely suited to serve as a general method for assessing changes in transcription caused by estrogen stimulation. First, GRO-seq is able to resolve instantaneous changes in the recruitment of RNA polymerases. This rapid temporal resolution can be used to identify the most immediate transcriptional effects of estrogen signaling. Second, GRO-seq is a direct measure of transcription. In contrast, methods that use RNA abundance as a measure of transcription, based either on microarray or deep sequencing technologies, measure a complex function of both RNA stability and processing. Third, unlike polymerase ChIP-seq, GRO-seq identifies the orientation of transcription, allowing the detection of antisense and divergent transcripts, which are clearly a significant fraction of the transcriptome and may have important regulatory roles. Fourth, GRO-seq allows the identification and characterization of transcripts lacking existing annotations in an unbiased way, which is also significant for detecting novel transcripts.

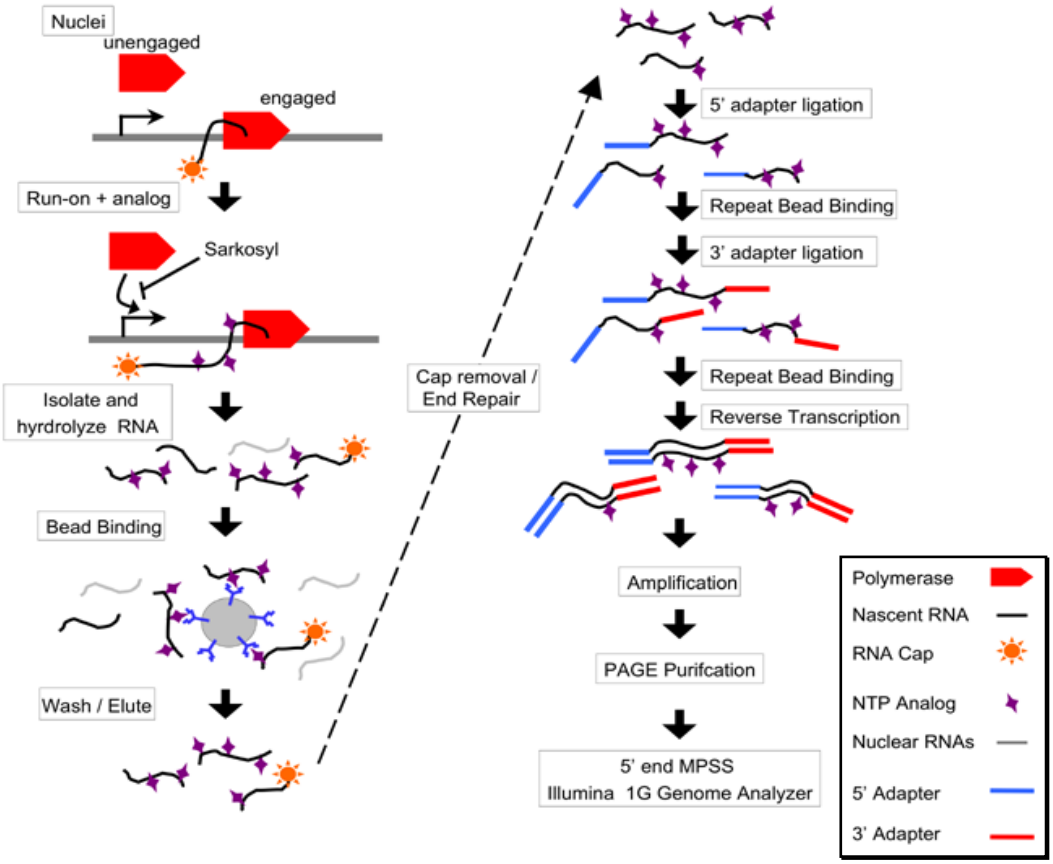


Figure 3.1. Schematic of the GRO-seq method (from Core et al., 2008).

Using GRO-seq in combination with a novel bioinformatic approach based on Hidden Markov Models (HMMs), we determined all (i.e., both annotated and unannotated) genomic regions in MCF-7 cells that are transcribed by Pols I, II, and III. In addition, we identified the primary transcriptional targets of E2 signaling by focusing on short treatments, prior to the activation of secondary targets (i.e., 0, 10, and 40 min.). Our unique approach has revealed a picture of estrogen-regulated gene expression that differs considerably from previous views based on expression microarray studies. We found that E2 signaling directly regulates a strikingly large fraction (~26%) of the MCF-7 transcriptome in a rapid, robust, and unexpectedly transient manner. In addition to protein coding transcripts, nearly every other class of transcript that has been described to date is also regulated by E2, including annotated microRNAs and non-coding RNAs (e.g., tRNAs, rRNAs), as well as a large number and variety of other unannotated, non-coding transcripts. Increased transcription of rRNA and tRNA genes by Pol I and Pol III, respectively, provided a means to accelerate the translation of newly synthesized protein-coding transcripts. Finally, comparisons of E2-regulated transcripts defined by GRO-seq to ERα binding sites defined by ChIP-seq suggests that many of the primary targets may be regulated by the direct actions of ERα at its binding sites, allowing us to categorize and understand a much larger fraction of the intergenic ERα binding sites. Collectively, the results presented here reveal many unexpected features of E2-regulation, providing the most comprehensive measurement of the primary and immediate effects of E2 signaling to date. Our results provide a model and resource for understanding rapid signal-dependent transcription in other systems.

3.3. Results and Discussion

3.3.1. Generation of GRO-seq libraries from estrogen-treated MCF-7 cells

To investigate the immediate effects of estrogen on the transcriptome of human cells, we treated estrogen-deprived ERα-positive MCF-7 human breast cancer cells with a short time course of 17β-estradiol (E2) (0, 10, 40, and 160 min) (Fig. 3.2A). MCF-7 cells are an excellent model for these analyses due to the plethora of available genomic data (Cheung and Kraus, 2010; Kininis and Kraus, 2008). The estrogen-deprived MCF-7 cells continued to grow actively (Fig. 3.2B) and the population of cells showed a normal distribution through the cell cycle (Fig. 3.2C).

Nuclei were isolated from two biological replicates of the E2-treated MCF-7 cells and subjected to the GRO-seq procedure to generate ~100 bp libraries representing nascent RNAs, which were sequenced using an Illumina Genome Analyzer (Fig. 3.2A). Short-reads were aligned to the human reference genome (hg18, NCBI36), including autosomes, the X-chromosome, and one complete copy of an rDNA repeat (GenBank ID: U13369.1). Approximately 13 to 17 million reads were uniquely mapped to the genome for each treatment condition and the biological replicates for each time point were highly correlated (average correlation coefficient = 0.98). GRO-seq returns data from all three RNA Polymerases (Pols I, II, and III). To validate if the reads mapping to the supposed loci transcribed by Pols I, II and III were correlated with the activities of each individual RNA polymerase, we carried out filter binding assays with combinations of polymerase inhibitors to isolate each polymerase. As expected, the activities detected by the filter binding assays were comparable to GRO-seq product fraction (Fig. 3.3).

Fig. 3.4 (top) shows a representative histogram of read counts versus genomic position for a locus containing the *LHX4* and *ACBD6* genes. Key features of the data

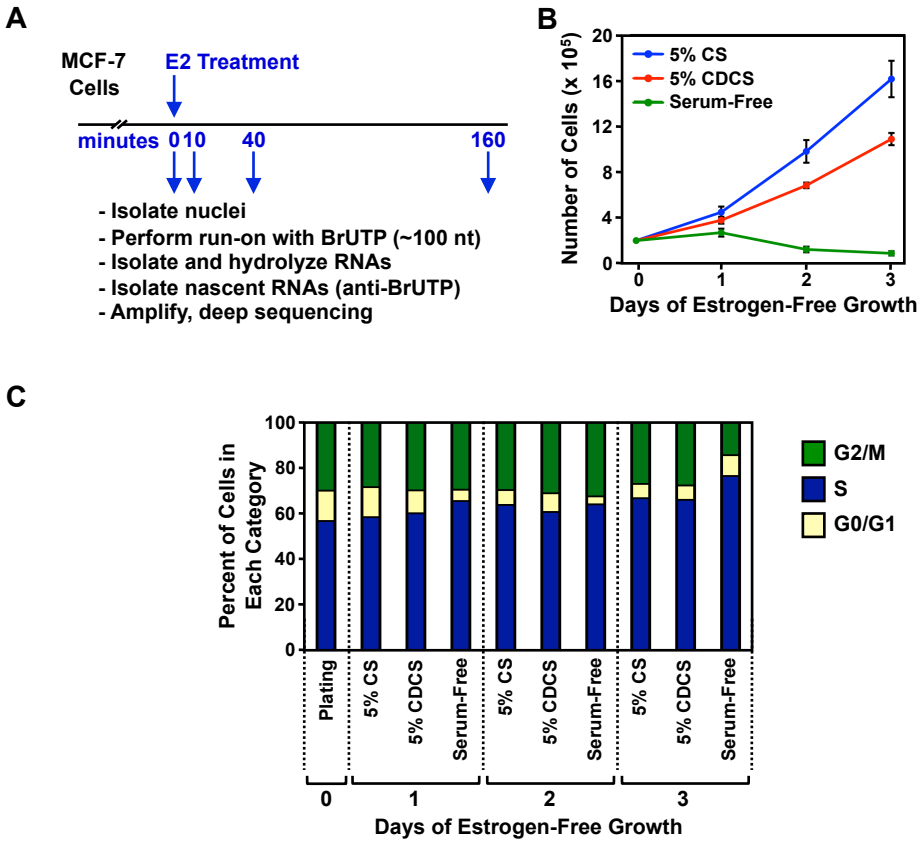


Figure 3.2. Experimental set up and conditions for GRO-seq analysis using MCF-7 cells. (A) Overview of the experimental set up for GRO-seq analysis using MCF-7 cells. **(B)** Proliferation of MCF-7 cells grown in medium with 5% calf serum (CS), 5% charcoal-dextran stripped CS (CDCS), or without serum. Cells were collected at the specified time point and counted. **(C)** Cell cycle analysis of MCF-7 cells grown in medium with 5% calf serum (CS), 5% charcoal-dextran stripped CS (CDCS), or without serum. The fraction of cells at each phase of the cell cycle was determined by flow cytometry.

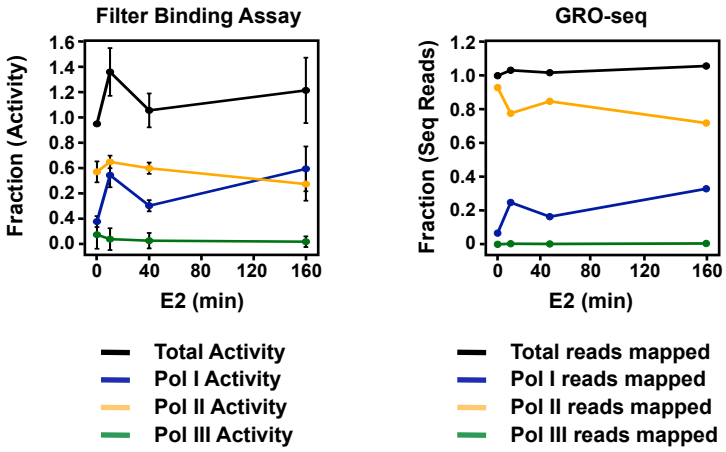


Figure 3.3. Comparison of the activities of RNA polymerases I, II, and III in MCF-7 cells. The activities of RNA polymerases I, II, and III measured by filter binding assays (left) and GRO-seq read counts (right). For the filter binding assays, the isolated nuclei were treated with (1) 1 μg/ml α-amanitin to block RNA Pol II or (2) 1 μg/ml α-amanitin plus 12 μM tagetin to block RNA Pols II and III. The nuclei were then subjected to run-on reactions in the presence of ^{32}P-CTP. The contribution of the different polymerases was calculated by subtraction from the total counts.

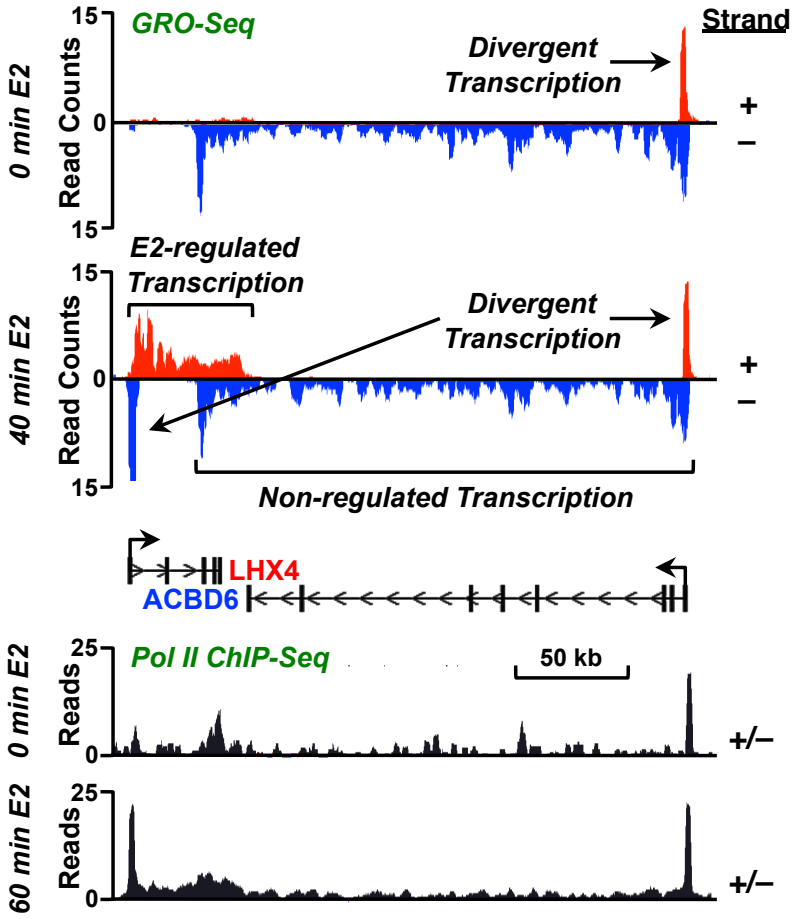


Figure 3.4. GRO-seq provides a detailed view of the E2-regulated transcriptome in MCF-7 cells. Genome browser view for a specific locus showing GRO-seq (top) and Pol II ChIP-seq (bottom) data illustrating the features of transcription and the effects of estrogen treatment.

set are illustrated in this representation, including strand-specific transcription, divergent transcription near transcription start sites (TSSs), and robust E2-dependentinduction for some genes (e.g., *LHX4*). These features are not readily apparent in ChIP-seq data from the same region (Fig. 3.4, bottom).

3.3.2. Unbiased assignment of GRO-seq reads to specific transcripts

To determine the effects of E2 on the entire transcriptome (i.e., annotated and unannotated; coding and non-coding), we developed an unbiased approach for calling transcripts using a two-state hidden Markov model (HMM). The model takes as input information about read counts across the genome and subsequently divides the genome into two states representing "transcribed" and "non-transcribed" regions (Fig. 3.5A, inset). In our approach, we combined sequence reads across all four time-points into a single read set. This combined set was then used to train the model and to construct a single set of transcripts that were expressed at one or more points during the E2 treatment time course. Importantly, this combined approach increased our power for detecting transcripts with low expression levels, allowing us to more accurately annotate a larger fraction of transcripts.

An example of the input and output of this algorithm for a gene-rich region of the genome is shown in Fig. 3.5A. The top panel shows the raw sequence read counts for the GRO-seq data, the middle panel shows the predicted transcripts, and the bottom panel shows the RefSeq annotations. To evaluate the robustness of our approach, we compared our predicted transcript calls to existing annotations when these were available. First, we determined whether our predictions reflect entire transcripts, as opposed to breaking each gene up into a series of smaller units. Then, we determined whether our approach can accurately identify non-transcribed intervals

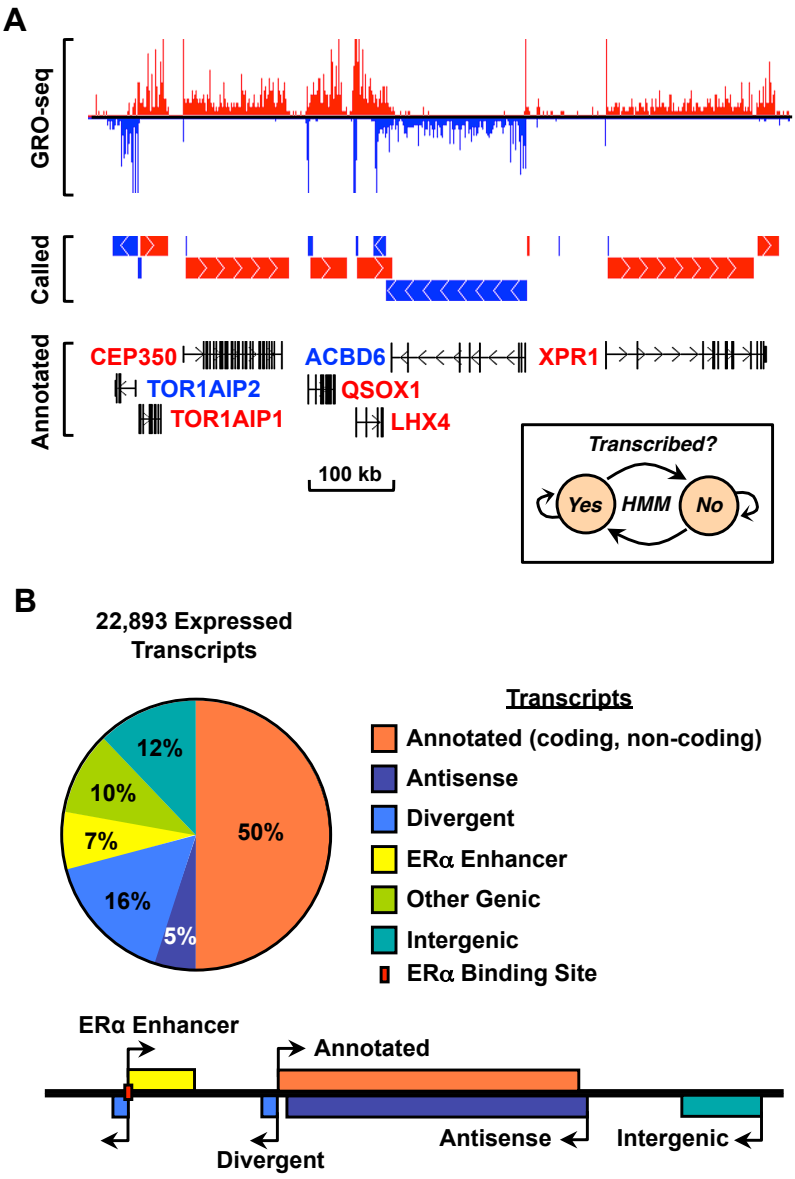


Figure 3.5. GRO-seq provides a detailed view of the E2-regulated transcriptome in MCF-7 cells. **(A)** De novo detection of transcripts using GRO-seq data (top) and an HMM (inset). Called transcripts (middle) match well to RefSeq annotations (bottom). **(B)** Classification of transcripts based on the annotation filter.

between neighboring, but distinct, gene annotations. We found that 90% of transcribed annotated genes overlap with exactly one transcript, and that 82% of called transcripts overlapping an annotated gene do so with exactly one annotation. Together, these results suggest that our HMM-based transcript calls largely recapitulate public annotations. In many cases our transcript calls provided new or more refined information about TSSs, 5' exons, and transcription termination sites than was available in existing databases. Using our algorithm, we assigned the genomic reads into 22,893 transcripts expressed in the MCF-7 genome at one or more points during the E2 treatment time course.

Transcripts called by the HMM were divided using a heuristic approach into six distinct, non-overlapping classes, which describe the best classification of each transcript given currently available annotations and other information (Fig. 3.6). The six classes of transcripts that we defined are illustrated in Fig. 3.5B and include: (1) annotated genic and non-coding RNA transcripts, (2) antisense (genic) transcripts, (3) divergent transcripts, (4) ERα enhancer transcripts, (5) other transcripts falling into annotated regions, but poorly matching the annotation, and (6) completely unannotated, intergenic transcription. Although each transcript is assigned to only one of these six classes, within each class multiple annotations could be applied, allowing the accurate annotation of miRNA genes that fall inside the introns of protein coding genes. We found that 50.1% of the called transcripts map to previously annotated genes or non-coding RNAs, 5.2% map to antisense transcripts, 16.4% map to divergent transcripts, 6.8% map to ERα binding enhancers, and 12.1% are entirely unannotated intergenic transcripts (Fig. 3.5B).

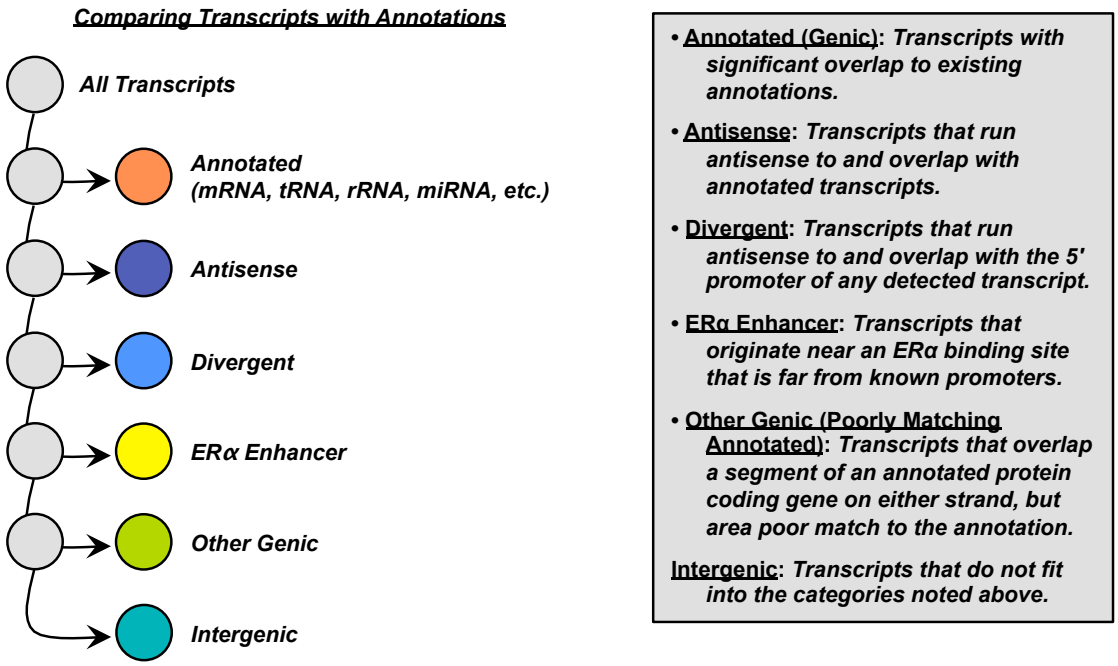


Figure 3.6. Summary and description of the heuristics used to define previously unannotated transcripts based on the most likely biological function.

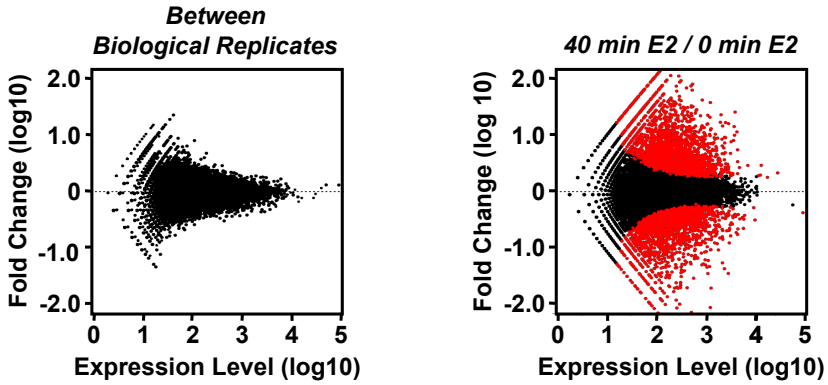


Figure 3.7. Determining E2 regulation of transcripts. Plots depicting the fold change of genes as a function of expression between two biological replicates (left) or between two treatment conditions (right). Red points indicate genes that fall outside of the expected variation are called "regulated' by E2 for the E2 treatment time point shown.

3.3.3. Extensive estrogen-dependent changes in the MCF-7 transcriptome

We determined which of the 22,893 transcripts change in response to E2 using a recently described model-based approach (Robinson et al., 2010). This approach fits a negative binomial model to read counts across all genes between our two biological replicates, and subsequently uses an exact test to identify genes whose change between two treatment conditions is beyond the global level of variation (Fig. 3.7). We chose to focus our analysis on a 12 kb window at the 5' end of each transcript, as we expect to observe changes during the first 10 minutes in this window that will not yet have spread to the 3' end of longer transcripts. Surprisingly, we found that transcription of an unexpectedly large fraction (~26%) of the MCF-7 transcriptome is altered (up- or down-regulated relative to the control/untreated condition) upon E2 treatment for at least one point in the time-course (Fig. 3.8A; comparisons are relative to the untreated condition). Large fractions of the genome are regulated even for the short treatments used in our experiments, strongly suggesting that these are direct actions of ERα. For example, at 10 minutes of E2 treatment, almost 10% of the MCF-7 cell transcriptome was significantly regulated at a false discovery rate of 0.1% (Fig. 3.8B). Another surprising finding concerns the dynamics of regulation for up and down-regulated transcripts. Through 40 minutes of E2 treatment, the time point at which the largest number of transcripts were regulated in our analyses, roughly equal numbers were up-regulated and down-regulated, but by 160 minutes ~75% of the transcripts were down-regulated (Fig. 3.8B). Those transcripts showing regulation at 10 or 40 minutes represent the most comprehensive and accurate definition of the immediate transcriptional targets of the estrogen signaling pathway described to date.

Next, we examined the regulation of the different classes of transcripts in more detail. Annotated protein coding and functional RNA transcripts as a group, as well as

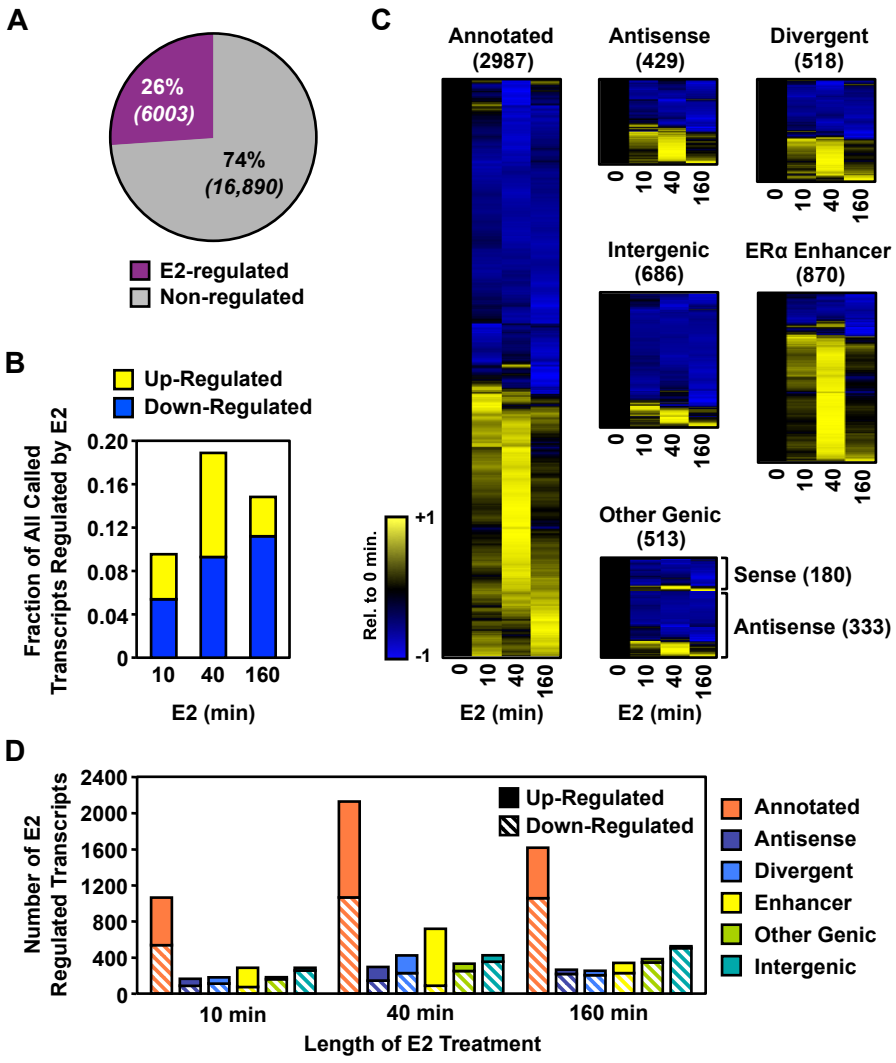


Figure 3.8. A large fraction of MCF-7 transcriptome is regulated by estrogen. **(A)** The fraction of all transcripts that are regulated by E2 at any time point. **(B)** The fraction of all transcripts that are up- or down-regulated by E2 at the time point shown. **(C)** Heatmap representations of time-dependent regulation by E2 for each transcripts class. Values are centered and scaled to the 0 minute time point. **(D)** The fraction of each class of transcripts that are up- or down-regulated by E2 at each time point.

those unannotated transcripts with possible roles in gene regulation (e.g., divergent and antisense), had approximately equal numbers of up-regulated and down-regulated transcripts at 40 minutes (Figs. 3.8C and 3.8D). In contrast, the ERα enhancer transcripts were predominately up-regulated, while the intergenic transcripts were predominantly down-regulated. Together, these results suggest a coordinated transcriptional response in which E2 signaling directs the transcriptional machinery from intergenic regions to those more critical to the estrogen response. In addition, they give a fundamentally different view of estrogen-regulated gene expression than has been obtained using expression microarrays, especially with respect to the timing, magnitude, and extent of regulation. In the following sections, we highlight some key aspects of our data related to non-coding transcripts, protein coding transcripts, microRNAs, components of the protein biosynthetic machinery, and ERα enhancers.

3.3.4. Regulation of unannotated non-coding transcripts by estrogen: divergent, antisense, and intergenic transcripts

As noted above, our GRO-seq data revealed extensive estrogen regulation of a large set of unannotated non-coding transcripts, including divergent, antisense, and intergenic transcripts. Although the functions of these transcripts are largely unknown, their regulation by E2 suggests a role in estrogen-dependent transcriptional responses.

The production and accumulation of divergent transcripts was first documented in recent studies using high-throughput genome-wide sequencing approaches with human fibroblasts (Core et al., 2008) and mouse embryonic stem cells (Seila et al., 2008). Divergent transcripts are transcribed in the opposite direction from primary transcripts at the promoters of actively transcribed genes. The function of divergent transcripts is unknown, but their production has been suggested to promote an open

chromatin architecture at promoters through the generation of a nucleosome-free region or negative superhelical tension (Core et al., 2008; Seila et al., 2008; Seila et al., 2009). We identified 518 divergent transcripts associated with the promoters of protein coding genes and other unannotated transcripts that are regulated by E2 for at least one time point (FDR q-value < 0.001). Using these annotations, we tested whether production of a given E2-regulated divergent transcript correlates with the synthesis of the corresponding primary transcript. To do so, we tested 844 primary/divergent transcript pairs for which either the divergent, primary, or both transcripts, were regulated by E2 for at least one time point. As shown in Fig. 3.9 (left), E2-dependent changes in divergent transcription were strongly correlated with E2-dependent changes in the corresponding primary transcripts (Pearson correlation: 0.744; $p < 2.2 \times 10^{-16}$). This result is consistent with a role for divergent transcription in facilitating E2-dependent transcription of the corresponding primary transcript.

Although not well characterized, antisense transcription has been shown to have roles in degradation of corresponding sense transcripts (Katayama et al., 2005; Werner et al., 2009), as well as gene silencing at the chromatin level (Liu et al., 2010; Morris et al., 2008). Of 1,197 transcripts annotated as antisense to a protein coding transcript, we identified 429 that are regulated by E2 (FDR q-value < 0.001). As with the divergent transcripts, we determined if production of a given E2-regulated antisense transcript correlates with the synthesis of the corresponding primary transcript. Based on 582 sense/antisense transcript pairs, we found a remarkably high correlation between genes and their antisense transcripts (Pearson correlation: 0.654; $p < 2.2 \times 10^{-16}$) (Fig. 3.9, right). This is particularly surprising given that, unlike divergent transcripts, antisense transcripts do not share a proximal promoter with the sense transcript. If antisense transcripts play a role in the degradation of the sense transcript, as has been suggested previously, then their E2-dependent production may

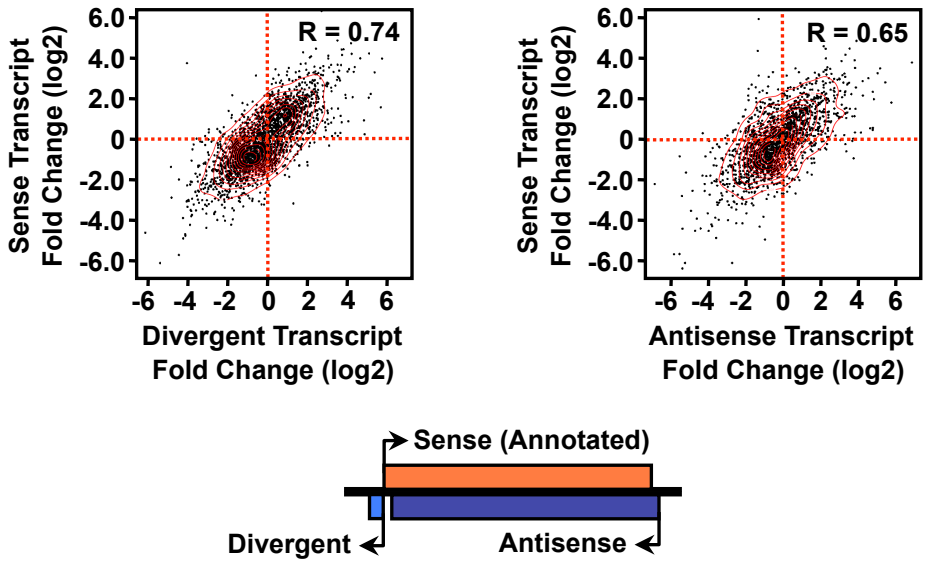


Figure 3.9. Determining correlation of sense transcripts with divergent and antisense transcripts. Scatter plots showing the correlation between primary transcripts and antisense transcripts (left) and between primary transcripts and divergent transcripts (right). Density isochrones are shown as red lines.

provide a "built in" means of attenuating the steady-state levels of a select set of estrogen-regulated transcripts. Alternatively, their production could also potentially attenuate expression at the level of primary transcription.

We also identified 2,761 transcripts that have no specific relation to previous genome annotations. Of these, 686 were regulated by E2 for at least one time point. Interestingly, the vast majority of these E2-regulated intergenic transcripts are down regulated by E2 treatment (Fig. 3.8D). The function of these transcripts is unknown. Some may represent currently unannotated protein coding transcripts or functional RNAs. Ascribing a function to these RNAs and determining their relative stability inthe steady-state cellular RNA pool will require additional studies. Their down-regulation by E2, however, suggests a link to the estrogen signaling program. Perhaps they act to antagonize E2-dependent transcriptional responses and must be shut down to achieve a full estrogen response. Alternatively, their antagonism by E2 may be a passive effect of RNA polymerases being diverted to bona fide transcriptional targets of the estrogen signaling pathway, as suggested previously (Carroll et al., 2006).

3.3.5. Rapid, extensive, and transient regulation of protein coding transcripts by estrogen

Numerous studies have examined the steady-state regulation of protein coding transcripts by E2 using expression microarrays. Given the sensitivity of our approach for detecting immediate transcriptional changes in response to short E2 treatments, we extracted and examined the protein coding transcripts in our GRO-seq data for comparison. We focused on annotations in the RefSeq database, because this set is among the most comprehensive collection of transcripts, and has extensive and well-documented overlap with expression microarrays. As noted above, we used read

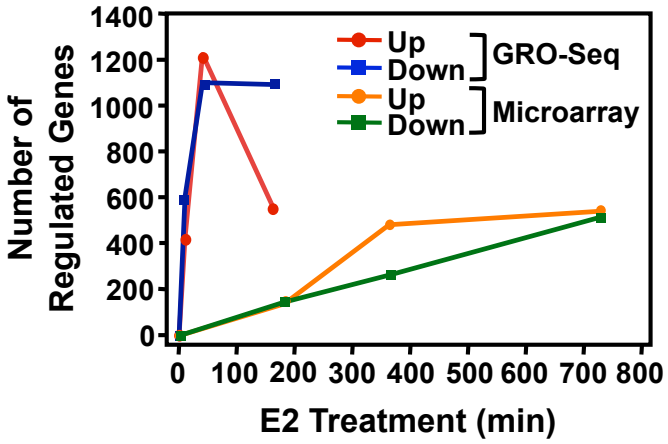


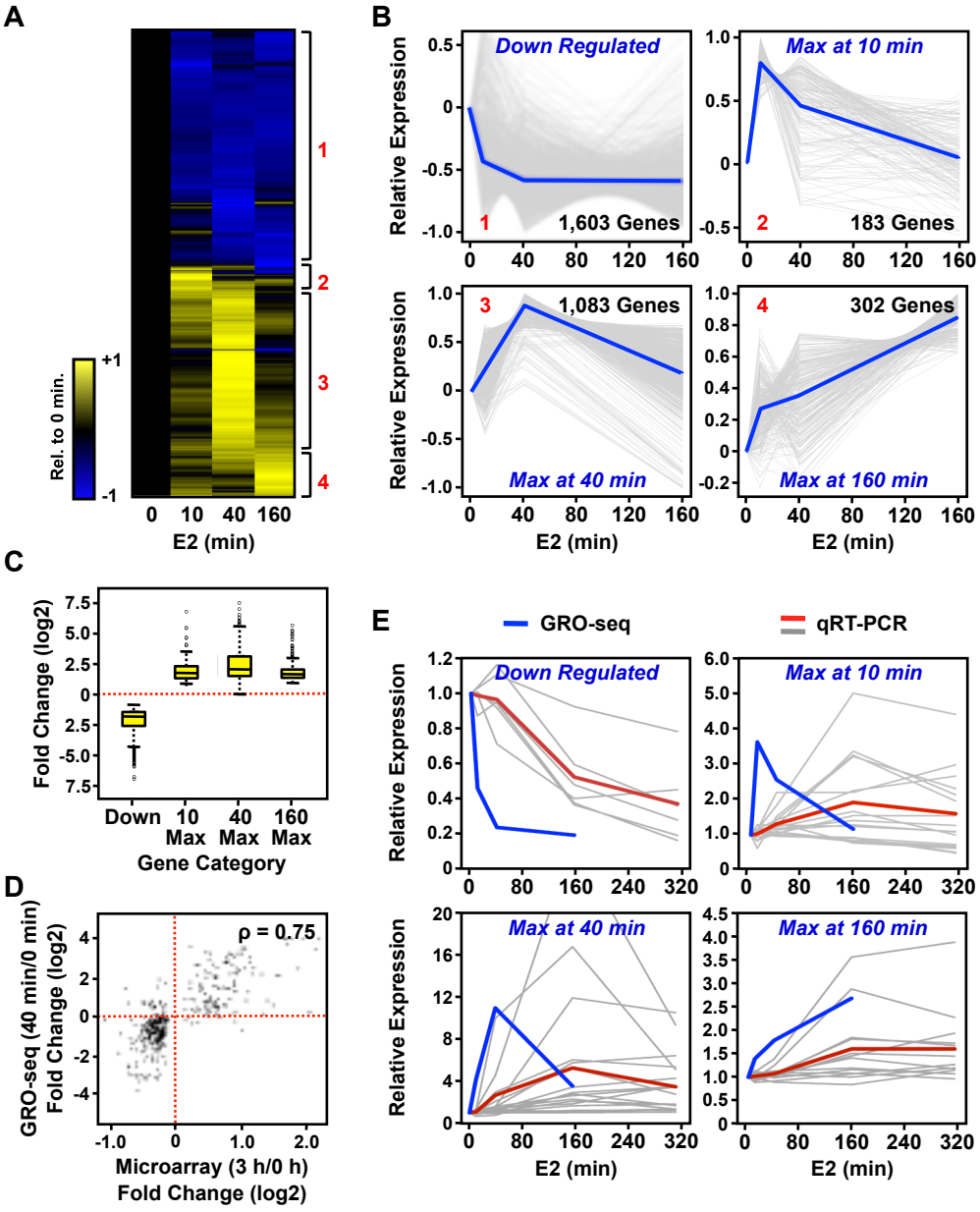
Figure 3. 10. Analysis of E2-dependent regulation of RefSeq genes and comparison with other measures of gene expression. The number of RefSeq genes that are detected as regulated by GRO-seq, as described herein, or by gene expression microarrays (fold change > 2 or < 0.6; p-value < 0.05; Carroll et al., 2006) following a time course of E2 treatment.

counts in a 12 kb window at the 5' end of each annotation and determined regulation by E2 using the edgeR package, filtering for a false discovery rate of 0.1%.

Using this approach, we detected a total of 3,098 protein coding transcripts whose levels changed relative to the control (untreated) condition at one or more of the points in the E2 treatment time course. In total, this represents ~15% of all genes annotated in RefSeq (~33% of 9,337 expressed genes) that are responsive to E2. This is a considerably larger number of genes than was detected previously at 1 or 3 hours of E2 treatment using expression microarrays (Cheung and Kraus, 2010; Kininis and Kraus, 2008; Fig. 3.10). Surprisingly, we found ~1000 genes total to be up- or down-regulated after only 10 minutes of E2 treatment. We used hierarchical clustering to define four classes of genes sharing similar patterns of regulation, including a class of rapidly down-regulated genes and three classes of genes with maximal transcription at the three E2 treatment time points (10, 40, or 160 minutes) (Fig. 3.11, A and B). The down-regulated class was the largest, comprising ~50% of the E2-regulated protein coding transcripts. The majority of genes in this class was rapidly down-regulated (by 10 minutes, on average) and tended (with a few exceptions) to stay down regulated throughout the time course. Up-regulated genes with maximal transcription at 40 minutes were the second largest class, comprising ~34% of the E2-regulated protein coding transcripts. Although the time course of induction or repression varied among the four classes, the magnitude of response did not differ between the classes (Fig. 3.11C). Interestingly, the genes in the "10 minute max" and "40 minute max" classes returned, on average, to the basal levels of transcription by the end of the E2 treatment time course (Fig. 3.11B), highlighting the rapid and transient nature of the transcriptional response for the majority of the up-regulated genes.

Biologically relevant changes in transcription should be accompanied, in most cases, by similar changes in the steady-state level of the corresponding mRNA. We

Figure 3.11. GRO-seq identifies four distinct classes of E2-regulated RefSeq genes. **(A)** Heatmap of the time course of E2-dependent regulation of RefSeq genes. Red numbers indicate the four different classes of regulation. **(B)** Centered-scaled traces showing the regulation of the four distinct classes of E2 regulation. Grey lines represent GRO-seq data for individual genes and blue lines represent the mean of the individual traces. **(C)** Box-plot showing the E2-dependent fold change for genes in each of the four classes. **(D)** Correlation between fold changes measured by GRO-seq and expression microarrays for genes that show a change in the microarray analyses. **(E)** Comparison of GRO-seq data to mRNA expression measured by RT-qPCR. Blue lines represent the mean of the GRO-seq data for the genes analyzed. Grey lines represent RT-qPCR data for individual genes and red lines indicate the mean.


A
Rel. to 0 min.
+1
-1
0 10 40 160
E2 (min)
1
2
3
4
B
Down Regulated
1
1,603 Genes
Max at 10 min
2
183 Genes
3
1,083 Genes
Max at 40 min
4
302 Genes
Max at 160 min
Relative Expression
E2 (min)
C
Fold Change (log2)
Down
10 Max
40 Max
160 Max
Gene Category
D
ρ = 0.75
GRO-seq (40 min/0 min) Fold Change (log2)
Microarray (3 h/0 h) Fold Change (log2)
E
GRO-seq
qRT-PCR
Down Regulated
Max at 10 min
Max at 40 min
Max at 160 min
Relative Expression
E2 (min)

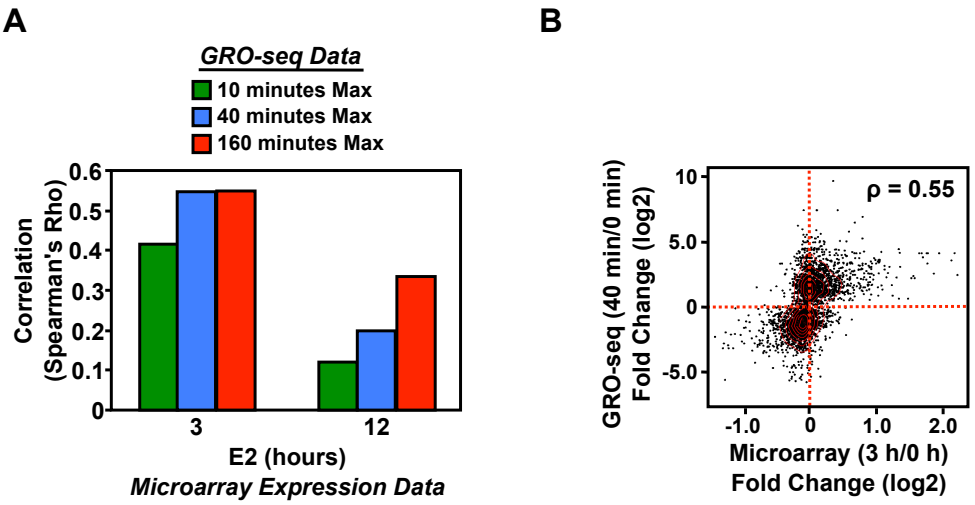


Figure 3.12. Analysis of E2-dependent regulation of RefSeq genes and comparison with other measures of gene expression. (A) Spearman's rank correlation between E2-regulated RefSeq genes detected by GRO-seq at each time point indicated and their regulation as detected by expression microarrays at 3 or 12 hours of E2 treatment. **(B)** Spearman's rank correlation of fold changes for RefSeq genes that are regulated in the GRO-seq data set compared to gene expression microarrays at the time points indicated.

tested this expectation using both genomic and gene-specific comparisons. First, we compared fold-changes in primary transcription detected using our GRO-seq data to fold-changes at the level of steady-state mRNA (3 or 12 hours of E2 treatment) from published expression microarray data for MCF-7 cells. For the subset of genes that we found to be regulated by GRO-seq, we find that the strongest correlations were between either the 40 or 160 minute GRO-seq time points with the 3 hour microarray time point (Fig. 3.12, A and B). Interestingly, if we limit the analysis to only genes that change in the microarray analysis (FDR corrected q-value < 0.05), we see an even higher correlation between GRO-seq and microarray data (Fig. 3.11D; Spearman's correlation: 0.75). This analysis suggests that the early actions of E2 are almost all mediated at the level of transcription, and that E2 does not affect RNA stability or degradation rate directly. These results provide a first indication that transcription, as determined by GRO-seq, is indeed robustly propagated to changes in the steady-state levels of the corresponding mRNAs.

Next, we randomly selected a set of 10 to 20 genes for each of the four classes (54 genes total) and measured the relative steady-state levels of mRNA from each gene over a 6 hour time course of E2 treatment using RT-qPCR. In general, the changes in transcription measured by GRO-seq were reflected in corresponding changes in the steady-state mRNA levels measured by RT-qPCR (Fig. 3.11E; Fig. 3.13, A – D). In almost all cases, we observed a delay of approximately 1 to 3 hours between the peak fold changes measured by GRO-seq and RT-qPCR. This delay reflects the time necessary for changes in Pol II (measured at the 5' end in GRO-seq) to reach the 3' end of the gene and for mRNA to accumulate (or degrade) by a detectable level. As with the comparisons to the microarray expression data, these results indicate that changes in transcription are efficiently translated into changes in the steady-state levels of the corresponding mRNAs. The correspondence was

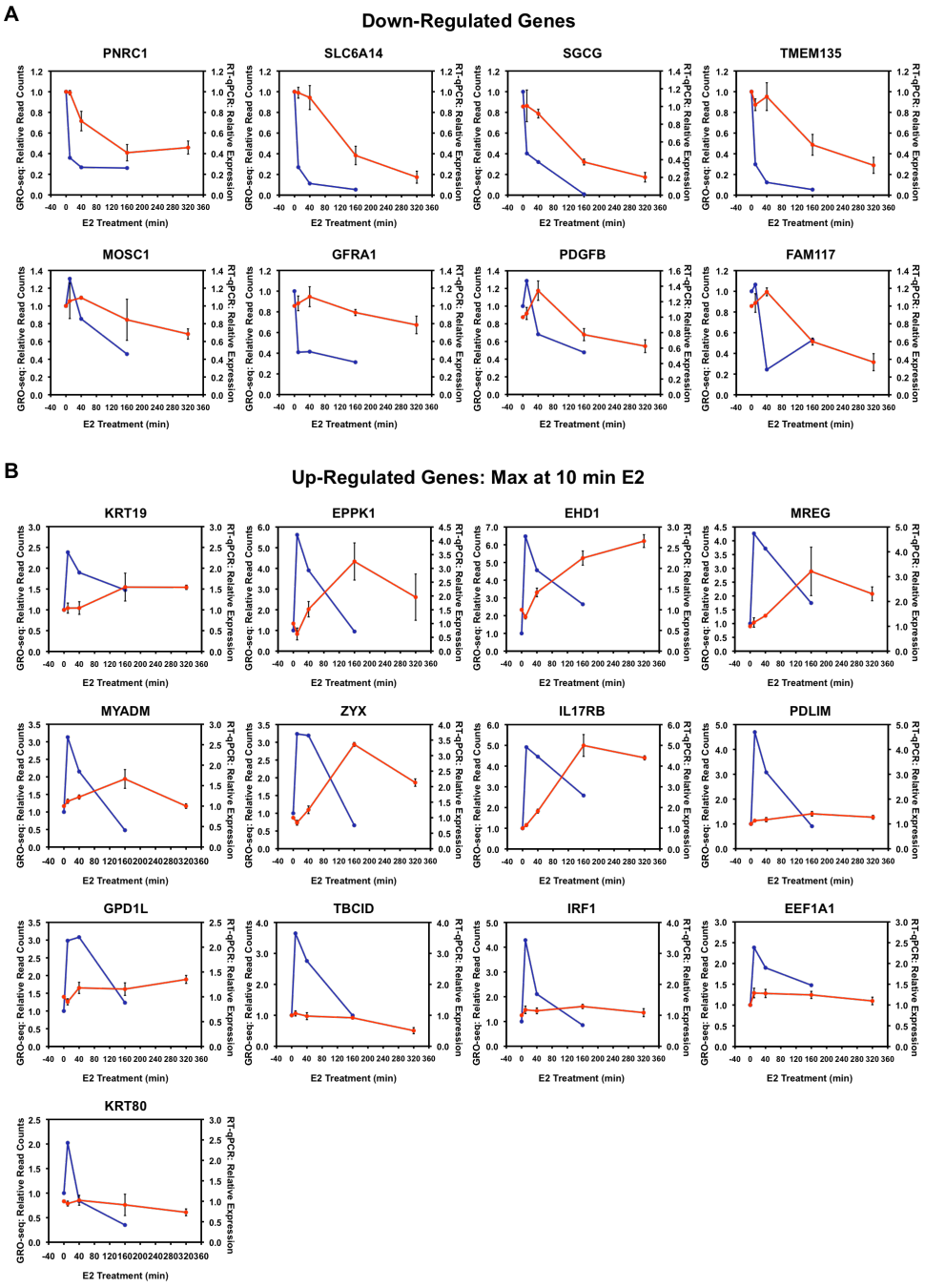


Figure 3.13. Analysis of E2-dependent regulation of RefSeq genes and comparison with other measures of gene expression.

(A-D) Comparison of RefSeq gene expression levels by GRO-seq (blue) or steady-state mRNA measured by RT-qPCR (red) for individual genes from the four different E2-regulated classes.

Figure 3.13. (*Continued*)

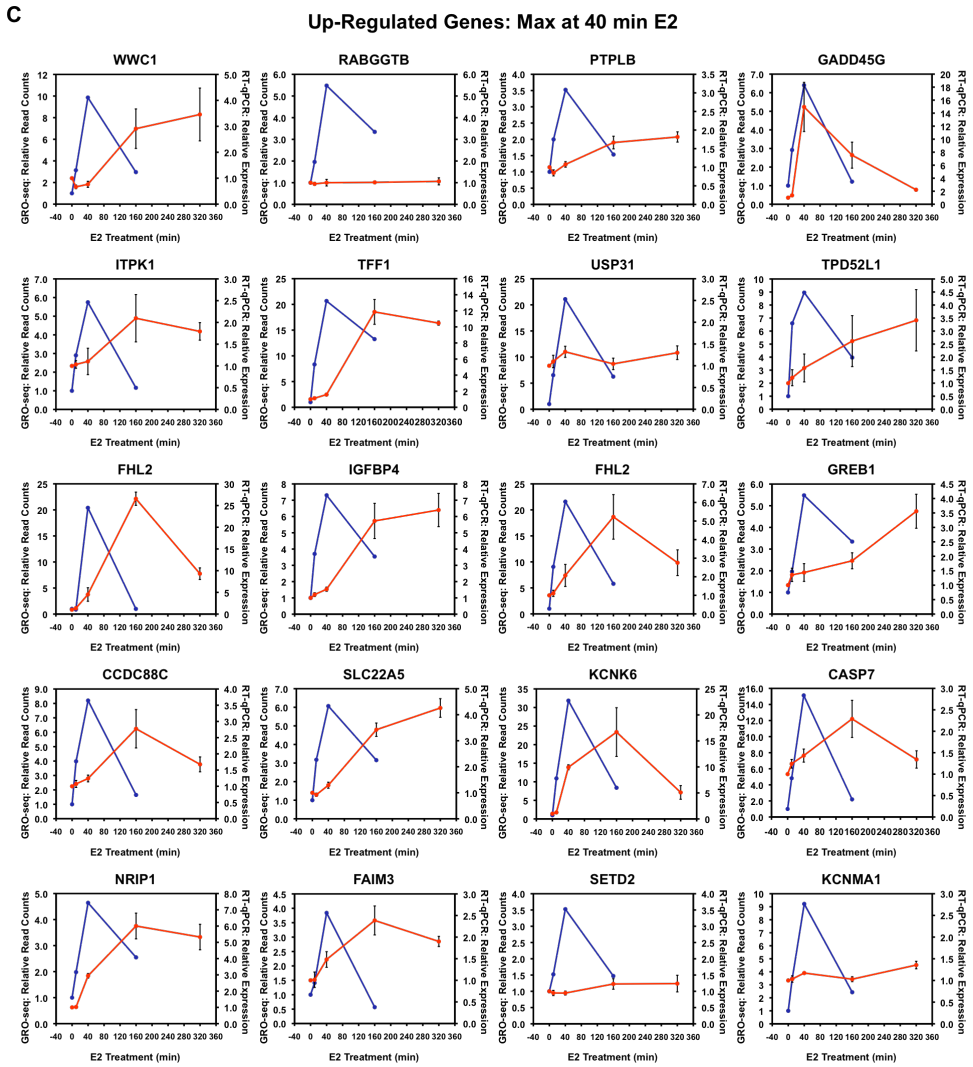


Figure 3.13. (*Continued*)

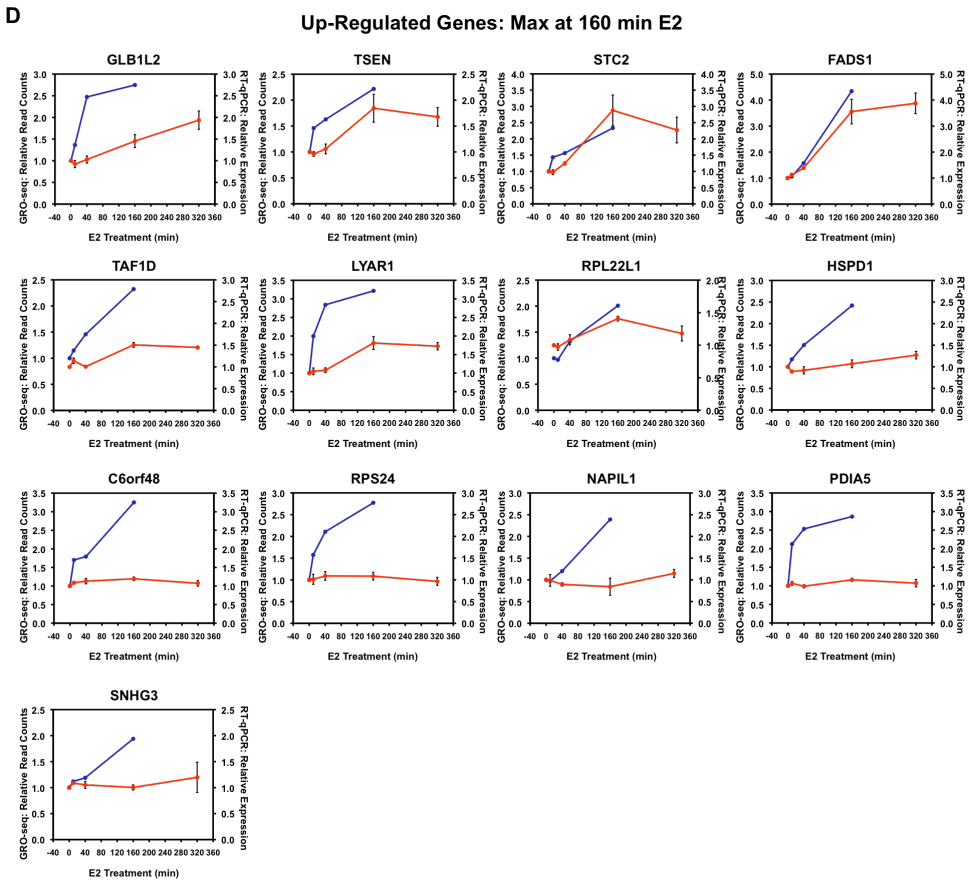


Table 3.1. GO analysis of the four classes of E2-regulated protein coding genes identified by GRO-seq[a].

A. Down-regulated

Accession	GO term	p-value
GO:0043170:	macromolecule metabolic process	9.63E-05
GO:0006396:	RNA processing	9.63E-05
GO:0048731:	system development	9.63E-05
GO:0006820:	anion transport	1.98E-04
GO:0006412:	translation	1.01E-03
GO:0015698:	inorganic anion transport	1.01E-03
GO:0010467:	gene expression	1.13E-03
GO:0048513:	organ development	1.13E-03
GO:0006139:	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.77E-03
GO:0006811:	ion transport	3.77E-03
GO:0007166:	cell surface receptor linked signal transduction	4.05E-03
GO:0007399:	nervous system development	6.23E-03
GO:0019226:	transmission of nerve impulse	6.95E-03
GO:0006281:	DNA repair	8.98E-03
GO:0007186:	G-protein coupled receptor protein signaling pathway	8.98E-03

B. Max at 10 minutes E2

Accession	GO term	p-value
GO:0048731:	system development	4.17E-04
GO:0048513:	organ development	2.53E-03
GO:0046883:	regulation of hormone secretion	5.14E-02

C. Max at 40 minutes E2

Accession	GO term	p-value
GO:0048731:	system development	3.48E-12
GO:0048513:	organ development	1.98E-09
GO:0006139:	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1.01E-08
GO:0043170:	macromolecule metabolic process	4.02E-08
GO:0010467:	gene expression	5.02E-08
GO:0043283:	biopolymer metabolic process	1.27E-06
GO:0015031:	protein transport	5.19E-06
GO:0007186:	G-protein coupled receptor protein signaling pathway	1.74E-05
GO:0007166:	cell surface receptor linked signal transduction	3.80E-05
GO:0045184:	establishment of protein localization	3.80E-05
GO:0006928:	cell motility	6.66E-05
GO:0016070:	RNA metabolic process	2.96E-04

Table 3.1. (*Continued*)

GO:0006396:	RNA processing	4.17E-04
GO:0006811:	ion transport	5.55E-04
GO:0007399:	nervous system development	1.49E-03
GO:0016477:	cell migration	1.49E-03
GO:0007266:	Rho protein signal transduction	2.61E-03
GO:0030001:	metal ion transport	2.91E-03
GO:0006886:	intracellular protein transport	4.46E-03
GO:0009888:	tissue development	4.46E-03
GO:0035023:	regulation of Rho protein signal transduction	4.82E-03
GO:0035295:	tube development	6.10E-03
GO:0001501:	skeletal development	6.34E-03
GO:0007600:	sensory perception	7.21E-03
GO:0006412:	translation	8.39E-03
GO:0048519:	negative regulation of biological process	8.54E-03
GO:0030308:	negative regulation of cell growth	8.81E-03
GO:0045321:	leukocyte activation	9.26E-03
GO:0007169:	transmembrane receptor protein tyrosine kinase signaling pathway	9.31E-03

D. Max at 160 minutes E2

Accession	**GO term**	**p-value**
GO:0044249:	cellular biosynthetic process	1.27E-10
GO:0006412:	translation	1.40E-03
GO:0009059:	macromolecule biosynthetic process	1.38E-02
GO:0022613:	ribonucleoprotein complex biogenesis and assembly	2.93E-02
GO:0044260:	cellular macromolecule metabolic process	3.47E-02
GO:0019538:	protein metabolic process	3.53E-02

---

[a] The four classes of genes are (A) down-regulated, (B) max at 10 minutes E2, (C) max at 40 minutes E2, and (D) max at 160 minutes E2.

Methods: Gene ontology analyses were performed using GoStat (http://gostat.wehi.edu.au/) (Beissbarth and Speed, 2004). All expressed genes were used as a background set to analyze GO term for each class ($p < 0.05$).

the strongest for the down-regulated and the 40 minute max GRO-seq classes (> 80% of genes assayed showed corresponding changes) and weaker for the 10 minute max and 160 minute max classes (~50% of genes assayed showed corresponding changes). The discrepancies between transcription and steady-state mRNA levels may be due to inherent instability of certain nascent transcripts, which prevents them from generating mature transcripts. Alternatively, it may reflect active post-transcriptional regulation of specific transcripts (e.g., by miRNAs; see below).

Gene ontology (GO) analyses of the four classes of genes revealed a similar pattern of enrichment in gene ontological categories for the down-regulated and 40 minute max classes (Table 3.1, A and C). Specifically, there was a significant enrichment in GO terms related to transcription, nucleic acid metabolism, cell surface receptor and G protein-coupled signaling. A priori, protein coding genes with these functions are the types of genes that one might expect to be regulated in an early transcriptional response. The fact that the same GO terms (but different genes) are regulated in both the major up- and down-regulated classes suggests a switch from one cellular signaling program (e.g., serum response) to another (i.e., estrogen signaling); each pathway may require the same functional categories of genes, but use a distinct set of genes within each category. Interestingly, the 160 minute max class was significantly enriched in GO terms related to ribosome biogenesis, translation, and protein synthesis (discussed and elaborated below) (Table 3.1D), whereas a very modest enrichment of GO terms was observed for the 10 minute max class (Table 3.1B).

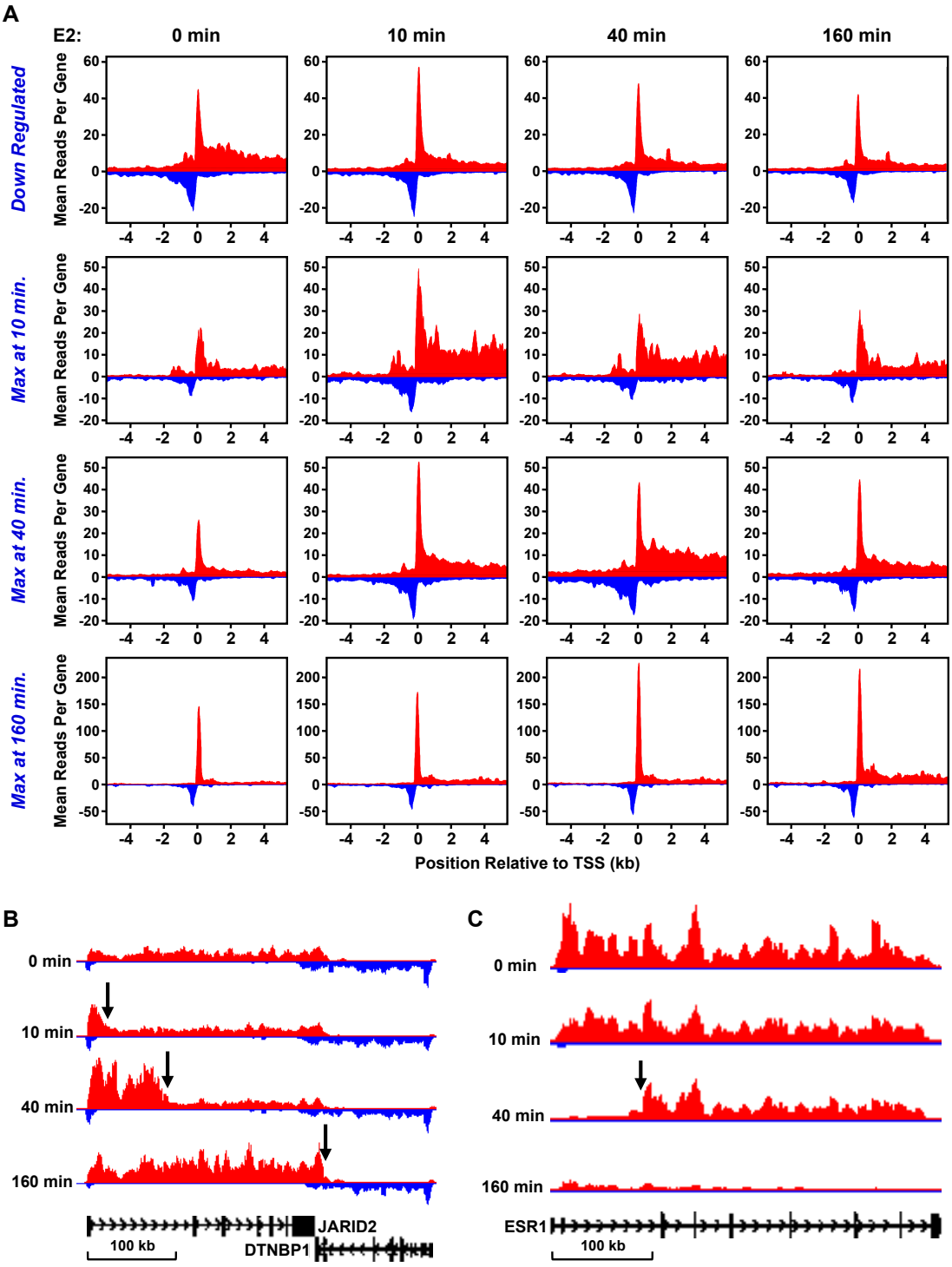
Together, our results show that the transcriptional response to estrogen signaling for protein coding genes (and other classes of transcripts, as well; see below) is rapid, extensive, and transient. This represents a different view of the estrogen response than has been provided by microarray expression studies, which have

suggested a continually increasing set of regulated genes in response to E2 treatment, many of which are likely to be secondary or tertiary effects (Fig. 3.10). In this regard, our GRO-seq results have identified a set of genes whose transcriptional responses were turned on and off before the first E2 treatment time point examined in the vast majority of expression microarray studies. Interestingly, the transcriptional response for more than half of the protein coding genes was a rapid down-regulation. This may be due to active repression mechanisms mediated by liganded ERα. Alternatively, this may reflect a passive mechanism where the Pol II transcription machinery is drawn from the down-regulated genes to those that are actively up-regulated in response to E2 treatment.

3.3.6. Pol II dynamics in response to E2

Since the transcriptional response for protein coding genes to estrogen signaling was rapid and transient, we explored the dynamics of Pol II at the promoters of the four classes defined in hierarchical clustering analysis. We performed metagene analyses across the promoter regions of each class from -4 kb to +4 kb for each treatment time point (Fig. 3.14A). The peak of reads in the immediate vicinity of TSS indicates the presence, on average, of engaged Pol II before and after E2 treatment. The decrease (or increase) of reads in the downstream region indicates the down-regulation (or up-regulation) of transcription in response to E2. This presentation of the GRO-seq results highlights the following: on average (1) loading of Pol II at the TSSs of up-regulated genes increases in response to E2 treatment, (2) divergent transcription of the up-regulated genes increases in response to E2 treatment, (3) down-regulation affects primarily Pol II in the gene bodies, and (4) loading of Pol II at the TSSs and divergent transcription largely follow the Pol II response in the body of the gene.

Figure 3.14. GRO-seq reveals the dynamics of E2-dependent transcription. (A) Metagene representations showing the average profile of GRO-seq sequence reads near and at the TSSs of RefSeq genes in each of the four classes during the E2 treatment time course. **(B and C)** Gene-specific views of the leading (B) and lagging (C) edges of a Pol II "wave" shown for the up-regulated gene *JARID2* (B) and the down-regulated gene *ESR1* (C), respectively, during the E2 treatment time course.


A
E2:
0 min
10 min
40 min
160 min
Down Regulated
Max at 10 min.
Max at 40 min.
Max at 160 min.
Mean Reads Per Gene
Position Relative to TSS (kb)
B
0 min
10 min
40 min
160 min
JARID2
DTNBP1
100 kb
C
0 min
10 min
40 min
160 min
ESR1
100 kb

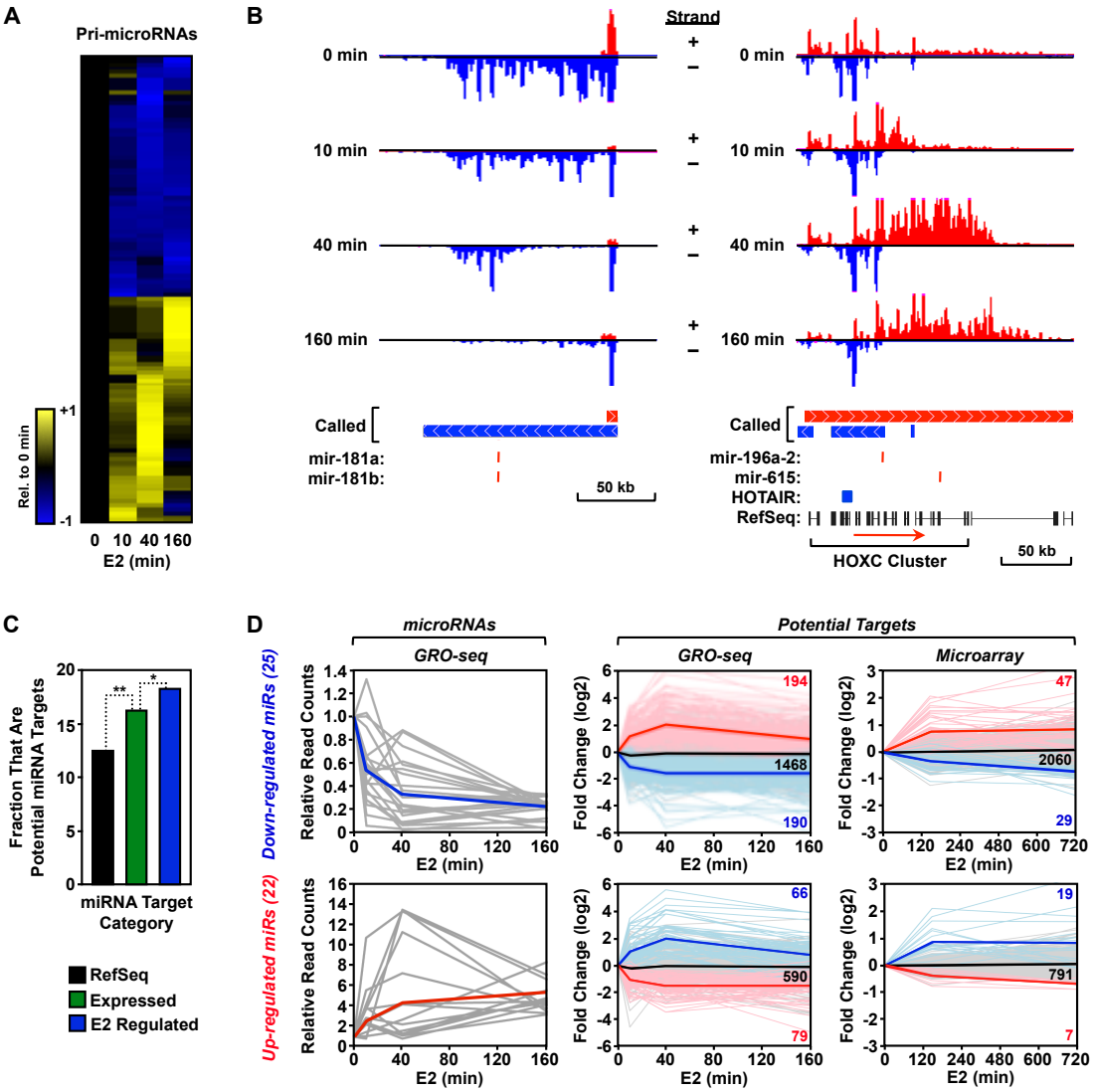
The increase in Pol II loading at the TSS in response to E2 suggests that Pol II loads more rapidly than it escapes into the body of the gene for these classes of E2-regulated genes. This is especially evident between the 10 and 40 minute time points for the 40 minute max genes and between the 40 and 160 minutes time points for the 160 minute max genes, where we see increased Pol II loading at the earlier time point followed by an appreciable increase in Pol II in the body of the gene at the later time point. This "delayed" pattern of loading and escape is perhaps unexpected for the 160 minute max genes, since the pausing of Pol II in the promoter proximal region is thought to allow rapid activation of transcription in response to cellular signaling (Lis, 1998). Alternatively, such a response fits well with a recent suggestion that pausing of Pol II in the promoter proximal region allows synchronous gene activation (Boettiger and Levine, 2009).

The dynamics of Pol II can also be clearly observed in examples from specific up- and down-regulated genes. With E2 up-regulated genes, the leading edge of a Pol II wave was observed traveling into the gene body upon E2 treatment (Fig. 3.14B). In contrast, with E2 down-regulated genes, the lagging edge of a Pol II wave was observed as the polymerases were cleared from the TSS (Fig. 3.14C). The results from our GRO-seq analysis have provided an unprecedented view of the Pol II dynamics in response to a sustained signal.

3.3.7. Regulation of miRNA gene transcription by estrogen: parallels to the regulation of protein coding genes

Our GRO-seq approach also provides considerable information regarding the transcriptional regulation of primary microRNA transcripts. Micro RNAs (miRNAs) are ~22 nt non-coding regulatory RNAs that mediate post-transcriptional regulation of gene expression by inhibiting the translation or promoting the degradation of target

Figure 3.15. E2 regulates the transcription of primary miRNA genes. (A) Heatmap of the time course of E2-dependent regulation of primary miRNA transcripts. **(B)** Gene-specific examples of down-regulated (left) and up-regulated (right) primary miRNA genes. Called transcripts and annotations are shown. **(C)** Fraction of the specified subset of annotated genes that are predicted to be targets of an E2-regulated miRNA based on TargetScan. **(D)** ***Right panels***, GRO-seq data for pri-miRNA transcripts that are up-regulated (top) or down-regulated) ≥ 3-fold by E2. Grey lines = data for individual genes. Blue lines = average for all genes. ***Middle and left panels***, GRO-seq (middle) and expression microarray (right) data for all of the potential targets of miRNAs encoded by the pri-miRNA transcripts shown in the left panels. Faded red, black, and blue lines = data for individual up-regulated, unregulated, and down-regulated genes, respectively (the counts for each type are listed). Bold red, black, and blue lines = averages for all up-regulated, unregulated, and down-regulated genes, respectively.


A
Pri-microRNAs
Rel. to 0 min
+1
-1
0 10 40 160
E2 (min)
B
Strand
0 min
10 min
40 min
160 min
Called
mir-181a:
mir-181b:
50 kb
mir-196a-2:
mir-615:
HOTAIR:
RefSeq:
HOXC Cluster
C
Fraction That Are Potential miRNA Targets
miRNA Target Category
RefSeq
Expressed
E2 Regulated
D
microRNAs
Potential Targets
GRO-seq
Microarray
Down-regulated miRs (25)
Up-regulated miRs (22)
Relative Read Counts
Fold Change (log2)
E2 (min)
194
1468
190
47
2060
29
66
590
79
19
791
7

mRNAs. miRNA precursor transcripts (pri-miRNAs) are generated by Pol II, or in some cases Pol III, either as part of a "host" gene in which they are embedded or from an intergenic region using their own promoter (Krol et al., 2010). Using our GRO-seq data set, we explored the regulation of pri-miRNA gene transcription by E2. We unambiguously identified 322 expressed miRNA-containing transcripts in our data set based on miRBase ver. 14. Of these, 119 (~37%) were regulated by E2 during at least one time point (FDR q-value < 0.001). Regulated pri-miRNAs included some previously published estrogen-regulated miRNAs, including mir-181a, mir-181b and mir-21 (Bhat-Nakshatri et al., 2009; Maillot et al., 2009). Overall, the pattern of regulation depicted in the heatmap shown in Fig. 3.15A mirrors that observed for the protein-coding transcripts (i.e., approximately half up-regulated and half down-regulated), which is consistent with a large fraction being processed from protein-coding transcripts. Examples of the transcriptional response of specific pri-miRNAs are shown in Fig. 3.15B. The primary transcript of both examples is considerably larger than the processed miRNA. Therefore, as with the protein-coding genes, the leading (or lagging) edge of the polymerase wave can be seen during the transcriptional response of the up-regulated (or down-regulated) genes. Together, these results suggest that the transcription of pri-miRNA genes are regulated by E2 in a similar pattern and with similar kinetics as protein-coding genes.

Previous reports indicate that transcripts regulated on a post-transcriptional level by miRNAs tend to be expressed in the same cell types as the miRNA transcripts which regulate them (Sood et al., 2006). We determined whether estrogen stimulation involves a coordinated response between pri-miRNA transcripts and the protein coding genes that they ultimately regulate. For this analysis, we reasoned that the subset pri-miRNAs undergoing long-lasting and relatively large regulatory changes are the most likely to be reflected as changes in processed, mature miRNA. Therefore,

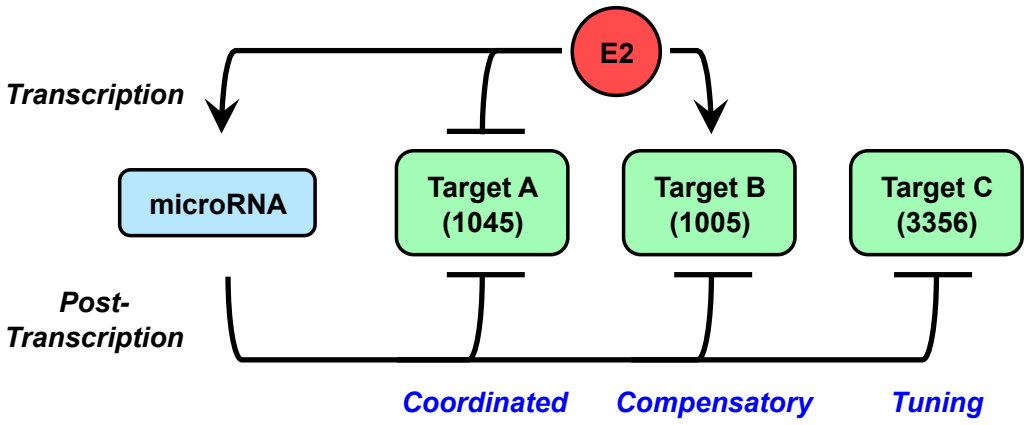
we focused on 47 of the 119 (~40%) regulated pri-miRNA transcripts that show more than 3-fold up- or down-regulation. These 47 robustly E2-regulated pri-miRNAs potentially target ~2,700 mRNAs according to the TargetScan database (Grimson et al., 2007; Lewis et al., 2005), or ~12.8% of RefSeq annotated mRNAs.

Interestingly, as shown in Fig. 3.15C, MCF-7 cells express a larger fraction of the ~2,700 target mRNAs than expected, such that 16.6% of expressed genes are targets of these miRNAs ($p = 3.7 \times 10^{-14}$; Fisher's exact test). This enrichment is consistent with an integrated regulatory program between the miRNAs expressed in a cell and the corresponding mRNA targets, consistent with previous suggestions (Farh et al., 2005). Importantly, the subset of genes regulated by E2 are enriched even further over those genes that are expressed by the cell, such that 18.6% of E2-regulated mRNAs are targets of E2 regulated pri-miRNAs ($p = 0.03$) (Fig. 3.15C). Moreover, this pattern of enrichment is robust to selecting a smaller set of miRNAs that are > 5-fold regulated by E2 ($p = 0.02$), or taking all miRNA transcripts regardless of their fold-change ($p = 0.003$)]. We found no evidence that E2 coordinates the direction (i.e. up or down) of regulation between pri-miRNAs and their potential target genes, either at the transcriptional level (per GRO-seq; Fig. 3.15D, middle panel) or at the steady state level (per expression microarrays; Fig. 3.15D, right panel). In fact, we found evidence for both coordinated and compensatory regulation (Fig. 3.15D, Fig. 3.16). Together, these results suggest an integrated regulatory program for E2-regulated transcription of pri-miRNA transcripts and the mRNAs targeted by the mature miRNAs.

3.3.8. Dramatic up regulation of the protein biosynthetic machinery by estrogen signaling

Because our GO analyses showed enrichment in genes with a primary

Figure 3.16. A circuit diagram of potential regulatory mechanisms involved in the E2-dependent regulation of miRNA transcripts and their targets. In this diagram, arrows indicate positive regulatory effects and blunted lines indicated inhibitory effects. This example shows a positive effect of E2 on the transcription of the miRNA gene (a similar representation could also be used to indicate a negative effect of E2 on the transcription of the miRNA gene). **"Coordinated effects"** are those in which the effects of E2 on the transcription of the miRNA gene and the target gene would be expected to produce the same outcome. For example, as shown, E2 up-regulates the transcription of a miRNA gene whose encoded miRNA inhibits (degrades or inhibits the translation of) Target A mRNA. At the same time, E2 down-regulates the transcription of Target A. Both processes ultimately act to decrease the amount of Target A protein, producing a coordinated response. **"Compensatory effects"** are those in which the effects of E2 on the transcription of a miRNA gene would be expected to reverse or attenuate the effects of E2 on the transcription of the target gene. For example, as shown, E2 increases the transcription of Target B, but at the same time increases the transcription of a miRNA gene whose encoded miRNA inhibits (degrades or inhibits the translation of) Target B mRNA. **"Tuning effects"** are those in which there is not a direct effect of E2 on a specific gene (Target C, in this case), but E2 nonetheless can regulate the expression of that gene by controlling the transcription of a miRNA gene whose encoded miRNA inhibits (degrades or inhibits the translation of) Target C mRNA.


E2
Transcription
microRNA
Target A
(1045)
Target B
(1005)
Target C
(3356)
Post-
Transcription
Coordinated
Compensatory
Tuning

biological function in protein biosynthesis, we asked whether E2 signaling has a broader effect on the protein biosynthetic machinery. GRO-seq provides a measure of all three eukaryotic polymerases; we therefore extracted and analyzed the data for changes in the 45S rRNA (RNA Pol I) and tRNAs (Pol III) annotated in the rnaGene track in the UCSC genome browser. Our analysis revealed that the transcription of Pol I and Pol III transcripts show a similar pattern of regulation by E2: (1) an initial burst at 10 minutes, (2) a slight decrease at 40 minutes, and (3) a maximal increase at 160 minutes (Fig. 3.17, A and D).

For individual tRNA genes, changes were strongly biased toward up-regulation, with the transcription of >90% of the tRNA genes showing up-regulation (Fig. 3.17B). Furthermore, this regulation unambiguously affects 158 of the 486 functional annotated tRNA genes (32%) in at least one of the time points. If the cell is indeed regulating tRNA genes in order to facilitate an increase in translation, one may expect that all 20 amino acids will be up-regulated. Indeed, we found that of the 158 up-regulated tRNA genes, at least one tRNA gene coding for each of the 20 amino acids is represented ($p = 0.0012$; Fisher's exact test) (Fig. 3.18). In addition to the 20 primary amino acids, we also found the tRNA coding for the amino acid variant selenocysteine, which is though to play a role in anti-oxidant activity and hormone biosynthesis (Stadtman, 1996), to be regulated by E2. One may also expect that up-regulation of the protein biosynthetic machinery may require a systematic increase across a wide spectrum of the 64 possible codon combinations. Because each three letter combination of codons is represented multiple times in the 486 annotated tRNA genes, we also asked whether E2 regulates a larger fraction of the 64 possible codon combinations than expected by chance. Indeed, we find 64% of the 64 codon combinations are unambiguously regulated by E2, which is more than expected based on our ability to call 32% of tRNA genes as regulated ($p = 0.0027$; Fisher's exact test).

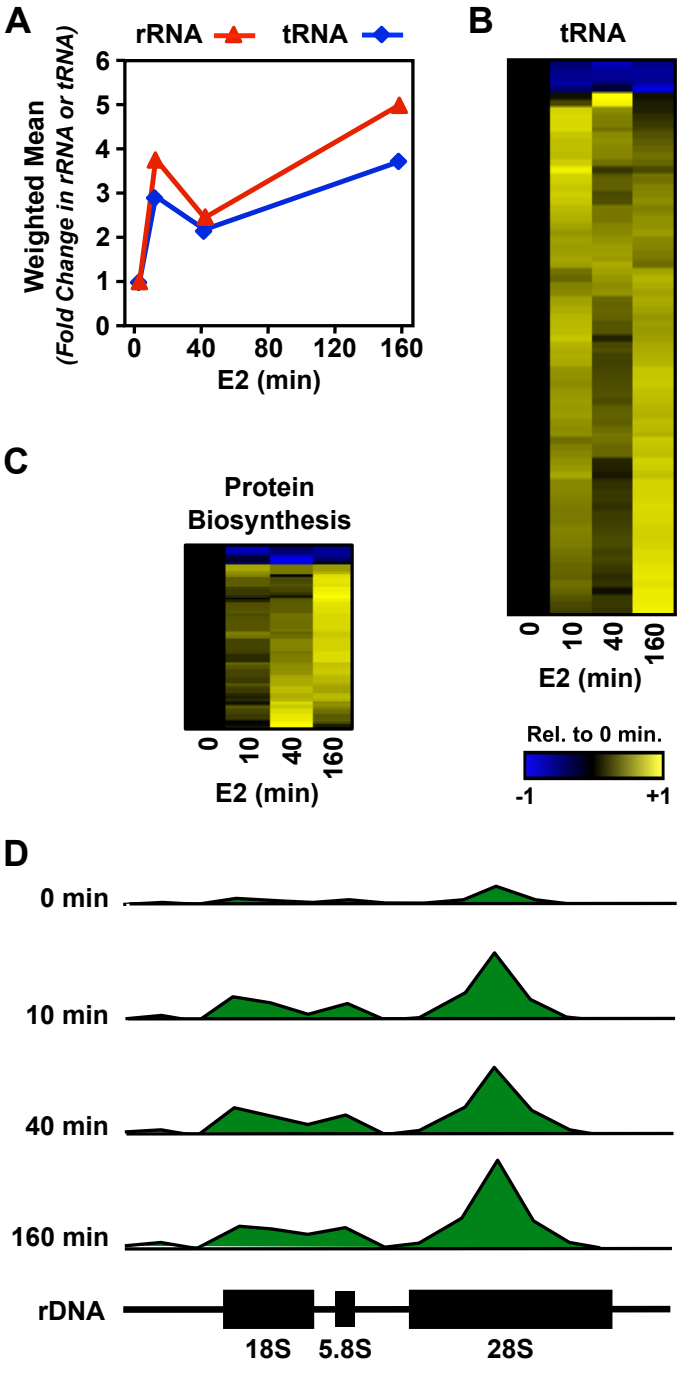
These results demonstrate that the observed changes in the protein biosynthetic machinery are applied in a robust and coordinated manner across amino acid and codon variations.

We also conducted a more focused analysis of protein coding genes with functions or cellular localization suggesting a role in protein biosynthesis (e.g. ribosome biogenesis, tRNA aminoacetylation, etc.; see Fig. 3.S5B for all GO terms used). As we observed for tRNA genes, protein coding genes represented in these groups are strongly biased toward up-regulation (Fig. 3.17C). As suggested by the GO analysis above, these genes are strongly enriched in the 160 minute max class ($p = 6.7 \times 10^{-13}$; Fisher's exact test), suggesting that these are sustained effects that translate the widespread changes observed in the cellular transcriptome to the proteome. Taken together, these results demonstrate a potent effect of estrogen signaling on the protein biosynthetic machinery. In addition, they highlight the fact the estrogen signaling has strong, immediate, and likely direct effects on transcription by all three RNA polymerases, not just Pol II. Up-regulation of the protein biosynthetic machinery is likely a means by which the estrogen signaling pathway prepares the cell for translation of the protein coding transcripts that are newly synthesized in response to estrogen signaling. It may also allow increased translation of extant stable transcripts in the cell, which may contribute to the overall estrogen signaling program. Previous studies have shown that actively dividing cells require increased protein synthesis to maintain their proliferate ability. This idea fits well with the known potent mitogenic effects of E2 on MCF-7 cells.

3.3.9. Relationship of ERα binding sites to primary estrogen target genes

Although most ERα binding sites are located distal to the promoters of protein coding genes, a small but highly significant, enrichment of ERα binding sites has been

Figure 3.17. E2 regulates transcription by Pol I and Pol III. (A) E2-dependent fold-change in the transcription of the 45S rDNA (Pol I) and tRNA (Pol III) genes. **(B and C)** Heatmap of the time course of E2-dependent regulation of tRNA transcripts (B) or protein coding transcripts encoding genes with a biological function or cellular compartment related to the synthesis, metabolism, or function of mature tRNAs or rRNAs (C). **(D)** GRO-seq reads mapped to the human rDNA gene (GenBank U13369.1) shown in 1 kb bins relative to the genome location during the time course of E2 treatment.


A
rRNA
tRNA
Weighted Mean
(Fold Change in rRNA or tRNA)
6
5
4
3
2
1
0
0
40
80
120
160
E2 (min)
B
tRNA
0
10
40
160
E2 (min)
Rel. to 0 min.
-1
+1
C
Protein
Biosynthesis
0
10
40
160
E2 (min)
D
0 min
10 min
40 min
160 min
rDNA
18S
5.8S
28S

A

First Position	Second Position: U		Second Position: C		Second Position: A		Second Position: G		Third Position
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG	STOP	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Annotated in hg18
Not annotated in hg18
N E2-Regulated

B

Protein coding genes with a primary biological function associated with protein translation, or otherwise associated with the ribosome

GO:0005840	Ribosome cellular compartement
GO:0042254	Ribosome biogenesis
GO:0016072	rRNA metabolic process
GO:000636:	rRNA processing
GO:0043039:	tRNA aminoacetylation
GO:0008033:	tRNA processing

Figure 3.18. E2-dependent transcriptional regulation of the protein biosynthetic machinery. (A) A representation of the genetic code showing specific the codons of tRNAs that are up-regulated by E2 treatment. Those codons highlighted in red are represented in tRNAs that are up-regulated by E2. Grey boxes indicate that no annotation is present for a tRNA representing the specified codon in hg18.

(B) Protein coding genes with GO terms specifying a biological function or cellular compartment related to the synthesis, metabolism, or function of mature tRNA or rRNAs.

observed in the proximal promoters of up-regulated genes (Carroll and Brown, 2006; Carroll et al., 2005), consistent with a direct role of ERα in mediating their regulation. This enrichment was strongest for up-regulated genes detected by expression microarrays following three hours of E2 treatment, with decreasing enrichment for genes regulated by longer treatment times (Carroll et al., 2006), likely reflecting an increasing number of indirectly regulated genes at later time points. Because our GRO-seq data reflects the direct transcriptional output of the cell, and because our shorter treatment times make it unlikely that we will detect secondary changes in transcription, we reasoned that we should observe a larger fraction of the genes that are regulated by GRO-seq are near ERα binding sites. To test this hypothesis, we used existing ERα ChIP-seq data (Welboren et al., 2009) to determine the fraction of E2 regulated genes with a proximal ERα binding site (<10 kb to the transcription start site). Indeed, we found that 46% of genes up-regulated by E2 at shorter time points (i.e., 10 and 40 minutes) contain an ERα binding site within 10 kb of the transcription start site.

Interestingly, when we analyzed the four classes (10, 40, 160 minute max, and down-regulated) of regulation separately, we found striking differences in binding site enrichment between these classes (Fig. 3.19A). In particular, almost half of the genes in the 40 minute max class are located within 10 kb of an ERα binding site, a striking enrichment over the ~10% found for RefSeq genes in general ($p < 2.2 \times 10^{-16}$; Fisher's exact test). Genes in the 10 minute max class are also substantially enriched for proximal ERα binding sites (33%; $p = 1.2 \times 10^{-12}$). Up-regulated genes that peak after 160 minutes have a lower level of enrichment that is not statistically significant (12%. $p = 0.24$), suggesting that a substantial fraction of this subset of genes reflects secondary effects. Conversely, down-regulated genes were slightly less likely than average to be located within 10 kb of an ERα binding site (8%, $p = 0.01$). This

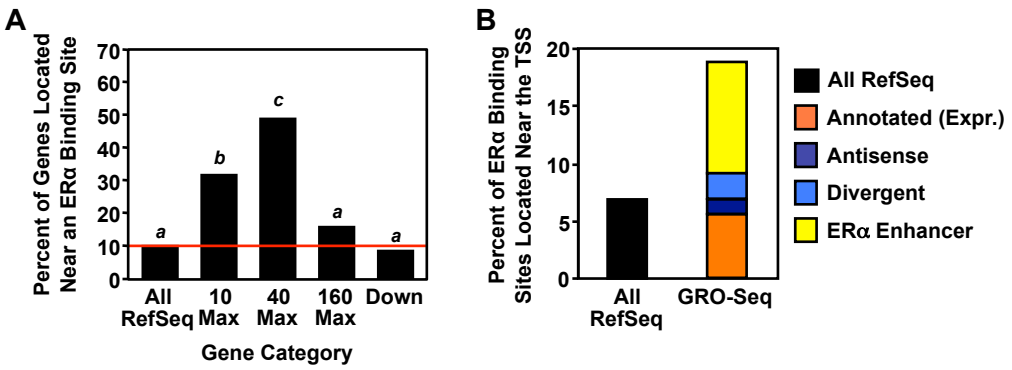


Figure 3.19. ERα binding sites are enriched in the promoters of primary E2 target genes. (A) The fraction of the specified subset of genes with an ERα binding site found within 10 kb of the TSSs. Bars with different superscripts are significantly different ($p < 1.2 \times 10^{-12}$). **(B)** The fraction ERα binding sites found within 1 kb of either all well-annotated RefSeq genes or the specified subset of de novo transcript annotations determined by GRO-seq analysis.

observation strongly suggests that E2 mediates up- and down-regulation by different mechanisms, and that immediate up-regulated genes tend to be the direct, genomic targets of ERα.

A surprising result from previous ERα genomic localization studies is that the vast majority of E2-induced ERα binding sites are located outside of the proximal promoter regions of protein coding genes (i.e., as many of 96% of sites located greater than 1 kb from the TSS; Carroll et al., 2005; Carroll et al., 2006). Using the de novo annotation approach reported here, we observe that only ~50% of the transcripts detected in MCF-7 cells correspond to classical, well-annotated protein-coding genes (Fig. 3.5B). Therefore, we determined whether a larger fraction of the ERα binding sites in MCF-7 cells are found proximal to one of the previously un-annotated transcripts that was detected in the present study. Using existing ERα ChIP-seq data (Welboren et al., 2009), we determined the fraction of ERα binding sites that map within the proximal promoter (<1 kb) for each class of transcript defined in our GRO-seq analysis. First, we compared ERα binding sites to the proximal promoter region of genes annotated in the RefSeq database, and found that ~6 to 7% of ERα binding sites are found proximal to an annotated gene (Fig. 3.19B), similar to what has been reported previously (Carroll et al., 2005; Carroll et al., 2006). Next, we calculated the fraction of ERα binding sites that fall near transcripts detected in MCF-7 cells in the present study. We found that ~18% of ERα binding sites fall near transcripts detected using our HMM in MCF-7 cells (Fig. 3.19B). This includes ~5 to 6% of ERα binding sites near transcripts matching annotated genes that were specifically found to be expressed in MCF-7 cells using our approach (Fig. 3.19B, orange bar). Furthermore, we find that an additional ~12% of ERα binding sites are found in the proximal promoters of genes producing transcripts that are not currently annotated in public databases, particularly ERα enhancer transcripts, but anti-sense and divergent

transcripts as well. While this finding still suggests that long-range enhancer-promoter interactions play a pivotal role in actions of ERα, as suggested previously (Fullwood et al., 2009; Pan et al., 2008; Theodorou and Carroll, 2010), it demonstrates a 3- to 4-fold increase in the fraction of ERα binding sites that are located near TSSs. Together with our observations of widespread regulation of unannotated and non-coding transcripts, this suggests that these previously unannotated classes of transcript play a pivotal role in the estrogen signaling pathway.

3.4. Materials and Methods

Cell culture. 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) with Hank's salts (Sigma) supplemented with 5% calf serum (CS), sodium bicarbonate, penicillin/streptomycin and gentamicin. Cells were plated for experiments in phenol red-free MEM (Sigma) supplemented with 5% charcoal-dextran treated calf serum (CDCS) prior to 17-β estradiol (E2) treatment.

Cell proliferation and cell cycle analyses. MCF-7 cells were plated at a density of 1 x 10^5 cells per well in a 6 cm plate in (1) MEM + 5% CS, (2) phenol red-free MEM + 5% CDCS, (3) or MEM without serum. For cell proliferation assays, the cells were trypsinized, collected, stained with trypan blue, and counted each day for 3 days. For cell cycle analyses, the cells were trypsinized and collected at various time points, washed twice with ice-cold PBS, and then fixed with ice-cold 70% ethanol for 1 h at −20°C. The ethanol-fixed cells were washed with ice-cold PBS and incubated with propidium iodide staining solution (40 μg/ml propidium iodide, 0.1% Triton X-100,

200 μg/ml RNase A) for 30 min at 37°C. Stained cells were analyzed with a BD-Biosciences LSRII flow cytometer and BD FACSAria™ software.

GRO-seq. GRO-seq was performed as described previously (Core et al., 2008), with limited modifications. Libraries were generated from two biological replicates.

Isolation of nuclei. MCF-7 cells were plated at a density of 1 x 10^6 cells per 15 cm diameter dish in phenol-red free MEM + 5% CDCS, using one dish per experimental condition. After three days, the cells were treated with 100 nM E2 as indicated and washed three times with ice cold PBS. Cells were collected in lysis buffer (10 mM Tris·HCl pH 7.4, 0.5% NP-40, 3 mM $CaCl_2$, 2 mM $MgCl_2$, protease inhibitors and RNase inhibitor) and pelleted by centrifugation at 500 x g for 5 min at 4°C. The cells were resuspended again in 1.5 ml of lysis buffer per 5x10^6 cells and pipetted up and down 20 times to release the nuclei completely. Release of the nuclei was checked by microscopy. The nuclei were pelleted again at 500 x g for 5 min., 4°C and resuspended in 100 μl of freezing buffer (50 mM Tris·HCl pH 8.3, 40% glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA) per 5 x10^6 nuclei.

Run-on and base hydrolysis. Nuclear run-on reactions were performed for 5 minutes at 30°C in the presence of NTPs (0.25 mM ATP and GTP, 1 μM α-^{32}P radiolabeled-CTP, and 0.25 mM 5'-bromo-UTP) and 0.5% Sarkosyl in run-on buffer (2.5 mM Tris·HCl pH 8.0, 75 mM KCl, 1.25 mM $MgCl_2$, 0.125 mM DTT, 0.2 U RNase inhibitor) to allow a run-on of ~100 nucleotides. The reaction was stopped by incubation with DNase I, followed by incubation with proteinase K. Finally, the newly synthesized nascent RNAs were isolated by acid phenol-chloroform extraction, followed by ethanol precipitation. After re-dissolving, the isolated RNAs were base-hydrolyzed with 0.2 N NaOH and the reaction was neutralized by the addition of 500

mM Tris·HCl, pH 6.8. The base-hydrolyzed RNAs were subjected to BioRad P-30 chromatography for buffer exchange.

Enrichment of nascent RNAs. The isolated and base hydrolyzed RNAs were subjected to three bead binding steps where the bromo-UTP incorporated nascent RNAs (BrU-RNAs) were enriched using anti-bromo-deoxy-U antibody conjugated beads (Santa Cruz Biotech). After each binding step, the BrU-RNAs were eluted, acid phenol-chloroform extracted, and precipitated. Each of the three bead binding steps also contained additional manipulations. After the first binding step, the BrU-RNAs were treated with tobacco acid pyrophosphatase (TAP; Epicentre) to remove 5'-methyl guanosine caps, then with T4 polynucelotide kiase (PNK; NEB) to remove 3'-phosphate group at low pH. The BrU-RNAs were treated with T4 PNK again at high pH in the presence of ATP to add 5'-phosphate group. 5'-adaptors were added to the end-repaired BrU-RNAs by T4 RNA ligase. After the addition of the 5'-adaptors, the BrU-RNAs were subjected to the second bead binding step to remove excessive adaptors and further enrich the BrU-RNAs. After the second bead binding step, 3' adaptors were ligated by T4 RNA ligase, followed by the third bead binding.

Amplification and purification of libraries. The affinity purified 5'- and 3'-adaptor-ligated BrU-RNAs were reverse transcribed into cDNAs usining annealed RT-oligo and Super Scripts III reverse transcriptase (Invitrogen). The RNAs were then degraded by incubation with RNase cocktail (RNases A/T1 and RNase H; Ambion). The cDNAs were then subjected to PCR-amplification using Phusion DNA Polymerase (Finzyme) with small RNA PCR primers. Samples of the amplified cDNAs were analyzed on a 2% agarose gel to assess yield and size. The remaining samples were extracted by phenol:chloroform:isoamyl alcohol (25:24:1). The purified cDNAs were run on a 6% native PAGE gel for further purification. The gel was stained with SYBRE gold and the cDNAs were visualized using a Dark Reader

transilluminator. The bands between size 100 bp to 250 bp were cut out and the cDNAs were eluted from the gel by incubating overnight in elution buffer (1x Tris•EDTA, 150 mM NaCl, 0.1% Tween 20). The eluted cDNA were extracted again with phenol-chloroform, resuspended in water, and quantified using a Nanodrop (Thermo Fisher). The final libraries were then sequenced using an Illumina Genome Analyzer.

GRO-seq data analyses - Transcript calling. Illumina sequencing reads were analyzed using available software, as well as a set of custom scripts written in the languages R (R Development Core Team, 2010), C/C++, and Perl. All custom software is available on request.

Short-read alignments. Short-reads were aligned to the human reference genome (hg18, NCBI36), including autosomes, X-chromosome, and one complete copy of an rDNA repeat (GenBank ID: U13369.1). The SOAP2 software package (Li et al., 2009) was used to align reads with the following options: (1) all non-unique 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). SOAP2 output was processed using custom Perl scripts, and imported into R for most the analysis.

Mappable regions. Regions of the genome (hg18) that were un-mappable at a read size of 44 bp were identified using the Tallymer program (Kurtz et al., 2008) in the Genome Tools package. Tallymer first builds an index of the genome using suffix arrays (suffixerator and tallymer programs). Subsequently, tallymer is used to search for all 44 bp sequences that occur more than once. The output of this program was imported into R for use in all subsequent analysis.

Transcript detection hidden Markov model. We detected transcribed regions de novo using a two-state hidden markov model (HMM) (Durbin, 1998). The model takes as input information about read counts across the genome, and subsequently divides the genome into two "states" representing "transcribed" and "non-transcribed" regions (Fig. 1C). Our objective was to detect a single set of transcripts that are active at any point during estrogen treatment. Read mapping positions were combined across all four time-points into a single read set. This combined set was used to train the model and to construct a single set of transcripts that were active during at least one point in the time course. Importantly, this combined approach increased our power for detecting transcripts with low expression levels, allowing us to more accurately annotate a larger fraction of transcripts. The genome was divided into non-overlapping windows of 50 bp. The number of reads mapping to each window was counted. Reads mapping to the "+" and "-" strand were counted separately (for the purposes of the model, the "+" and "–" strand are effectively treated as separate chromosomes). Free parameters of the model were estimated using the Baum-Welch Expectation Maximization (EM) algorithm. Two parameters were not trained using EM and were reserved for model tuning (described below). The Viterbi algorithm was used in combination with the final model parameters to obtain a set of transcript positions and orientations across the genome.

HMM parameters. The emission probabilities of our HMM represent the probability of observing a particular number of reads mapping to any 50 bp window of the genome. These were modeled using gamma distributions. Because the gamma distribution is always 0 when evaluated at $x = 0$, we used pseudocounts at every window in the model, incrementing the count in each window by 1. Emission probabilities for the transcribed state were modeled using a gamma distribution with two parameters, shape (k) and scale (ϑ). Both parameters were fit using the Baum-

Welch Expectation Maximization algorithm (Durbin, 1998). Starting parameters were set arbitrarily to: shape= 0.5 and scale= 10. Emission probabilities for the non-transcribed state were modeled using a constrained gamma distribution. Because we expect non-transcribed regions to have very few reads, we constrained the mean of the distribution to be one, which left only one free parameter, shape (k), which effectively defines its variance. We reserved the value of this parameter for model tuning, and fit it using the approach described below.

Transition probability represents the probability of switching from one state to another. Our two state model includes four transition probabilities, shown by the arrows in Fig. 1C. These include two self-transitions, in which either the transcribed or non-transcribed state remains in its current state. The probabilities associated with switching between the transcribed and non-transcribed states are also represented in the model. We determined the transition probability of moving from the non-transcribed to the transcribed state using the EM algorithm (Durbin, 1998). The transition of switching from the transcribed state to the non-transcribed state was held out as a second tuning parameter, which we fit using the methods described below.

Tuning the HMM. We set the values of the tuning parameters to optimize the performance of the HMM on annotated genes. To choose optimal values for the tuning parameters, we assume that GRO-seq annotations should largely be in agreement with annotations in annotated regions. To determine how well a particular set of parameters fit with annotations, we defined two distinct types of error and optimize the performance of the HMM over their sum. The two types of error are: (1) The fraction of genes that have two or more transcripts on the same strand end and start up again inside of a single gene annotation (in this case, the HMM is said to have 'broken up' a single annotation), and (2) The fraction of transcripts that continue between two non-overlapping genes on the same strand (the HMM is said to 'run

genes together'). These two types of error trade off, in that parameters which improve error rates on one parameter typically increase error rate of the other. For example, higher penalties on the transition probability tuning parameter favors annotating the genome with fewer, longer transcripts, and will therefore improve performance with respect to the second type of error at the expense of the first. Therefore, our strategy was to choose a fixed set of tuning parameters that minimize the sum of these error types.

Tuning parameters include: (1) the shape (or variance) of emissions in the non-transcribed state, and (2) the probability of switching from a transcribed to a non-transcribed state. We calculated the sum of the two types of error described above over a two-dimensional grid, and took the parameter settings that minimized the sum of the errors. We evaluated model performance in cases where the shape setting was set to 5, 10, 15, or 20, and the –log of the transition probability between transcribed and non-transcribed was set to 100, 150, 200, 250, 300, or 500. Final values selected by the model include a shape setting of 5 and a –log transition probability of 200.

Transcript annotation. Transcripts detected by the HMM were divided into six distinct, non-overlapping classes, which are intended to describe the function of each transcript, given currently available information. Annotations were made using the decision tree outlined in Fig. S1E. Each transcript was tested by a set of rules to determine membership in each class (as indicated below). Rules were applied in order, such that each transcript was assigned exactly one of the following six annotation types. Within each of the six types, multiple annotations were applied, allowing protein coding genes that contain non-coding RNAs (e.g. miRNAs) in the introns.

Our annotation pipeline is heavily dependent on the similarity of our MCF-7 GRO-seq transcripts to existing annotations. To make the annotation pipeline more

accurate, we first conditioned transcripts on all available gene annotations. To this end, all significantly sized (>5 kb) annotations in the RefSeq, ENSEMBL, and UCSC Known Gene databases were obtained from the UCSC genome browser (Rhead et al., 2010). Annotations were collapsed into contiguous genes using the featureBits program and the "-and" option (available from: http://hgwdev.cse.ucsc.edu/~kent/src/). Transcripts which overlapped multiple, non-overlapping annotations were broken at the most 5' base of the upstream transcript. Similarly, if multiple transcripts overlapped a single annotation, we joined them together, such that the final transcript had the 5' and 3' most positions of transcripts annotated by the HMM. This procedure effectively reduced the two types of error that were used to choose tuning parameters to 0 prior to annotation.

Annotation definitions. Transcripts were annotated as one of the following six classes according to the following rules:

(1) Annotated genic and non-coding RNA transcripts - Transcripts on the same strand and with significant overlap to existing annotations, including annotated protein coding genes, non-coding tRNA, rRNA, snoRNA, miRNA, or any other non-coding RNA annotations. Transcripts were classified as "annotated" if: (i) over 20% of any transcript overlapped more than 20% of any protein coding gene or (ii) any part of a transcript overlapped annotations for any functional RNA genes. Genic transcripts that contain miRNA or other functional RNA annotations in introns were counted as both a gene and the RNA transcript, and are stored in separate tables so that they are not counted twice in subsequent analysis. The following annotation sources were used: refGene, ensGene, knownGene, rnaGene (obtained using the UCSC genome browser; Rhead et al., 2010), and mirBase release 14 (Griffiths-Jones et al., 2006).

(2) Antisense (genic) transcripts - Transcripts that run anti-sense to gene annotations in refGene, ensGene, or knownGene. To be classified as antisense, >20%

of a transcript was required to overlap >20% of a well-annotated gene on the opposite strand.

(3) Divergent transcripts - Transcripts that overlap the 5' promoter of any detected "primary" transcript or annotated gene. Transcripts were included if >10% of a transcript overlapped the proximal promoter window (+/- 500 bp) of any transcript >1 kb in size on the opposite strand. The divergent transcript was also required to be <50% the size of its "primary" transcript, which excluded divergent enhancer-transcript pairs.

(4) ERα enhancer transcripts - Transcripts that overlap a previously defined binding site for ERα (Welboren et al., 2009). These transcripts overlap an ERα binding site, or begin within 1 kb of an ERα binding site that are located either (i) intergenic (distal from the 5' or 3' end of a gene), or (ii) within a gene.

(5) Other genic transcripts - This class was designed to include transcripts that have an extremely poor match to existing annotations, but can not be unambiguously classified as "unannotated" or "intergenic". Transcripts in this category overlap any segment of an annotation on either strand, but have a very poor match (<20%) to the annotation. Examples of these annotations are likely to include: (i) proximal promoter pausing on genes with very low levels of transcription in the gene body, (ii) divergent transcription from internal start sites (antisense), (iii) transcribed transposable elements inside of annotated genes, or (iv) reads systematically misaligned inside of genes.

(6) Intergenic transcripts - This class includes transcripts that do not have any overlap with either annotations or ERα binding sites. Examples of these annotations are likely to include: (i) active transcription of transposable elements, (ii) transcription at other distal enhancers (not ERα enhancers), (iii) post poly(A) transcription for well-

annotated transcripts, and (iv) systematically misaligned reads that happen to fall outside of existing annotations.

GRO-seq data analyses - Determining estrogen regulation of called transcripts. After calling transcripts, we identified transcripts that change in expression following estrogen treatment. We analyzed changes in transcripts annotated in RefSeq, the rnaGene track in NCBI36 (tRNAs), the UCSC Genome Browser (Rhead et al., 2010), and transcripts detected using our de novo HMM separately. In the discussion that follows, all are referred to simply as 'transcripts'.

Identifying estrogen-regulated transcripts. We detected estrogen dependent changes in gene expression using a method that samples the variation between two biological replicates and identifies genes in the different treatment conditions that fall outside of this expected variation. To determine E2-dependent changes in longer transcripts, we began by counting reads in the window between +1 kb and +13 kb relative to the TSS. This window was chosen for the following three reasons: (1) it prevented the counting of reads originating from RNA polymerases engaged at the promoter, but that are not productively elongating (i.e. paused polymerases) (Core et al., 2008), (2) it captured the region whose transcripts will be up-regulated at 10 min. of E2 treatment, assuming that Pol II elongation rates may be as low as ~1.2-1.3 kb/min (Ardehali and Lis, 2009), and (3) using this window over all points in the time course allowed the results to be compared between time points directly.

Next, regulation was determined using the edgeR package (v.1.4.1) (Robinson et al., 2010) for R/Bioconductor (Durinck et al., 2009; Kauffmann et al., 2009; Ritchie et al., 2009). The edgeR package was designed to detect changes in gene expression using short-read "digital" data from a small number of biological replicates. First, we used edgeR to fit the parameters of a negative binomial distribution to the variation in

read counts between combinations of the two biological replicates. Subsequently, this model was used in a negative binomial-based "exact test" (Robinson et al., 2010; Robinson and Smyth, 2008). This "exact test" assigns a p-value to each gene which reflects the probability that the observed change in read counts (between two different biological conditions) will occur given the level of global variation observed between biological replicates (Robinson et al., 2010). We used EdgeR to detect changes between the vehicle control and either the 10, 40, and 160 minute time points separately in annotated genes and de novo transcript predictions. P-values returned by edgeR were adjusted for multiple hypotheses testing using the false discovery rate (FDR) correction in R (function p.adjust). Transcripts with a FDR corrected q-value lower than 0.001 (approximate FDR under the model ~0.1%) were selected for further analysis.

Clustering, time course, and classification of temporal profiles. We selected all genes with an FDR corrected q-value of 0.001 at any point during the time course for inclusion in the temporal analysis. The results of the temporal analysis were plotted graphically as heatmaps and center-scaled traces of expression profile during the time course (e.g., Fig 2C). Genes in the heatmap were ordered using hierarchical clustering, with Ward's method used to measure the distance between genes (Danko and Pertsov, 2009). Expression measurements were centered and scaled for each gene using the R function "scale". We used expression measurements reported by the edgeR package (Robinson et al., 2010; Robinson and Smyth, 2008). These measurements represent quantile-normalized, model-adjusted transcript quantity and are considered slightly more robust than pooled, normalized read counts. After evaluating many different versions of the heatmap and dendrograms, we chose to use four clusters for the analysis of RefSeq genes, which provided a good compromise between minimizing the number of clusters and maximizing how well the clusters

describe the data. The clusters were generated using the R function "cutree" in the cluster package. Center-scaled traces of expression profiles for each cluster are plotted in Fig. 2.

Additional genomic analyses. In addition to the analyses described above, we performed a set of more focused analyses, as described below.

Gene Ontology analyses. Gene ontology analyses were performed using GoStat (http://gostat.wehi.edu.au/) (Beissbarth and Speed, 2004). All expressed genes were used as a background set to analyze GO term for each class ($p < 0.05$).

Protein biosynthesis-associated protein coding genes. Protein coding genes with a primary biological function or cellular compartment associated with the ribosome were identified using the Gene Ontology (GO) website (http://www.geneontology.org/). We selected the following GO terms for inclusion in this analysis: (1) ribosome cellular compartment (GO:365926), (2) ribosome biogenesis (GO:0042254), (3) rRNA metabolic process (GO: 0016072), (4) rRNA processing (GO:000636), (5) tRNA aminoacetylation (GO:0043039), and (6) tRNA processing (GO:0008033). P-values of enrichment in the gene class that peaks at 160 minutes were calculated using Fisher's exact test (fisher.test in R).

Comparing E2-induced changes in transcripts called by GRO-seq to changes observed by expression microarrays and Pol II ChIP-seq. All existing microarray datasets using the Affymetrix U133A or U133 2.0 platforms and representing the effects of E2 treatment on MCF-7 cells on were obtained from the Gene Expression Omnibus website (Cheung and Kraus, 2010; Kininis and Kraus, 2008). Raw CEL files were normalized for all data sets together using an approach described previously described (Danko and Pertsov, 2009). Briefly, samples collected using the U133A and

U133 2.0 platform were RMA normalized separately, and then combined by stripping off extra probes that were added to the U133 2.0 platform. Samples were corrected for batch effects using an Empirical Bayes based approach (Aryee et al., 2009; Gottardo et al., 2003; Johnson et al., 2007; Pan et al., 2008) and subsequently averaged to get final expression values for each gene, in each condition. In most of the analysis described below, we focused on microarray time points supported by data from more than one lab (including 0, 3, and 12 hours). Affymetrix probe sets corresponding to known RefSeq genes analyzed in the GRO-seq experiments were identified using a lookup table obtained from BioMart (Haider et al., 2009). All correlations were calculated using R.

We also compared our results to Pol II ChIP-seq data taken from a recent study (Welboren et al., 2009). To this end, mapped reads were obtained from Gene Expression Omnibus (Barrett et al., 2009) (accession numbers: GSM365929 and GSM365930, for control/vehicle and E2 treated, respectively**)**. Reads mapping to the +1 to +13 kb window of genes regulated by E2 during at least one point in the GRO-seq time course were counted. We computed the raw correlation between fold-changes in GRO-seq and Pol II ChIP-seq and generated scatterplots using R.

Comparing the GRO-seq results to known ERα binding sites. The complete list of 10,205 ERα binding sites defined by Welboren et al. (2009) were obtained from Gene Expression Omnibus (accession number GSM365926). Fourteen ERα binding sites mapping to the Y-chromosome were removed prior to the analysis, as this chromosome was not included in the GRO-seq mapping. Next, we calculated two fractions of interest: (1) the fraction of genes in a particular class that are found within 10 kb of an ERα binding site (Fig. 7A), or (2) the fraction of ERα binding sites mapping to within either 1kb, 5kb or 10kb from the 5' end of the nearest transcript identified de novo (using the HMM described above), or in a public database (Fig.

7B). For all analyses, the position of the ERα binding site was defined as the maximum of the enriched region, as reported in the file 'GSM365926.peaks.txt' available on GEO. Fractions of transcripts or ERα binding sites were calculated in R.

Correlations between primary transcripts and antisense/divergent transcripts. Correlations between primary transcripts and antisense/divergent transcripts were calculated using R. First, we identified the set of all "primary", or annotated protein-coding, transcripts that changed in response to E2 treatment in at least one time point. Then, for each primary transcript, we identified the set of all transcripts that were annotated as "antisense" or "divergent" using the annotation engine (described above). Pairs of primary and antisense/divergent transcripts were identified using the overlapSelect program (available from: http://hgwdev.cse.ucsc.edu/~kent/src/), allowing multiple antisense/ divergent transcripts to overlap with each primary transcript. All antisense/divergent transcripts were included in the analysis, regardless of whether or not their levels changed in response to E2 using the analysis described above. We then constructed a vector representing the expression levels of primary and antisense/divergent transcripts at each point in the time course. All points in the time course were included in the analysis, regardless of the time point at which the levels of the primary transcript changed. Correlations and scatterplots were generated in R (cor.test). Density isochrones were constructed using the kde2d function in the MASS package (Venables and Ripley, 2002).

Metagene analyses. We used metagene representations to illustrate the distribution of reads near a "typical" transcription start site. Intuitively, one can think of a metagene as a smoothed average of read density weighted by expression over the set of transcription start sites included in the analysis. Representations are made either over all transcription start sites in the genome, or over the genes in a particular class.

Here, separate metagene representations were generated for the 0, 10, 40, and 160 minute time points for each of the four clusters identified in the time course analysis. To compare different time points, all metagenes were scaled to a library size of 15 million reads. All plots were made in R.

Mathematically, we define a metagene as follows: Let M be a vector representing the number of GRO-seq reads falling a given distance from any generic transcription start site. Let m_i be an element in M where the subscript i denotes the position relative to the generic transcription start site. In this notation, i can take any real integer, where negative numbers represent positions upstream of the transcription start site, positive numbers represent positions downstream, and 0 denotes the transcription start site.

We define m_i as:

$$m_i = \sum_{c \in C} \sum_{s}^{[+1,-1]} \sum_{t \in T_{c,s}} \sum_{r \in R_{c,s}} I[(i-w) < s(P_r - P_t) \leq (i+w)]$$

Where:

- C is the set of all chromosomes included in the analysis;
- s denotes the strand along the DNA, and can take the values +1 (which denotes the positive strand) or -1 (negative strand);
- $T_{c,s}$ and $R_{c,s}$ denote the subset of transcription start sites or reads, respectively, mapping to strand s of chromosome c.
- P_t and P_r denote the position of transcription start sites or reads, respectively.
- I is an indicator function, taking the value 1 if the specified condition is met, and 0 otherwise.
- w denotes the window size, and is a free parameter which controls the amount of smoothing. Here, we use a value of 100 in all analyses.

MicroRNA analyses. We identified E2-regulated primary transcripts from our HMM transcript prediction algorithm that contain known miRNAs as described above. Each of these E2-regulated pri-miRNAs was associated with its regulatory targets using the TargetScan database (Lewis et al., 2005). All predicted targets were selected with a context score less than or equal to -0.3 (Grimson et al., 2007). When matching miRNA names between our dataset and TargetScan, we removed the final suffix (if present), which designates identical mature miRNAs that are coded by multiple genes. To test the enrichment of targets of E2 regulated pri-miRNAs, we calculated the fraction of all genes, expressed genes, or E2-regulated genes that are predicted targets of E2-regulated miRNAs. We focused this analysis on 47 of the 119 (~40%) regulated pri-miRNA transcripts that show more than three-fold up- or down-regulation at the 160 minute time point. Similar results were obtained using a more stringent cutoff (5-fold regulated) or using all E2-regulated miRNA transcripts. We also looked for evidence of coordination in the direction of regulation between miRNAs and the putative target transcripts (as illustrated in Fig. S4). For this analysis, we separated the 47 regulated miRNAs into 25 up-regulated and 22 down-regulated miRNAs. Next, we looked at the time course by GRO-seq or microarray of the putative target genes for up- and down-regulated miRNAs separately.

Filter binding assays. Nuclei were isolated from MCF-7 cells above after E2 treatment, as described. The nuclei were treated with (1) 1 μg/ml α-amanitin (Krackeler) to block RNA Pol II or (2) 1 μg/ml α-amanitin plus 12 μM tagetin (Epicentre) to block RNA Pols II and III. The nuclei were then subjected to run-on reactions in the presence of ^{32}P-CTP, the reactions were stopped, and the RNAs were isolated by acid phenol-chloroform extraction followed by ethanol precipitation, as described above. The isolated RNAs were then redissolved, mixed with sheared

salmon sperm DNA as a carrier nucleic acid, and precipitated by adding ice-cold 10% trichloroacetic acid (TCA) with incubation on ice for 10 min. The precipitated RNAs were collected by vacuum filtration on Whatmann GF/C glass fiber filters. The incorporated ^{32}P collected on the filters was quantified by liquid scintillation counting. The data were expressed as total transcription, Pol I transcription (signal remaining in the presence of α-amanitin and tagetin), Pol II transcription (signal calculated by subtracting activity inhibited by α-amanitin from total activity), and Pol III transcription (signal calculated by subtracting Pols I and II transcription from total transcription).

RT-qPCR gene expression analyses. Changes in the steady-state levels of the E2-regulated genes were analyzed by RT-qPCR, as previously described (Kininis et al., 2007) with a few modifications. MCF-7 cells were harvested at the specified time points using TRIZOL regent (Invitrogen) according to manufacturer's specifications. Two micrograms of isolated total RNA were reverse-transcribed from annealed oligo(dT) using 600 units of MMLV reverse transcriptase (Promega). The synthesized cDNA was treated with 3 units of RNase H (Ambion) for 30 min at 37°C and then analyzed by qRT-PCR using the primer sets listed below and a LightCycler® 480 real-time PCR thermocycler (Roche) for 45 cycles. The fold expression changes were normalized to *GAPDH* as an internal standard.

Primers for RT-qPCR analyses. We used the following primer sets for RT-qPCR:

AKAP1 mRNA Fwd:	5' – CTTGCCGAAGATCAGAGTCC -3'
AKAP1 mRNA Rev:	5' - ATTTGGCTTCACCACCTGTC -3'
C6orf48 mRNA Fwd:	5' - GGACAATGCACCTTCACAGA -3'
C6orf48 mRNA Rev:	5' -TGAGGGGGAGATTCCAAAC -3'

CASP7 mRNA Fwd:	5’ - CTGGGTGGGTACTTCCTTCA -3’
CASP7 mRNA Rev:	5’ - TGTGGTCTCCTAGACGTTGC -3’
CCDC88C mRNA Fwd:	5’ - AAAACTTTTGGCCCGTTTG -3’
CCDC88C mRNA Rev:	5’ - GCCGTCCACTAAATCCATGT -3’
EEF1A1 mRNA Fwd:	5’ - GCCAAAATGGGAAAGGAAA -3’
EEF1A1 mRNA Rev:	5’ - CCCGAATCTACGTGTCCAAT -3’
EHD1 mRNA Fwd:	5’ - ACAAGCCTATGGTGCTCCTC -3’
EHD1 mRNA Rev:	5’ - GTGTCGGATGAAGGTGGTCT -3’
EPPK1 mRNA Fwd:	5’ - CAGAGCAGGCCAGTGTACC -3’
EPPK1 mRNA Rev:	5’ - TACACCCCAGCTATGCTCCT -3’
FADS1 mRNA Fwd:	5’ - GCTACTTCACCTGGGACGAG -3’
FADS1 mRNA Rev:	5’ - ATCACTAGCCACCGCTCCT -3’
FAIM3 mRNA Fwd:	5’ - CTGGGCATTCATCCAGAAAT -3’
FAIM3 mRNA Rev:	5’ - AAGCTGACCTTGCTGATGGT -3’
FAM117B mRNA Fwd:	5’ - ATTCTACCCAGGCCTCCAGT -3’
FAM117B mRNA Rev:	5’ - CAGAGGAGATTGGCATGTGA -3’
FHL2 mRNA Fwd:	5’ - TGCCGTGAGTACCTCCAAC -3’
FHL2 mRNA Rev:	5’ - GCTCTGCTCCCCTCTCCT -3’
GFRA1 mRNA Fwd:	5’ - GCCATATTTGGCTGTGGTCT -3’
GFRA1 mRNA Rev:	5’ - CGGAGGACAATCAGCTCTTC -3’
GLB1L2 mRNA Fwd:	5’ - GTGACTTCCCCTTTCCCTTC -3’
GLB1L2 mRNA Rev:	5’ - CCTGCTAACTTGGGAAGACG -3’
GPD1L mRNA Fwd:	5’ - GATCTTCTGCAAAGGCCAAG -3’
GPD1L mRNA Rev:	5’ - CCTCCGTAACAGGTGGTGAT -3’
GREB1 mRNA Fwd:	5’ - GACCTGCCAAATGGAAGAAG -3’
GREB1 mRNA Rev:	5’ - AAAGCCATGTCCTTCCACAC -3’

HSPD1 mRNA Fwd:	5’ - CACAGTCTTTCGCCAGATGA -3’
HSPD1 mRNA Rev:	5’ - ACAGCATCGGCTAAAAGGTC -3’
IGFBP4 mRNA Fwd:	5’ - CCCACGAGGACCTCTACATC -3’
IGFBP4 mRNA Rev:	5’ - ATCCAGAGCTGGGTGACACT -3’
IL17RB mRNA Fwd:	5’ - TTGAGGGACCTCCGAGTAGA -3’
IL17RB mRNA Rev:	5’ - CCGGAGTACCCAGCTTACAT -3’
IRF1 mRNA Fwd:	5’ - TACTGTGGAGGCTGAAGCAC -3’
IRF1 mRNA Rev:	5’ - TGGCACAGTCTCAGCTCACT -3’
ITPK1 mRNA Fwd:	5’ - GAGGAGGAAGATGCAGACCT -3’
ITPK1 mRNA Rev:	5’ - TTCTCGCTCAGCCAGTAGC -3’
KCNK6 mRNA Fwd:	5’ - CGTGCTTGCTAACGCTTC -3’
KCNK6 mRNA Rev:	5’ - GAAGAGAGCAGAGGCGAAGT -3’
KCNMA1 mRNA Fwd:	5’ - AAGGGCTGTCAACATCAACC -3’
KCNMA1 mRNA Rev:	5’ - CCTGCAGCGAAGTATCATCA -3’
KRT19 mRNA Fwd:	5’ - CTGACACCATTCCTCCCTTC -3’
KRT19 mRNA Rev:	5’ - AACCCTGGTCTCAGAAGCTG -3’
KRT80 mRNA Fwd:	5’ - GCTGGCACTATCTCCAAGGT -3’
KRT80 mRNA Rev:	5’ - CCTTCTCCTGGTTCTTCAGC -3’
LPP mRNA Fwd:	5’ - CTTTTATTTGGGGGTGTGGA -3’
LPP mRNA Rev:	5’ - AGGCCGATCTTGTCTTGAAA -3’
LYAR mRNA Fwd:	5’ - GGCCCCAGACAATGAAATAA -3’
LYAR mRNA Rev:	5’ - TCTTCGGATCTGTGATGCTC -3’
MREG mRNA Fwd:	5’ - CTTGCTCGTCGAACTTACCC -3’
MREG mRNA Rev:	5’ - TGACTGCTGAGTCCTCATGG -3’
MYADM mRNA Fwd:	5’ - CCCCTCAGGTGCTCTTACAG -3’
MYADM mRNA Rev:	5’ - GGAGACGGATTAAGCCACAC -3’

MYC mRNA Fwd:	5’ - AGGGATCGCGCTGAGTATAA -3’
MYC mRNA Rev:	5’ - TGCCTCTCGCTGGAATTACT -3’
NAP1L1 mRNA Fwd:	5’ - GAAGGAGGACGGTTCAGTTG -3’
NAP1L1 mRNA Rev:	5’ - CTTGTAGTCTCCCAGCAACG -3’
NRIP1 mRNA Fwd:	5’ - AAATGGCCAGAAGGATGTTG -3’
NRIP1 mRNA Rev:	5’ - GCCATCAGTTTGCTTGATGA -3’
OBFC2A mRNA Fwd:	5’ - CCACCACTCATCTGACAGGA -3’
OBFC2A mRNA Rev:	5’ - ATTCTGGCCTGCTAGGAACA -3’
P2RY2 mRNA Fwd:	5’ - CAGGCACGTGGGTCTCTG -3’
P2RY2mRNA Rev:	5’ - CTGCGCTTCTCCTCTCAGG -3’
PDGFB mRNA Fwd:	5’ - GAAGGGAAAAGATCCCCAAG -3’
PDGFB mRNA Rev:	5’ - AGGGGGAAAGTGCAGTAGGT -3’
PDIA5 mRNA Fwd:	5’ - CTTGGTGCCCACACTGTAAG -3’
PDIA5 mRNA Rev:	5’ - CATCTTTGAAGGCATCAGCA -3’
PDLIM7 mRNA Fwd:	5’ - GCATCATGGATTCCTTCAAA -3’
PDLIM7 mRNA Rev:	5’ - GAGAGGGGCACATTGAAGTC -3’
PNRC1 mRNA Fwd:	5’ - CTGAGACGGGAGCATCTCTT -3’
PNRC1 mRNA Rev:	5’ - CAGTGGCCTGATCCTAGCTC -3’
PTPLB mRNA Fwd:	5’ - ATCAAATGGGCCAGGTACAC -3’
PTPLB mRNA Rev:	5’ - ACACTCCCATTGGGTACAGC -3’
RABGGTB mRNA Fwd:	5’ - GCCCGGCTCCTAAGTCTAC -3’
RABGGTB mRNA Rev:	5’ - AACAGGGAAAGGAGAGAGCA -3’
RPL22L1 mRNA Fwd:	5’ - GCAATTTCTACGGGAGAAGG -3’
RPL22L1 mRNA Rev:	5’ - CATTCCCGAGATTTCCAGTT -3’
RPS24 mRNA Fwd:	5’ - CACACCGGATGTCATCTTTG -3’
RPS24 mRNA Rev:	5’ - AAATCATGCCAAAGCCAGTT -3’

SETD2 mRNA Fwd:	5’ - GGGCTGTTTACAAACCCAAA -3’
SETD2 mRNA Rev:	5’ - TTACCTGACCACCCATCCTC -3’
SLC22A5 mRNA Fwd:	5’ - GGATTGTTGTGCCTTCCACT -3’
SLC22A5 mRNA Rev:	5’ - GGACTGCTGCTTCTTGGAAC -3’
SLC6A14 mRNA Fwd:	5’ - GTAGGGTGCTCAATGGTGGT -3’
SLC6A14 mRNA Rev:	5’ - TCTTGCCACTGTGGAATCTG -3’
SLC9A1 mRNA Fwd:	5’ - TCCCAGCTCCTCAGTCACTT -3’
SLC9A1 mRNA Rev:	5’ - ATGTGGGAAACTGAGCCCTA -3’
SNHG3 mRNA Fwd:	5’ - AGATGCACCGGACACTTTG -3’
SNHG3 mRNA Rev:	5’ - TTGAGCCATAGGATGCGATA -3’
STC2 mRNA Fwd:	5’ - CGACGGGAGGTTAATACCAA -3’
STC2 mRNA Rev:	5’ - AAGGTGGCCAACACCAAA -3’
TAF1D mRNA Fwd:	5’ - GAGTTGGAGGGCTGAGGTC -3’
TAF1D mRNA Rev:	5’ - ACTGGCCTGGTGTCCTAGAG -3’
TBC1D2 mRNA Fwd:	5’ - ACATGTCCCTTCCCCTCTCT -3’
TBC1D2 mRNA Rev:	5’ - GTGGCATCCCTGGGTAAGTA -3’
TFAP2C mRNA Fwd:	5’ - CGCCATGTTGTGGAAAATAA -3’
TFAP2C mRNA Rev:	5’ - CGCAGTCCTCTTCGTACTTG -3’
TFF1 mRNA Fwd:	5’ - GGAGCAGAGAGGAGGCAAT -3’
TFF1 mRNA Rev:	5’ - GGCGCAGATCACCTTGTT -3’
TMEM135 mRNA Fwd:	5’ - AGTCCTGGAGTCCCAAGGTT -3’
TMEM135 mRNA Rev:	5’ - ACTGGAGGCCTGGGTAGAAT -3’
TPD52L1 mRNA Fwd:	5’ - GAGGTAACCAGAAGCGGCTA -3’
TPD52L1 mRNA Rev:	5’ - GTGCTTCCCAGAGCAGATG -3’
TSEN2 mRNA Fwd:	5’ - GGGAGGAGAAAGATCCCTGT -3’
TSEN2 mRNA Rev:	5’ - GAAAAGACTGACGCCAGAGG -3’

USP31 mRNA Fwd:	5’ - CAGCAACACCGAGCTCTTC -3’
USP31 mRNA Rev:	5’ - CCAGCTGCTCAGTGACCTC -3’
WWC1 mRNA Fwd:	5’ - GCACTCCCAGCTGAAAAGTC -3’
WWC1 mRNA Rev:	5’ - CTGGCTGAAGTGGAGGAGTC -3’
ZYX mRNA Fwd:	5’ - CGATCTCCGTTTCGGTCT -3’
ZYX mRNA Rev:	5’ - ACAGGGCCGAACTTCTTCT -3’

REFERENCES

Acevedo, M.L., and Kraus, W.L. (2004). Transcriptional activation by nuclear receptors. Essays Biochem *40*, 73-88.

Ardehali, M.B., and Lis, J.T. (2009). Tracking rates of transcription and splicing in vivo. Nat Struct Mol Biol *16*, 1123-1124.

Aryee, M.J., Gutierrez-Pabello, J.A., Kramnik, I., Maiti, T., and Quackenbush, J. (2009). An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). BMC Bioinformatics *10*, 409.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., Marshall, K.A.*, et al.* (2009). NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Res *37*, D885-890.

Beissbarth, T., and Speed, T.P. (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics *20*, 1464-1465.

Bhat-Nakshatri, P., Wang, G., Collins, N.R., Thomson, M.J., Geistlinger, T.R., Carroll, J.S., Brown, M., Hammond, S., Srour, E.F., Liu, Y.*, et al.* (2009). Estradiol-regulated microRNAs control estradiol response in breast cancer cells. Nucleic Acids Res *37*, 4850-4861.

Boettiger, A.N., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471-473.

Carroll, J.S., and Brown, M. (2006). Estrogen receptor target gene: an evolving concept. Mol Endocrinol *20*, 1707-1714.

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.

Chang, E.C., Frasor, J., Komm, B., and Katzenellenbogen, B.S. (2006). Impact of estrogen receptor beta on gene networks regulated by estrogen receptor alpha in breast cancer cells. Endocrinology *147*, 4831-4842.

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.

Creighton, C.J., Cordero, K.E., Larios, J.M., Miller, R.S., Johnson, M.D., Chinnaiyan, A.M., Lippman, M.E., and Rae, J.M. (2006). Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors. Genome Biol *7*, R28.

Danko, C.G., and Pertsov, A.M. (2009). Identification of gene co-regulatory modules and associated cis-elements involved in degenerative heart disease. BMC Med Genomics *2*, 31.

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

Durbin, R., Eddy, S., Krogh, A., Mitchison, G. (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids (Cambridge University Press).

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc *4*, 1184-1191.

Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817-1821.

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

Fullwood, 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.

Gottardo, R., Pannucci, J.A., Kuske, C.R., and Brettin, T. (2003). Statistical analysis of microarray data: a Bayesian approach. Biostatistics *4*, 597-620.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res *34*, D140-144.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

Haider, S., Ballester, B., Smedley, D., Zhang, J., Rice, P., and Kasprzyk, A. (2009). BioMart Central Portal--unified access to biological data. Nucleic Acids Res *37*, W23-27.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J.*, et al.* (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564-1566.

Kauffmann, A., Rayner, T.F., Parkinson, H., Kapushesky, M., Lukk, M., Brazma, A., and Huber, W. (2009). Importing ArrayExpress datasets into R/Bioconductor. Bioinformatics *25*, 2092-2094.

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.

Kininis, M., Isaacs, G.D., Core, L.J., Hah, N., and Kraus, W.L. (2009). Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol Cell Biol *29*, 1123-1133.

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

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet *11*, 597-610.

Kurtz, S., Narechania, A., Stein, J.C., and Ware, D. (2008). A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. BMC Genomics *9*, 517.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

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., Strom, A., Vega, V.B., Kong, S.L., Yeo, A.L., Thomsen, J.S., Chan, W.C., Doray, B., Bangarusamy, D.K., Ramasamy, A.*, et al.* (2004). Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. Genome Biol *5*, R66.

Lis, J. (1998). Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. Cold Spring Harb Symp Quant Biol *63*, 347-356.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science *327*, 94-97.

Maillot, G., Lacroix-Triki, M., Pierredon, S., Gratadou, L., Schmidt, S., Benes, V., Roche, H., Dalenc, F., Auboeuf, D., Millevoi, S.*, et al.* (2009). Widespread estrogen-dependent repression of micrornas involved in breast tumor cell growth. Cancer Res *69*, 8332-8340.

Morris, K.V., Santoso, S., Turner, A.M., Pastori, C., and Hawkins, P.G. (2008). Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet *4*, e1000258.

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.

Platet, N., Cathiard, A.M., Gleizes, M., and Garcia, M. (2004). Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. Crit Rev Oncol Hematol *51*, 55-67.

R Development Core Team. (2010). R: A language and environment for statistical computing. In R Foundation for Statistical Computing (Vienna, Austria).

Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J.*, et al.* (2010). The

UCSC Genome Browser database: update 2010. Nucleic Acids Res *38*, D613-619.

Ritchie, M.E., Carvalho, B.S., Hetrick, K.N., Tavare, S., and Irizarry, R.A. (2009). R/Bioconductor software for Illumina's Infinium whole-genome genotyping BeadChips. Bioinformatics *25*, 2621-2623.

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.

Robinson, M.D., and Smyth, G.K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics *9*, 321-332.

Ruhl, D.D., and Kraus, W.L. (2009). Chapter 5 biochemical analyses of nuclear receptor-dependent transcription with chromatin templates. Prog Mol Biol Transl Sci *87*, 137-192.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell Cycle *8*, 2557-2564.

Sidhu, J.S., and Omiecinski, C.J. (1998). Protein synthesis inhibitors exhibit a nonspecific effect on phenobarbital-inducible cytochome P450 gene expression in primary rat hepatocytes. J Biol Chem *273*, 4769-4775.

Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. Proc Natl Acad Sci U S A *103*, 2746-2751.

Stadtman, T.C. (1996). Selenocysteine. Annu Rev Biochem *65*, 83-100.

Theodorou, V., and Carroll, J.S. (2010). Estrogen receptor action in three dimensions - looping the loop. Breast Cancer Res *12*, 303.

Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S, Fourth Edition edn (New York, Springer).

Warner, M., Nilsson, S., and Gustafsson, J.A. (1999). The estrogen receptor family. Curr Opin Obstet Gynecol *11*, 249-254.

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.

Werner, A., Carlile, M., and Swan, D. (2009). What do natural antisense transcripts regulate? RNA Biol *6*, 43-48.

Widelitz, R.B., Duffy, J.J., and Gerner, E.W. (1987). Accumulation of heat shock protein 70 RNA and its relationship to protein synthesis after heat shock in mammalian cells. Exp Cell Res *168*, 539-545.

CHAPTER 4.

Transcription at Estrogen Receptor α Enhancers Specifies Functional Estrogen Receptor Binding Sites

Dr. Charles Danko developed the bioinformatic and computational approaches used to analyze the GRO-seq data. His key contributions are represented in Figs. 4.2, 4.7, 4.8, 4.10, and 4.11.

4.1. Summary

ERα is a ligand-dependent transcription factor that binds to *cis*-acting DNA regulatory elements to modulate transcription. A majority of ERα binding sites are located in genomic regions distal from transcription start sites (TSSs). As such, it is not clear how ERα enhancers contribute to the estrogen-dependent expression of target genes. In this study, I have used Global Nuclear Run-On and Massively Parallel Sequencing (GRO-seq) to investigate ERα enhancer functions in estrogen-regulated transcription in MCF-7 human breast cancer cells. Using a novel bioinformatic approach based on the Hidden Markov Model (HMM) described in the previous chapter, I have observed a large number of unannotated transcripts that are initiated in proximity to the ERα binding sites. Strikingly, 7% of all transcripts detected by GRO-seq in MCF-7 cells originate from an ERα binding site. Although nearly all ERα binding sites are bound by Pol II, only a subset of those sites produce transcripts from ERα enhancers, indicating that the production of transcripts may specify functional enhancers. Also, the regulation of enhancer transcript expression correlates with the level of ERα binding, which suggests that transcripts generated from ERα binding sites are likely to be directly regulated by estrogen. Interestingly, short transcripts from ERα enhancers resemble divergent transcription from RefSeq Pol II promoters, which may indicate that transcribing enhancers keeps the chromatin in an active confirmation similar to active promoter, whereas long transcripts from ERα binding sites may be functional non-coding RNAs or novel unannotated protein coding. Additionally, the regulation of enhacer transcripts is likely to be correlated with genes that are regulated through looping machanism, suggesting that synthesis of enhancer transcripts are involved in promoting transcriptional regulation of genes that are distally located from functional trnascriptional binding sites. Together, these studies

have characterized the transcripts generated at distal ERα enhancers, will should provide a greater understanding of how liganded ERs regulate transcription networks.

4.2. Introduction

Estrogen has diverse roles in human physiology by regulating transcriptional outcomes through nuclear action of estrogen receptors (ERs), which serve as ligand-regulated transcription factors. Extensive genome-wide studies using ChIP-chip, ChIP-DSL, ChIP-Seq, and ChIP-PET have been used to map the location of ERα binding (Cheung and Kraus, 2010). ERα cistromes in a various cell type have been mapped to define the genome-wide mechanism of ERα binding and gene regulation (Carroll et al., 2006; Cheng et al., 2006; Jin et al., 2004; Kininis et al., 2007; Kwon et al., 2007; Laganiere et al., 2005; Welboren et al., 2009). Interestingly, the cistrome data sets using the same MCF-7 ERα positive breast cancer cells showed a limited overlap between the experiments, perhaps due to experimental variations such as the cellular growth conditions and the methods of data analysis. Despite all the differences, the published ERα cistrome analyses have provided us with a better understanding of ERα binding throughout the genome. One of main conclusions from the analyses is that only small fraction of ERα binding sites are located in the promoter-proximal regions of estrogen target genes. The vast majority of estrogen target genes defined by expression microarrays lack promoter-proximal ERα binding sites (Carroll et al., 2006; Kininis et al., 2007; Welboren et al., 2009).

Epigenetic modifications and factor recruitment have also been useful tools to define promoters and enhancers. The promoters of active genes are marked by tri-methylated histone H3 lysine 4 (H3K4me3), which promotes recruitment of chromatin remodeling complexes and histone modifying enzymes, whereas repressed genes are

(Welboren et al., 2009) genome-wide studies of epigenetic modifications have indicated that functional enhancers are marked by mono- and dimethylated, whereas non-functional enhancers showed elevated H3K9me2 (Heintzman et al., 2009). However, regardless of the histone marks found at the enhancers, it has been shown that Pol II co-localizes with ERα binding sites across the genome, indicating that enhancer activities are not only dictated by histone marks, but also determined by the regulation of coregulators and Pol II activities (Heintzman et al., 2007).

The molecular functions and characteristics of enhancers are still largely unknown. Enhancers have been implicated in locus control regions (LCRs) (Li et al., 2002), as well as the regulation of neighboring promoters for gene expression, in which a distal enhancer communicates with a proximal promoter through chromosomal looping (Wang et al., 2005). Recently, a genome-wide study using ChIA-PET has shown that the most high-confidence distal ERα enhancers are associated with promoters through chromatin looping, which represent long-range chromatin interactions used for coordinated transcriptional regulation (Fullwood et al., 2009). Other recent studies have revealed the production of transcripts from neuronal enhancers using RNA-Seq, suggesting that transcript production at the enhancer regions is involved in promoting mRNA synthesis at the promoters of nearby target genes (De Santa et al., 2010; Kim et al., 2010).

In the current study, I examined the estrogen-dependent production of transcripts from ERα binding sites using GRO-seq. I found that 7% of all transcripts detected by GRO-seq originate from ERα binding sites. Although nearly all ERα binding sites are bound by Pol II, only a subset of those sites produce transcripts from ERα enhancers, indicating that the production of transcripts may specify functional enhancers. As described below, I explored the nature and fucntion of ERα transcription using a variety of addition gene-specific, genomic, and bioinformatic

approaches. These studies have allowed me not only to identify novel transcripts that are generated from enhancer regions, but also to provide more mechanistic clues about how distal enhancers can influence estrogen-regulated transcription.

4.3. Results

4.3.1. Defining estrogen-regulated ERα enhancer transcription

Genome-wide studies have shown that a majority of ER binding sites is distally located from proximal transcription start sites (Carroll et al., 2006; Kininis et al., 2007; Welboren et al., 2009). First, we have applied several criteria to define ERα enhancers using ERα ChIP-seq data from Welboren *et al.* that can be unambiguously analyzed to call transcripts. These fell into two categories; (1) intergenic ERα binding sites, which are located >10 kb away from any annotated gene, and (2) ERα binding sites near or in genes, but that are associated with transcripts that are unambiguously distinguishable from other transcripts, such as antisense transcript or non-overlapping transcripts. In this study, we only focused on distal intergenic ERα enhancers that are >10 kb away from the nearest annotated genes (a total of 3,191 intergenic ERα enhancers) (Fig 4.1).

To further investigate whether these enhancers are enriched with Pol II, we utilized published Pol II ChIP-seq data from Welboren *et al*. (Welboren et al., 2009), counting reads in a 4 kb window centered on the maximal ERα ChIP-seq peak position. We then tested for enrichment using a statistical test (i.e., Poisson enrichment relative to a null model of redistributing reads uniformly across human genome sequence). We found that ~ 46% ERα enhancers have significant enrichment of Pol II. The correlation between ERα binding and Pol II localization also showed

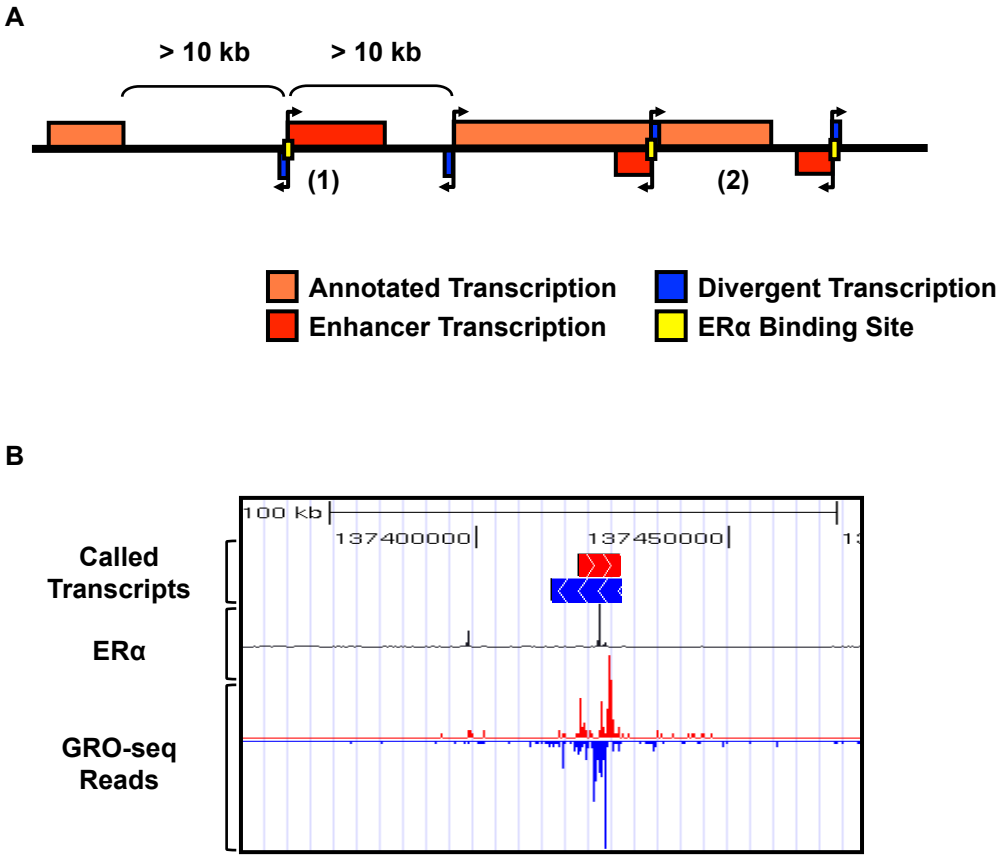


Figure 4.1. Defining ER enhancer transcription. (A) Two categories of ERα binding sites: (1) Intergenic ERα Enhancer that is > 10 kb away from any annotated genes, (2) near or in genes, but distinguishable from other transcripts (e.g., antisense). (B) Examples of GRO-seq reads at intergenic ERα enhancers.

high correlation with a correlation coefficient 0.65 ($p < 10^{-300}$) (Fig. 4.2). This was also observed in the ChIP-chip data sets from the Brown lab (Carroll et al., 2006), showing that Pol II and ERα are recruited to more than 9,000 independent high confidence sites with FDR < 1% across the genome of breast cancer cells upon estrogen-treatment.

Next, we looked at whether the enrichment of Pol II at the enhancers reflects the production of transcripts. We found that 978 ERα binding sites are associated with at least one transcript detected by using the two-state hidden Markov model (HMM) described in Chapter 3. In total, 829 unique transcripts are near this "intergenic" subset of ERα binding sites and the number of transcripts is less than the number of ERα enhancers due to the proximity or grouping of some of the ERα binding sites. Interestingly, approximately 67% of enhancers are enriched with Pol II generated transcripts, indicating that these may the functional enhancers. Among these ERα enhancers, approximately 570 ERα binding sites are associated with at least one transcript that is also regulated by estrogen. It is noteworthy that this is less than the total number of enhancer transcripts (~53.7% of them) since it does not include enhancer transcription that is inside of or near to RefSeq annotated genes. Also, ERα and Pol II ChIP-seq data from Welboren *et al.* were done at 60 minutes of estrogen treatment (Welboren et al., 2009), which may reduce the apparent correlation in some extent as peaks may have changed.

4.3.2. Regulation of ERα Enhancer Transcription

To investigate whether the transcription at the enhancer regions is regulated and how it might be affected upon estrogen stimulation, we first analyzed the regulation of all enhancer transcripts using a model-based approach called edgeR, where the variation in the expression measure across all transcripts between two or

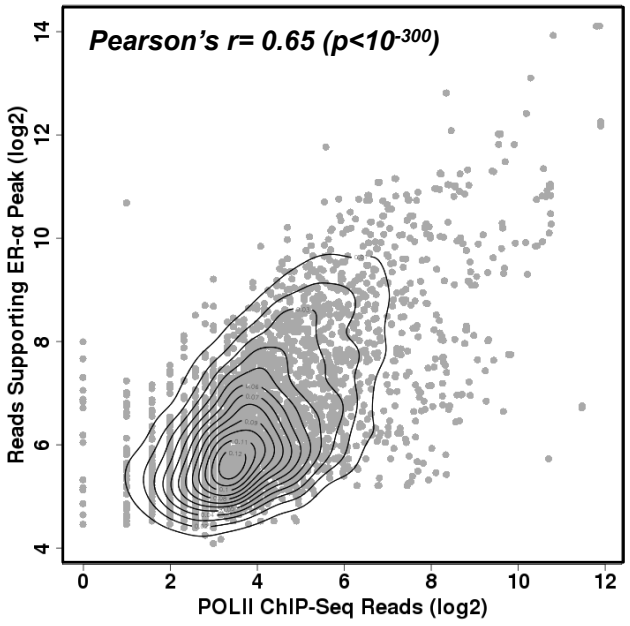


Figure 4.2. Correlation between ERα enhancer transcription and RNA polymerase II. Pearson correlation analysis was applied to determine the correlation between ERα binding sites and Pol II localization from ChIP-seq data sets.

more biological replicates is fitted in a statistical model (Robinson et al., 2010). Approximately 56% of the identified enhancer transcripts are regulated at least one time point of estrogen stimulation, which is significantly more than transcripts in general (Fisher's Exact test, p value = 4.14^{e-58},) (Fig. 4.3). The fraction that is unregulated may be due to the limits of our approach to detect regulation at a low number of sequence reads, unrelated to ER binding or estrogen action, or pre-estrogen action of pioneer factors such as FoxA1 (Eeckhoute et al., 2009; Lupien et al., 2008).

When we aligned the regulated enhancer transcripts using hierarchical clustering method over estrogen treatment time (Fig. 4.4A), we made an interesting observation that approximately ~90% of transcripts are significantly up-regulated at 40 min of estrogen treatment among the E2-regulated enhancer transcripts (Fig. 4.4B). Since enhancer transcripts are more likely to be regulated by ERα than other transcripts in general, we performed chromatin immunoprecipitation (ChIP) analysis to monitor the binding of ERα. This allowed us to characterize the relationship between ERα binding dynamics and the regulation of enhancer transcription (Fig. 4.5). Interestingly, the time course of maximal ERα binding corresponds to maximal enhancer transcription, which indicates that it is likely that enhancer transcripts are true direct estrogen targets regulated via direct ERα binding.

4.3.3. Characterizing ERα enhancer transcription

To further characterize the transcription at ERα enhancers, we performed a meta-gene analysis on 3,191 intergenic enhancers that are >10 kb away from any neighboring genes. For the majority of enhancers, GRO-seq metagenes appear similar to metagenes from transcription start sites. At 40 min estrogen treatment, the reads on the plus and minus strands peak at 50-100 bp upstream of the ERα binding site (shown in the meta-gene plots) (Fig. 4.6A right). This is consistent with a model

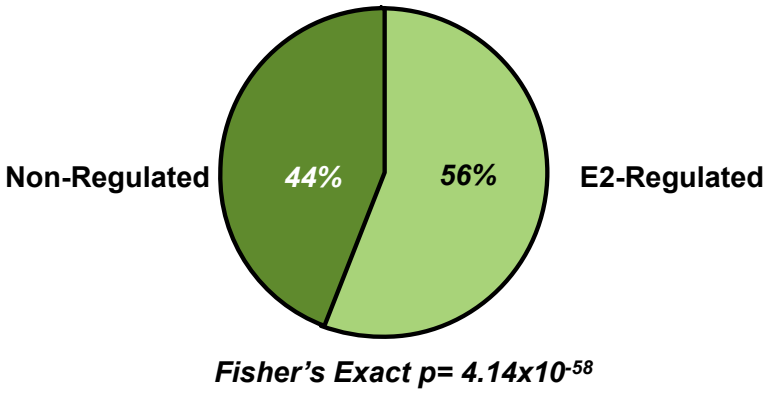


Figure 4.3. Determining estrogen-regulated ERα enhancer transcription. The fractions of ERα enhancer transcripts that are regulated by estrogen stimulation at least one time point of estrogen stimulation. The regulation was determined by using a statistical model, edgeR. (p-value= 4.14 x 10^{-58})

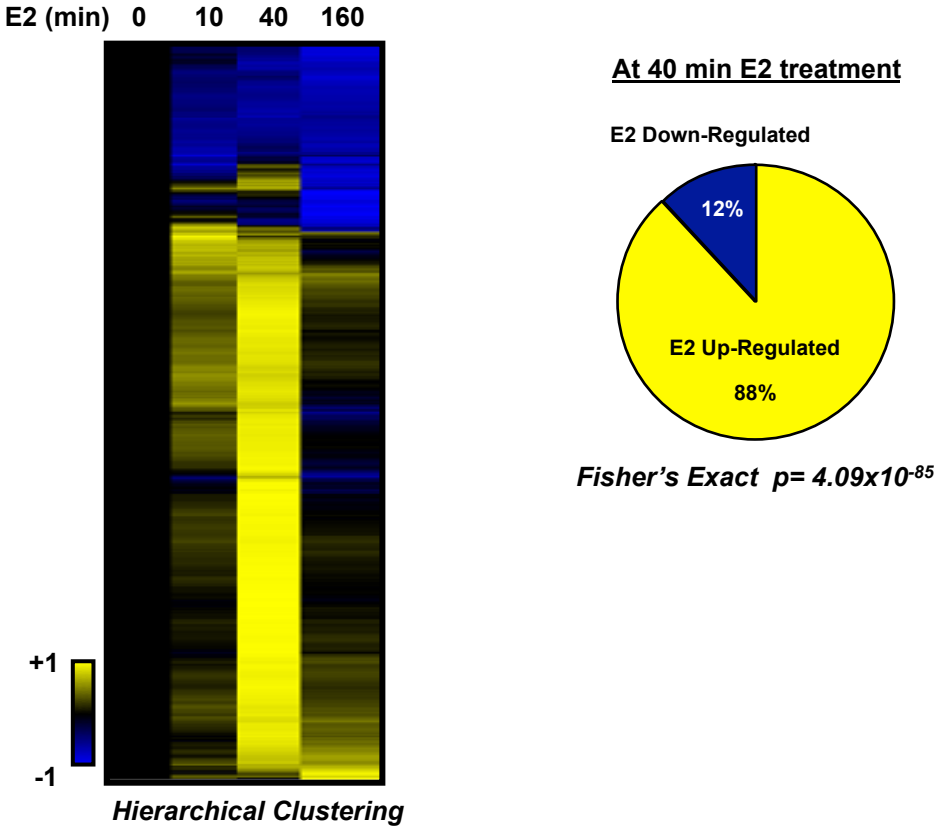


Figure 4.4. Determining estrogen-regulated ERα enhancer transcription. All the estrogen-regulated ERα enhancer transcripts that we detected was grouped based on hierarchical clustering and the regulation was analyzed over time.

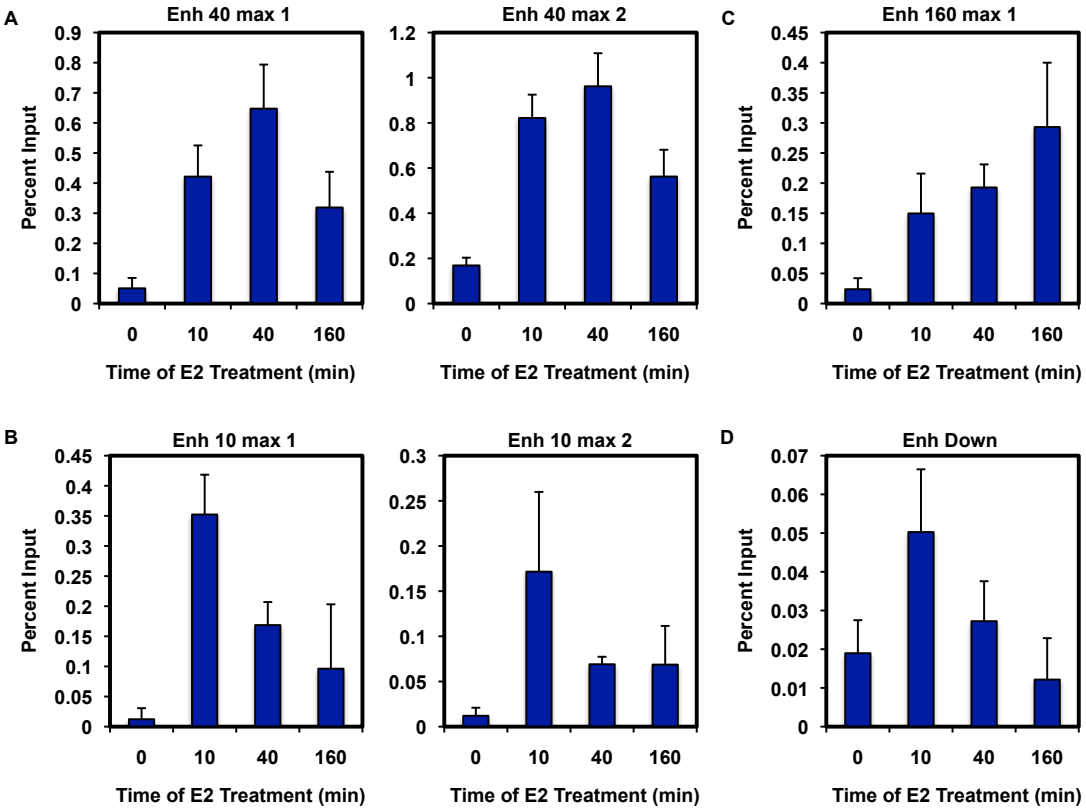


Figure 4.5. The dynamics of ERα binding at enhancers. ERα binding at enhancers whose transcript levels are (a) max at 40 min, (B) max at 10 min, (C) > 160 min, and (D) down-regulated upon estrogen stimulation were analyzed by ChIP-qPCR. Each bar = mean + SEM, n = 3.

in which Pol II initiates transcription at the presumptive ERα binding site, but the transition into productive elongation is blocked. The distance of the Pol II peak (50-100 bp away from the ERα binding site) is also very similar to that observed at the promoter regions (Fig. 4.6A left). Interestingly, at the 0 minute time point, an excess of reads can be detected over background, but there is no offset relative to the binding site; rather the reads seem to mark the ERα binding site directly. We also observed a significant correlation (Pearson correlation: 0.44) between the ERα binding and the number of GRO-seq reads at the 40 minute time point in a 4 kb window surrounding these ERα peak (data not shown). This is a much higher correlation compared with other time points (i.e., 0.22, 0.34, and 0.35 for the 0, 10, and 160 min times points, respectively).

To determine whether the polymerase at enhancer regions is predominantly Pol II, we carried out metagene analysis on the Pol II ChIP-seq dataset at ERα enhancer regions from Welboren et al. (Welboren et al., 2009). We found that Pol II is significantly enriched at ERα binding sites, implying that a sizable fraction of this transcription is likely to be the result of Pol II transcription, rather than Pol I or Pol III (Fig. 4.6B).

We also investigated the length distributions of transcripts that are generated from enhancers. Interestingly, our analysis of the size distribution of enhancer transcripts at both sense and antisense strands showed that about ~40% of the enhancer transcripts are longer than 5 kb and about 25% are longer than 10 kb . It is possible that these long transcripts may be non-coding functional RNAs or unannotated protein coding transcripts (Fig. 4.7). We further investigated the size difference of transcripts that are generated at both strands. To analyze this, total 369 ERα binding sites that are associated with two transcripts on both strands were analyzed. Overall, the length of both transcripts tend to be very similar (Perason's

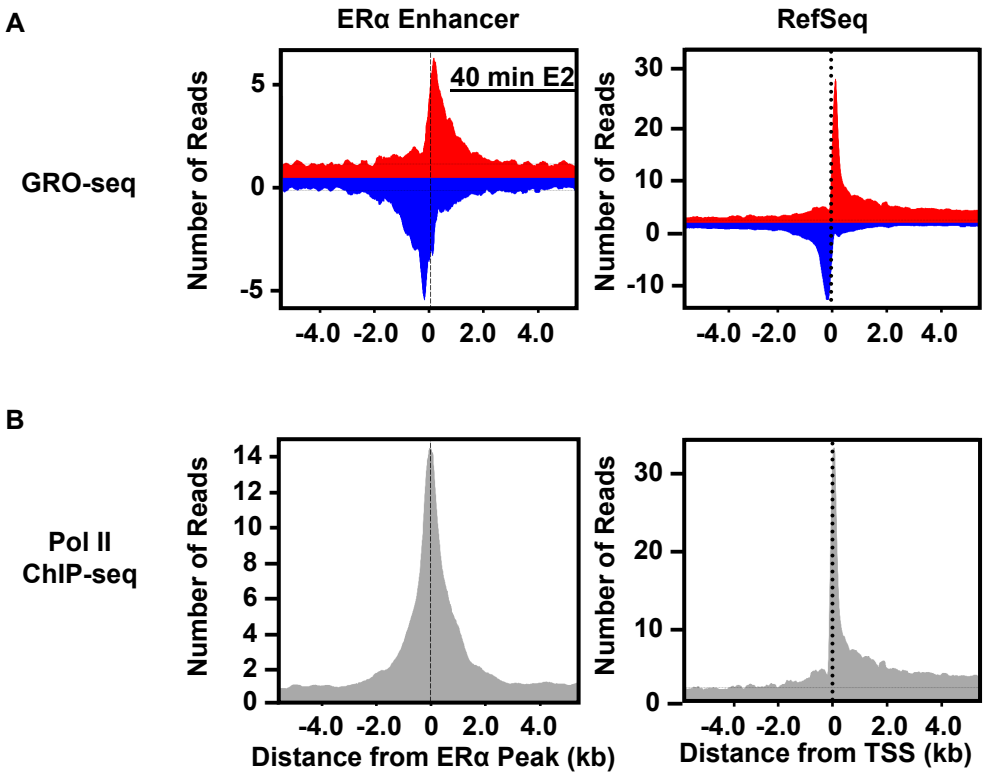


Figure 4.6. Metagene profiles of ERα Enhancer Transcription. (A) Metagene analysis of GRO-seq read counts at ERα enhancers (right) and at RefSeq promoters (left). (B) Metagene analysis of Pol II ChIP-seq tag counts at ERα enhancers (right) and at RefSeq promoters (left).

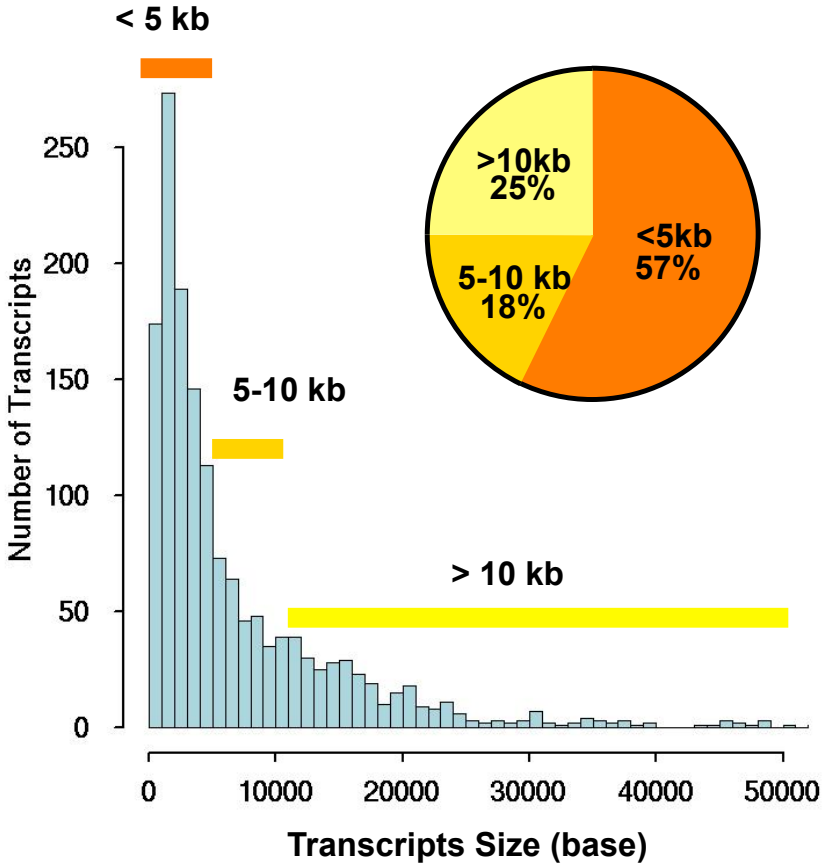


Figure 4.7. The length distribution ERα enhancer transcripts. All detected transcripts within 1 kb of the ERα enhancer were distributed based on size with 1 kb bins.

R=0.64), suggesting that the transcript size may reflect the size of open chromatin boundaries near each ERα binding sites (Fig. 4.8, black circles). There are also 53 transcripts (14.4% of 369 ERα binding sites) with a major size difference between the size of the two transcripts and one is longer than 10 kb in length, which is much like a gene and its divergent transcript (Fig 4.8, red circles). It is very likely that they represent novel, unannotated protein coding or non-coding transcripts.

We next determined the characteristics of ERα enhancers upon estrogen treatment using genomic metagene analyses of existing data sets from MCF-7 cells. Histone H3 dimethylated at lysine 4 (H3K4me2), which has been shown to be enriched at non-promoter-regulatory elements (Heintzman et al., 2007), as well as associated with DNase I hypersensitivity (Heintzman et al., 2007), was enriched upon estrogen treatments at ERα enhancers (Fig 4.9A). Also, TRIM24, a protein that is found to be overexpressed and bind to ERα to regulate gene expression in breast cancer cells (Tsai et al., 2010), is also enriched at ERα enhancers upon estrogen treatment (Fig 4.9B), indicating that estrogen stimulation may recruit enhancer-specific cofactors to generate functional transcripts.

4.3.4. Exploring functional connections of ERα enhancer transcription

Since one of the suggested functions for enhancers is to regulate gene expression by long-range interactions, we explored if there is any correlation between transcripts synthesis at the enhancer region and expression of estrogen-regulated target genes by a chromatin looping mechanism. Using ERα enhancer transcripts and ChIA-PET data set from Fullwood et al. (Fullwood et al., 2009), we first looked how many estrogen-regulated genes are shown to be involved in looping interactions. Approximately 10% of E2-regulated genes (531 E2-regulated genes) are involved in long-range interactions through looping. Among these genes, 75% (406 genes) show

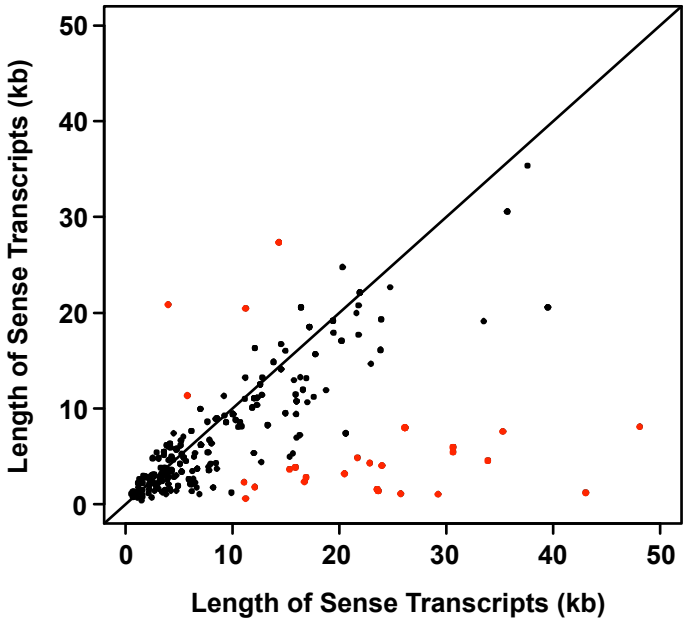


Figure 4.8. The size differences of enhancer transcripts. **A** total 369 ERα binding sites that are associated with transcripts originating on both strands were analyzed. Red circles indicate ERα binding sites with a major size difference between the two transcripts and one is longer than 10 kb in length.

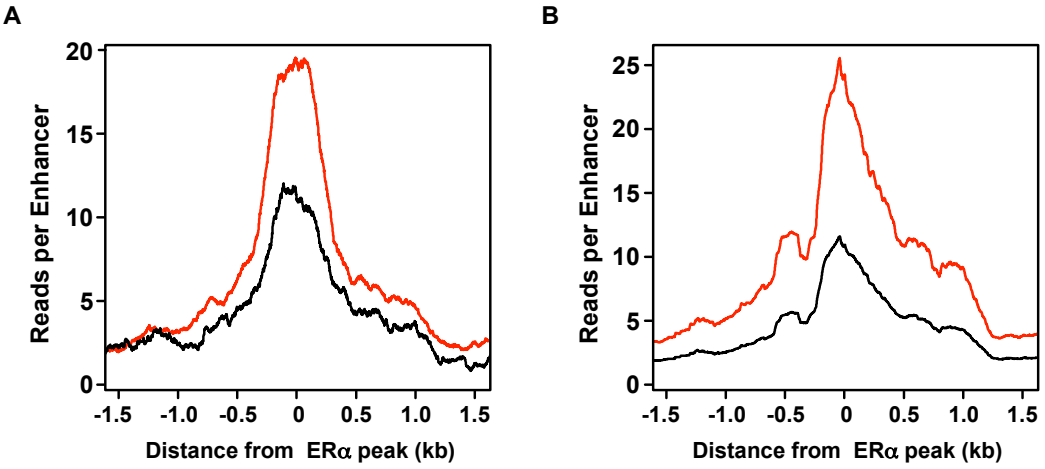


Figure 4.9. Metagene Analysis of TRIM24 and H3K4me2 at ERα enhancers. (A) TRIM24 and (B) H3K4me2. The black line indicates the data from vehicle treatment and the red line indicates the data from estrogen treatment (6 hrs).

peak transcription at 40 minutes. Also, of the 531 E2-regulated genes, there are 345 of looping interactions between ERα binding sites and a regulated gene (~10% of distal enhancers).

We further investigated the enhancers that loop to functional genes. They show a much higher read count relative to those enhancers that do not loop to estrogen-regulated genes (Fig. 4.10). Also, ERα binding as well as Pol II localizations were enriched at these enhancers. This suggests the possibility that the extent of transcript production is an indicator of functional enhancers. In this regard, distal ERα enhancers may serve as a docking platform for the binding of other factors that are necessary for target gene transcription. We also examined the correlation between the fold-change of GRO-seq reads at the enhancers and the fold-change of the regulated transcripts (pearson correlation = 0.57); the results indicate that changes in transcription at the enhancer are a predictive of the changes at the regulated target gene (Fig. 4.11).

We next examined whether the transcripts production at the enhancer are required for the looping, as well as ER binding. Since most of ERα enhancers generate a low level of transcripts prior to estrogen treatment, we tested treated MCF-7 cells with α-amanitin, an RNA polymerase inhibitor, prior to E2 stimulation to determine if the production of these transcripts affects ERα binding (Fig. 4.12). Interestingly, ERα binding is minimally affected by the blocking of transcription at enhancer regions, indicating that the binding of ERα is independent of Pol II localization and transcriptional elongation.

4.4. Discussion

Since the genomic ERα binding site analyses revealed that the that majority of ERα binding sites are located in distal regions from transcription start, it has been a long-standing question of how distal ERα enhancers contribute estrogen-dependent gene expression. Gene-specific and genomic studies have shown that the ERα-bound

enhancers are required for estrogen-dependent target gene transcription (Pan et al., 2008; Wang et al., 2005). A recent study has used the ChIA-PET method to investigate the communication between distal enhancers and promoter regions through long-range looping interactions (Fullwood et al., 2009). More recently, genome-wide studies using RNA-seq and ChIP-seq have shown the production of transcripts from the enhancer regions (Kim et al., 2010).

In this study, we used GRO-seq to investigate transcription events at ERα enhancer region with greater sensitivity and in unbiased ways. Strikingly, we have found many transcripts that are associated with ERα enhancers. However, it is interesting that only a subset of those sites were enriched with Pol II and about 67% of Pol II-enriched enhancers produce transripts from ERα enhancers, indicating that the enhancers that are not only bound by transcription factors, but also enriched with Pol II that can actually elogate and generate transcsripts may be the funtional enhancers.

One of the most intriguing observations is that the enhancers that loop to E2-regulated genes appear to produce more enhancer transcripts than those that do not loop to regulated genes. This suggests the possibility that distal ERα enhancers may serve as a docking platform for other factors binding that are necessary for the transcription of target genes. It is also possible that the production of transcripts at enhancers is a consequence, rather than a cause, of looping. Thus, it is important to find the causative relationship between looping and generation of transcripts.

The data also showed that short transcripts produced from ERα enhancers resemble divergent transcription from RefSeq Pol II promoters, which may indicate that transcribing enhancers keeps the chromatin in an active confirmation similar to active promoter. Another distinct feature we have identified is that most of the ERα transcriptions are uniformly up-regulated at 40 min estrogen stimulation. Gene-specific ERα binding analysis by ChIP suggests that receptor binding dynamics are

correlated with maximal regulation of transcripts, indicating that it is likely that these enhancer transcripts are direct targets of estrogen signaling.

In spite of our new findings, our analysis in conjunction with published studies, have raised more questions, such as (1) what is the regulatory curcuit to control Pol II to generate transcripts at a subset of enhancers, but not from all the enhancers?, (2) what coregulators are recruited?, and (3) what is the specific biological function of these enhancers?. From the correlation analysis and extensive genomic analyses of promoter regions, we can speculate that the generation of transcripts at the enhancers might be involved in promoting the formation of an open chromatin structure to facilitate the communication with target promoters. However, it is also possible that long transcripts that are generated from ERα enhancer region are themselves functional. The products could be direct estrogen-regulated functioanl non-coding regulatory RNAs or novel unannotate protein-coding transcripts.

The measurement of direct changes in nascent transcripts using GRO-Seq allows us not only to identify novel transcripts that are generated from enhancer regions, but also provides more mechanistic clues of how distal enhancers can influence estrogen-regulated transcription. To address the questions arising from our studies (e.g., what is the molecular mechanism of enhancer transcription?, what other factors are involved to specify functional enhancers? what are the biological functions of distal enhancers?), it is important to dissect the regulation processes more carefully by genomic, gene-specific, and cell-based analysis.

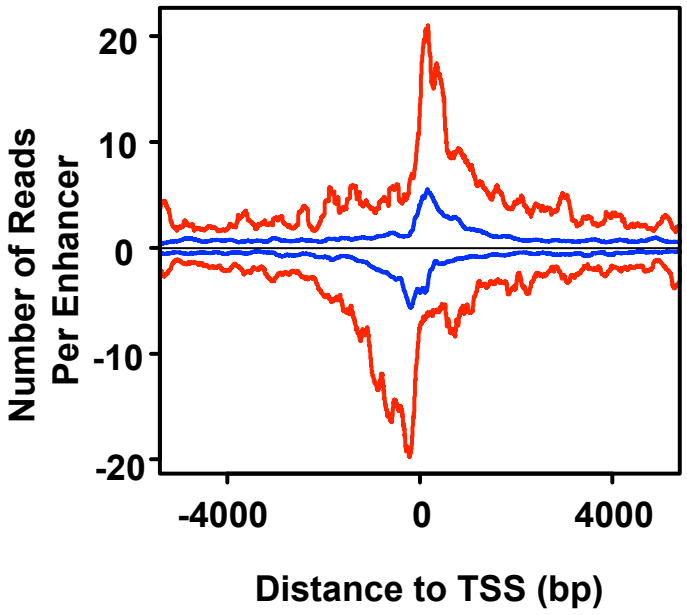


Figure 4.10. Metagene profile of ERα enhancer transcription in relation to gene looping. Metagene profiles of GRO-seq read counts at ERα enhancers that do not loop to and estrogen-regulated genes (blue) and those that do loop to an estrogen-regulated gene (red).

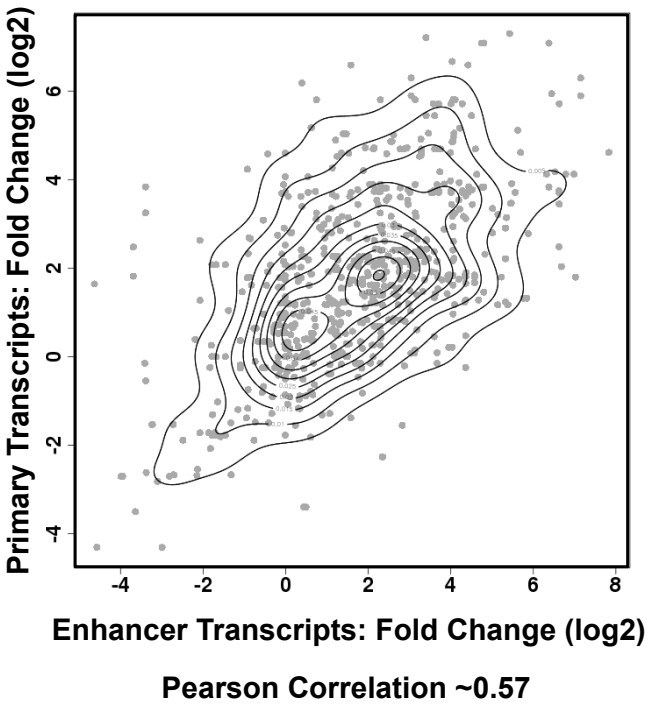


Figure 4.11. Correlation between ERα enhancer transcription and primary transcript regulation. Pearson correlation analysis was applied to determine correlation between the fold changes in ERα enhancer transcription and the fold changes in E2-regulated primary transcripts.

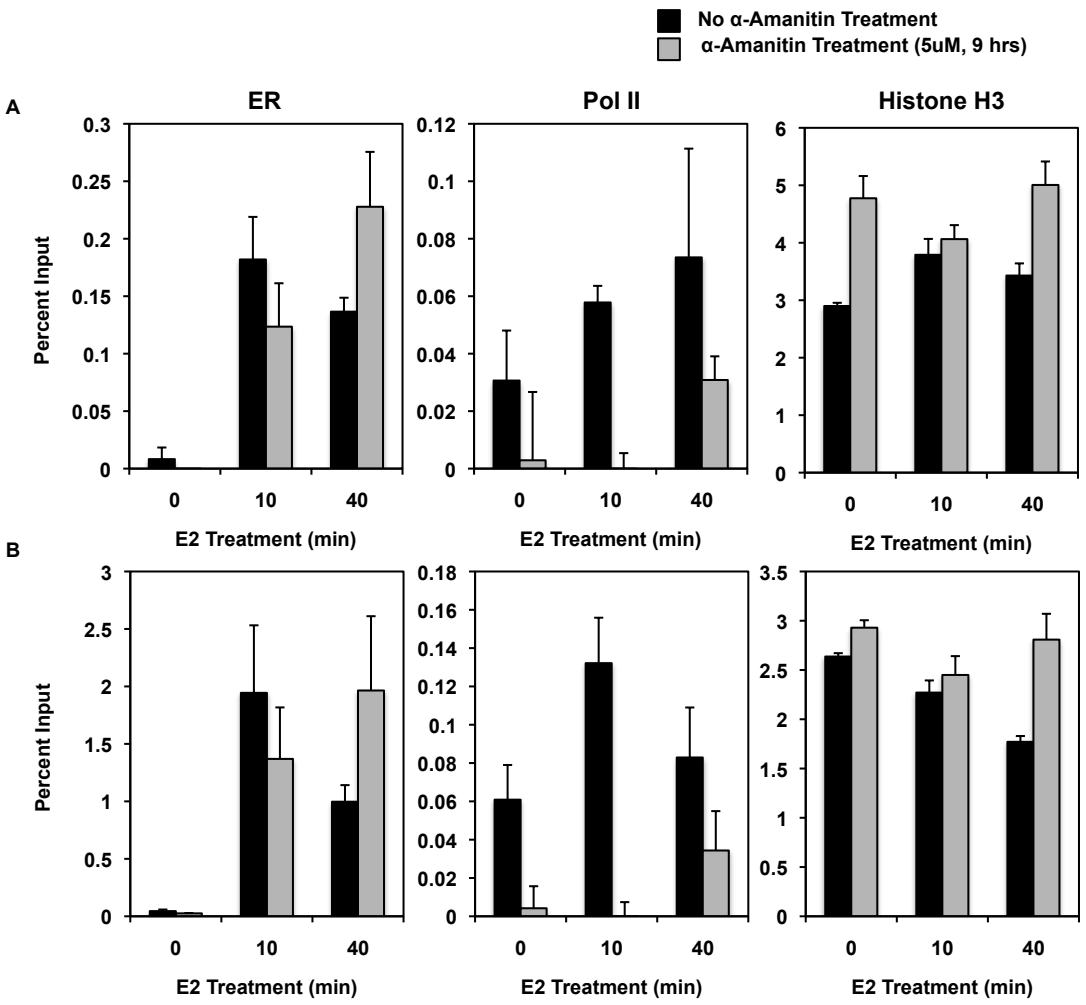


Figure 4.12. The functional connection between transcripts generation and ERα binding at the enhancers. The effects of transcript generation on ERα binding and the level of histones at enhancers were tested by ChIP-qPCR analyses. Each bar = mean + SEM, n = 3.

4.5. Method and Materials

The GRO-seq data sets. The GRO-seq data sets were used from our previous studies in Chapter 3. The data set was used to analyze specifically ERα enhancers.

Calling regulation in ERα enhancer transcription. Regulation of enhancer transcription was determined using edgeR (v.1.4.1) for R/Bioconductor using the same method and settings described in Chapter 3. Briefly, reads mapping to +1 to +13kb of each transcript prediction (mapped using the HMM described in Chapter 3) were passed as input to the edgeR program. We filtered enhancer transcripts to a 0.1% false discovery rate along with other transcripts in the genome. Following this procedure, transcripts that were annotated as "Enhancer transcription" using the annotation engine were collected and analyzed separately.

Correlation between ERα enhancer and RNA polymerase II. We looked at the Pearson correlation between RNA Pol II ChIP-seq and ERα ChIP-seq following 60 minutes of E2 at 3,191 ERα enhancers located >10 kb from the nearest gene. At each enhancer, ChIP-seq reads were counted in a window of 4 kb (Pol II) or 2 kb (ERα) centered on the location of maximum read depth in the previously reported peaks. Results are robust to a range of window sizes for each Pol II and ERα. Pearson correlation and scatterplots were generated using R.

ChIP analysis Chromatin immunoprecipitation (ChIP) assays. ChIP-qPCR assays in MCF-7 cells were performed as described previously. Briefly, cells were grown to ~80 to 90% confluence, cross-linked with 1% formaldehyde in PBS for 10 min. at 37°C, and quenched in 125 mM glycine in PBS for 5 min. at 4°C. The cells were

collected by centrifugation and sonicated in Lysis Buffer [Tris•HCl pH 7.9, 10 mM EDTA, 50 mM NaCl, 1% SDS, and 1x protease inhibitor cocktail (Roche, cat no. 04 693 124 001)] to generate chromatin fragments of ~500 bp in length. The material was clarified by centrifugation, diluted 10-fold in Dilution Buffer (20 mM Tris•HCl pH 7.9, 2 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 1x protease inhibitor cocktail), and precleared by incubation with protein A-agarose beads. The pre-cleared, chromatincontaining supernatant was used in immunoprecipitation reactions with antibodies against the factor of interest, or without antibodies as a control. The immunoprecipitated genomic DNA was cleared of protein and residual RNA by digestion with proteinase K and RNase H, respectively. The DNA was then extracted with phenol:chloroform:isoamyl alcohol and precipitated with ethanol. RT-qPCR was used to determine the enrichment of immunoprecipitated material relative to the input material using gene-specific primer sets to the specified regions (see primer sequences listed below). Each ChIP experiment was conducted a minimum of three times with independent chromatin isolates to ensure reproducibility.

α-Amanitin treatments. Prior to E2 treatment, the cells were treated with 5 μM of α-amanitin for the indicated times. E2 was then added in the presence of α-amanitin for 0, 10 and 40 min.

Primers for ChIP analyses

Enh-40 max1 Fwd: 5' - AGGGACAATCTGCCCTGTAA - 3'

Enh-40-max1 Rev: 5' - ATTTGAACCCAGGCACATTC - 3'

Enh-40-max2 Fwd: 5' - CCAACTTAGCCACACTGGAA - 3'

Enh-40-max2 Rev: 5' - CTGGGTGACACGTGTTAGGT - 3'

Enh-10-max1 Fwd: 5' - GGGCCTTGAATTCTCCTG - 3'

Enh-10-max1 Rev: 5' - TAAGTGGAGGGACACCCAAT - 3'

Enh-10-max2 Fwd: 5' - TCAGACATGCCTTTCATTCC - 3'

Enh-10-max2 Rev: 5' - TGGGACCAACATCTAGAGCA - 3'

Enh-160-max1 Fwd: 5' - CACTGTCAATGTCCCAATGC - 3'

Enh-160-max1 Rev: 5' - CCCAGTGGGTAGCAGGTAAA - 3'

Enh-Down-Fwd: 5' - TTCTTACAAAGGCGCTGCTA - 3'

Enh-Down-Rev: 5' - CTGGGATGAGCCTGTCACTA - 3'

REFERENCES

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.

Cheng, A.S., Jin, V.X., Fan, M., Smith, L.T., Liyanarachchi, S., Yan, P.S., Leu, Y.W., Chan, M.W., Plass, C., Nephew, K.P*., et al.* (2006). Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters. Mol Cell *21*, 393-404.

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

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.

Eeckhoute, J., Lupien, M., Meyer, C.A., Verzi, M.P., Shivdasani, R.A., Liu, X.S., and Brown, M. (2009). Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. Genome Res *19*, 372-380.

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.

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.

Jin, V.X., Leu, Y.W., Liyanarachchi, S., Sun, H., Fan, M., Nephew, K.P., Huang, T.H., and Davuluri, R.V. (2004). Identifying estrogen receptor alpha target genes using integrated computational genomics and chromatin immunoprecipitation microarray. Nucleic Acids Res *32*, 6627-6635.

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.

Kwon, Y.S., Garcia-Bassets, I., Hutt, K.R., Cheng, C.S., Jin, M., Liu, D., Benner, C., Wang, D., Ye, Z., Bibikova, M.*, et al.* (2007). Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. Proc Natl Acad Sci U S A *104*, 4852-4857.

Laganiere, J., Deblois, G., and Giguere, V. (2005). Functional genomics identifies a mechanism for estrogen activation of the retinoic acid receptor alpha1 gene in breast cancer cells. Mol Endocrinol *19*, 1584-1592.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood *100*, 3077-3086.

Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958-970.

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.

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.

Tsai, W.W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.Y., Shi, X., Schwarzer, D., Plunkett, W.*, et al.* (2010). TRIM24 links a non-canonical histone signature to breast cancer. Nature *468*, 927-932.

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.

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.

APPENDIX 1

Global Analysis of the Membrane-Initiated Transcriptional Effects of Estrogens

Dr. Charles Danko developed the bioinformatic and computation approaches used to analyze the GRO-seq data. His key contributions are represented in Fig. A1.2.

A1.1. Summary

To investigate the immediate direct effect of extranuclear roles of estrogen, characterize, I applied Global Run-On and sequencing (GRO-seq) method to on global transcriptional regulation using estrogen dendrimer conjugates (EDCs), which specifically initiate membrane actions of estrogens. Interestingly, unlike our previous observation that short estrogen treatments can up and down regulate significant numbers of genes, short treatments of EDCs did not changes much in gene regulation although the membrane-initiated kinase cascade signaling pathways can be activated at these time points. Collectively, our results and further analyses will provide the most comprehensive measurement of the immediate direct estrogen effects to date and a resource for understanding rapid signal-dependent transcription in other systems.

A1.2. Introduction

Estrogens are a class of endogenous steroidal hormones that are required for the normal development and function of the female reproductive organs and mammary glands (Deroo and Korach, 2006). Estrogens also act as potent mitogens in breast cancer initiation and progression (Platet et al., 2004). The actions of estrogens are mediated by two estrogen receptor (ER) isoforms, ERα and ERβ (Leclercq, 2002). ERα functions primarily as a nuclear transcription factor, which dimerizes upon the binding of its natural ligand, 17β-estradiol (E2), and acts as a potent regulator of gene expression. The nuclear actions of ERα occur through two distinct molecular pathways: (1) direct binding to DNA sequence motifs called estrogen response elements (ERE) (Klinge et al., 2004) and (2) tethering through other DNA-bound transcription factors (e.g., including AP-1, or c-Fos and C-Jun heterodimers) and

modulating their activity (Cheung et al., 2005). These same ERs, however, are also found in the cytoplasm and at the plasma membrane, where they can mediate "membrane initiated" or "extranuclear" cellular signaling responses through kinase cascades (Levin, 2005; Madak-Erdogan et al., 2008; Song, 2007). Other proteins associated with the plasma membrane may also bind estrogens and mediate estrogen responses (e.g., the orphan G protein-coupled receptor, GPR30) (Mizukami, 2010), although these are more controversial. The end result of both the nuclear and membrane initiated estrogen signaling pathways is the regulation of specific gene sets across the genome. To characterize the set of membrane-initiated ER target genes, previous studies have used steady-state gene expression patterns in the presence and absence of E2 as well as estrogen-dendrimer conjugates (EDC). However, the use of expression microarrays with a long ligand time leaves the possibility that the regulation of target genes are mediated by secondary response of nuclear and extranuclear actions considering our observations of global and immediate effects of estrogen. To have a better understanding, I applied GRO-seq method to investigate the immediate direst effect of extranuclear roles of estrogen on global transcriptional regulation using EDCs.

A1.3. Results and Discussion

A1.3.1. Generation of GRO-seq libraries from estrogen- and estrogen-dendrimer conjugate (EDC)-treated MCF-7 cells

To investigate the immediate effects of nuclear and extranuclear actions of estrogen on the transcriptome of human cells, we treated ERα-positive MCF-7 human breast cancer cells with a short time course of 17β-estradiol (E2) or estrogen-dendrimer conjugates (EDC) (0, 10 and 40 min) (Fig. A1.1). Nuclei were isolated

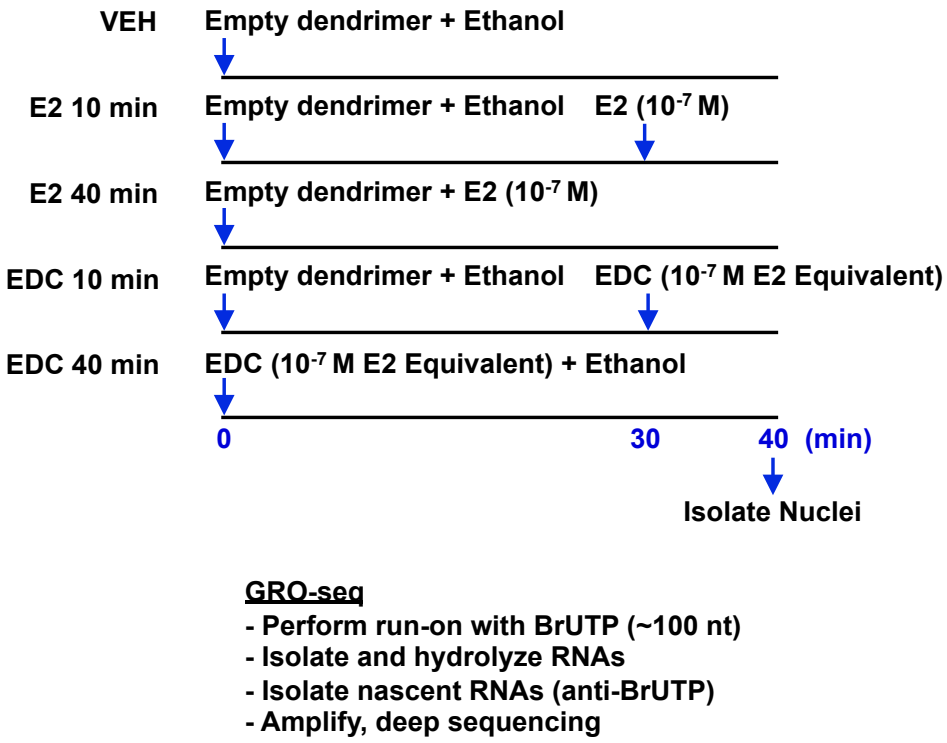


Figure A1.1. Overview of experimental set up and conditions for GRO-seq analysis using MCF-7 cells. MCF-7 cells were plated estrogen-deprived media for three days prior to ligand treatments. MCF-7 cells were treated with vehicle (ethanol and EDC equivalent empty dendrimer), 10^{-7} M E2 with empty dendrimer, or 10^{-7} M E2 equivalent EDC. Nuclei isolations and GRO-seq procedures were performed in the same condition as described previously.

Table A1. Summary of short-reads alignment to the human reference genome.

	Total Reads	Uniquely mappable reads	Fraction of reads mapped in rRNA	Fraction of reads mapped in Chr
VEH	48452911	22172633	0.25	0.75
E2 10 min	48975416	25538017	0.26	0.74
E2 40 min	51086044	27395433	0.30	0.70
EDC 10 min	48608001	25777624	0.27	0.73
EDC 40 min	49013120	23829134	0.31	0.69

from two biological replicates of the E2-treated MCF-7 cells and subjected to the GRO-seq procedure to generate ~100 bp libraries representing nascent RNAs, which were sequenced using an Illumina Genome Analyzer (Fig. A1.1). Short-reads were aligned to the human reference genome (hg18, NCBI36), including autosomes, the X-chromosome, and one complete copy of an rDNA repeat (GenBank ID: U13369.1) (Table A1.1). Approximately 22-26 million reads were uniquely mapped to the genome for each treatment condition and the biological replicates for each time point were highly correlated (average correlation coefficient = 0.98).

A.1.3.2. Determining the regulation of protein coding transcripts by E2 and EDC regulation.

We determined which of the annotated protein coding genes change in response to E2 or EDC using a recently described model-based approach. This approach fits a negative binomial model to read counts across all genes between our two biological replicates, and subsequently uses an exact test to identify genes whose change between two treatment conditions is beyond the global level of variation (Fig. A1.2). We chose to focus our analysis on a 12 kb window at the 5' end of each gene, as we expect to observe changes during the first 10 minutes in this window that will not yet have spread to the 3' end of longer transcripts. As we expected from the previous study described in Chapter 3, a large number of RefSeq gene are significantly regulated at a short 10 and 40 min E2 treatment (up- or down-regulated relative to the control/untreated condition) at a false discovery rate of 0.1%. On the other hand, surprisingly, EDC treatment for 10 min and 40 min did not change transcription of protein coding gene compare to E2 treatment. At 0 min EDC treatment, no genes are found to be regulated, indicating that all the protein coding genes that are regulated by 10 min E2 treatment is the direct targets of estrogens. At a 40 min EDC treatment

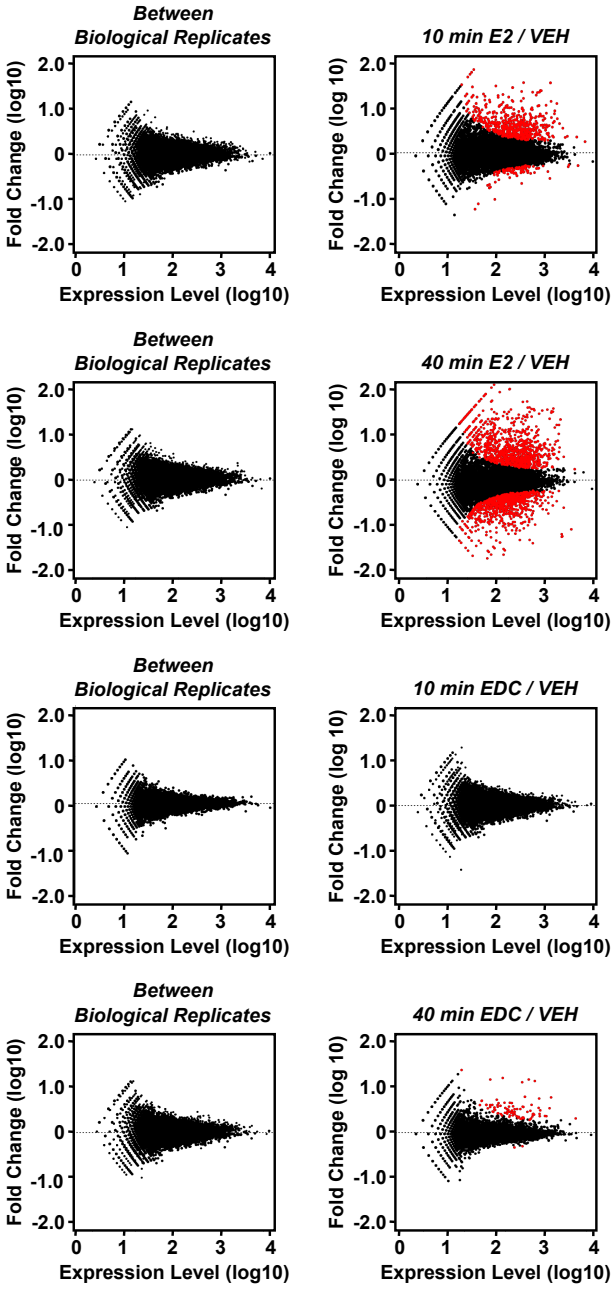


Figure A1.2. Determining E2 regulation of transcripts. Plots depicting the fold change of genes as a function of expression between two biological replicates (left) or between two treatment conditions (right). Red points indicate genes that fall outside of the expected variation are called "regulated" by E2 or EDC for the treatment time point shown.

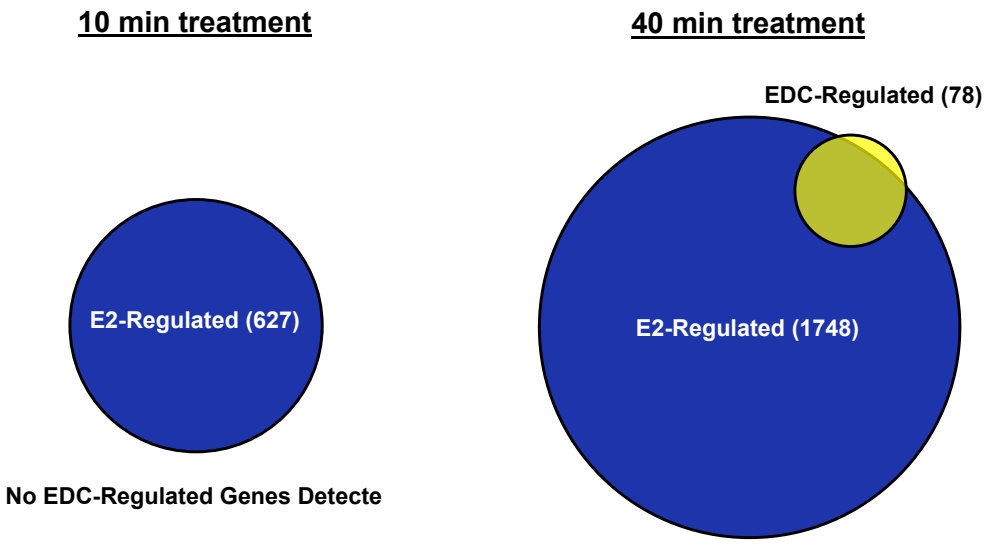


Figure A1.3. Distinguishing nuclear and extranuclear actions of estrogens. Venn diagram indicating the overlap of the genes that are regulated by 40 min E2 treatment and 40 min EDC treatment. The number of genes that are regulated by E2 or EDC at 40 min treatment is determined by 12 kb window (1 kb – 13 kb) at the 5' end of each gene as previously described in Chapter 3.

condition, total 79 genes are significantly regulated by EDC and most of these genes (71 genes) are regulated by 40 min E2 treatment (Fig. A1.3), indicating these genes are likely to be regulated through membrane initiated actions rather than by ER direct binding, or both. These genes showing regulation at 10 or 40 minutes by E2 treatment represent the most comprehensive and accurate definition of the immediate, direct transcriptional targets of the estrogen signaling pathway described to date.

A1.4. Future Directions

Over the past three decades, the mechanisms of ER-dependent transcriptional regulation have been intensively studied. Although, these studies have elucidated and provided of information of estrogen biology, several fundamental questions remain to be addressed in the field. One of the biggest tasks is to define the direct targets of estrogen signaling pathway in human cells. Here, I used an innovative new method called Global Nuclear Run-On and Massively Parallel Sequencing to provide a map of estrogen-regulated transcriptome. In addition to defining all the estrogen-regulated genes in our first study, I also used EDC to further define and distinguish nuclear and extranuclear actions of estrogen signaling. These data set will require further analyses that are not only limited to protein coding genes, but also expanded to any transcripts (i.e. rRNAs, tRNA, miRNA, ncRNAs, etc.) that are regulated by Pol I, II and III. Using the unbiased transcripts calling methods will further empower this data set to provide more comprehensive estrogen-regulated transcriptome. Also, it is important to further investigate how transcripts can be turned on and off upon short E2 treatments as well as how ER binding sites-lacking transcription start sites can be regulated through nuclear estrogen action. Defining the direct estrogen targets, estrogen binding sites, as well as other transcription factors that are involve in any

estrogen-regulated transcription will enable to provide more comprehensive pictures of estrogen signaling pathway, which will be utilized for developing hormone therapy as well as anti-estrogen drugs.

A1.5. Methods

Cell culture. 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) with Hank's salts (Sigma) supplemented with 5% calf serum (CS), sodium bicarbonate, penicillin/streptomycin and gentamicin.

EDC and E2 treatments. Cells were plated for experiments in phenol red-free MEM (Sigma) supplemented with 5% charcoal-dextran treated calf serum (CDCS) prior to ligand treatments. For control, vehicles of E2 (0.1 % ethanol) and EDC (empty dendrimer that is equivalent of EDC) were treated for 40 min. For 40 min E2 treatment, the same concentration of empty dendrimer was treated with 10-7 M E2. For 40 min EDC, 0.1 % of ethanol was treated with EDC equivalent with E2 concentration. For 10 min treatments, prior to E2 or EDC treatments, cell were treated with vehicles for 30 min.

GRO-seq. GRO-seq was performed as described previously in Section 3.4. Material and Methods in Chapter 3.

GRO-seq data analyses. Short-read alignments and mappable regions were determined as described in Section 3.4. Material and Methods in Chapter 3. Also,

calling regulation was based on a negative binomial model to read counts across all genes between our two biological replicates, and subsequently uses an exact test to identify genes whose change between two treatment conditions is beyond the global level of variation, which is also described in detail in Section 3.4. Material and Methods in Chapter 3.

Acknowledgements

I would like to thank Dr. John A. Katzenellenbogen (University Of Illinois At Urbana-Champaign) for providing EDCs.

REFERENCES

Cheung, E., Acevedo, M.L., Cole, P.A., and Kraus, W.L. (2005). Altered pharmacology and distinct coactivator usage for estrogen receptor-dependent transcription through activating protein-1. Proc Natl Acad Sci U S A *102*, 559-564.

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

Klinge, C.M., Jernigan, S.C., Mattingly, K.A., Risinger, K.E., and Zhang, J. (2004). Estrogen response element-dependent regulation of transcriptional activation of estrogen receptors alpha and beta by coactivators and corepressors. J Mol Endocrinol *33*, 387-410.

Leclercq, G. (2002). Molecular forms of the estrogen receptor in breast cancer. J Steroid Biochem Mol Biol *80*, 259-272.

Levin, E.R. (2005). Integration of the extranuclear and nuclear actions of estrogen. Mol Endocrinol *19*, 1951-1959.

Madak-Erdogan, Z., Kieser, K.J., Kim, S.H., Komm, B., Katzenellenbogen, J.A., and Katzenellenbogen, B.S. (2008). Nuclear and extranuclear pathway inputs in the regulation of global gene expression by estrogen receptors. Mol Endocrinol *22*, 2116-2127.

Mizukami, Y. (2010). In vivo functions of GPR30/GPER-1, a membrane receptor for estrogen: from discovery to functions in vivo. Endocr J *57*, 101-107.

Platet, N., Cathiard, A.M., Gleizes, M., and Garcia, M. (2004). Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. Crit Rev Oncol Hematol *51*, 55-67.

Song, R.X. (2007). Membrane-initiated steroid signaling action of estrogen and breast cancer. Semin Reprod Med *25*, 187-197.

APPENDIX 2

Mapping of the Estrogen-Regulated Phosphoproteome Network

Dr. Marcus Smolka performed the mass spectrometry analysis.

A2.1. Summary

Recent studies have suggested links between both the nuclear and membrane-initiated estrogen signaling pathways to kinase-dependent cellular signaling pathways. This interplay is likely to lead to extensive changes in the phosphoproteome upon estrogen treatment. To investigate the changes in phosphoproteomes, I applied Mass spectrometry (MS) in combination with SILAC (Stable isotope labeling with amino acids in cell culture) and IMAC (immobilized metal affinity chromatography) for the quantitative measurement of changes in phosphorylation conditions. In this study, the preliminary results proved the experimental system is sensitive to detect changes in site-specific phosphorylation in a very quantitative manner. This will yield a great resource to understand estrogen-regulated phosphoproteome.

A2.2. Introduction

The cellular actions of estrogens are mediated through estrogen receptors (ERs), ERα and ERβ, which bind estrogens with high affinity. ERs regulate transcription of estrogen-responsive target genes through direct binding of liganded ERs to EREs and indirect association of liganded ERs with AP-1-responsive elements via heterodimers of Fos and Jun (Cheung et al., 2005; Klinge, 1999). ERs are also found in the cytoplasm and at the plasma membrane, where they can also play a role in mediating membrane-initiated cellular signaling responses (Levin, 2005; Madak-Erdogan et al., 2008; Song, 2007).

Growing evidence has linked both the nuclear and membrane-initiated estrogen signaling pathways to kinase-dependent cellular signaling pathways

(Aksamitiene et al., 2010; Harrington et al., 2006; Madak-Erdogan et al., 2011; Wong et al., 2002). For example, ERα is a target for phosphorylation by various cellular kinases, including mitogen activated protein (MAP) kinases, protein kinase A, and the p90 ribosomal S6 kinase (Clark et al., 2001). In addition, ERα coregulators, such as SRC-3/AIB1, are also targets for phosphorylation (Feng et al., 2006; Zheng et al., 2005). Furthermore, estrogen signaling stimulates a variety of kinase-dependent pathways on a rapid time scale and in a manner that does not require genomic estrogen actions. Recent studies have suggested an interplay between ERα and signaling kinases at the promoters of estrogen target genes (Madak-Erdogan et al., 2011). This interplay is likely to lead to extensive changes in the phosphoproteome upon estrogen treatment.

To investigate the changes in phosphoproteomes, I applied Mass spectrometry (MS) to analyze thousands of sites of phosphorylation in MCF-7 breast cancer cells. To further allow quantitative analysis using MS, I also used SILAC (Stable isotope labeling with amino acids in cell culture) for the incorporation of stable isotopes into peptides. The use of isotopic labeling allows the comparison of two different samples and quantification of the changes in the abundance of different phosphopeptides under different experimental conditions. The low abundance of phosphopeptides and low degrees of phosphorylation, however, typically necessitate the isolation and concentration of phosphopeptides prior to MS analysis. The enrichment of phosphopeptides with immobilized metal affinity chromatography (IMAC) is a good approach to reduce the complexity of proteome sample and decrease the dynamic range. I used SILAC coupled with multidimensional chromatography technology, including IMAC, to map the E2-responsive phosphoproteomes of MCF-7 cells (Fig A2.1).

A2.3. Results and Discussion

A2.3.1. Optimization of MCF-7 cell growth, incorporation of stable isotope labeled amino acids, and ligand treatment condition for SILAC.

In order to incorporate heavy amino acids during cell culture, cells for experimental (E2 treated) group are grown in heavy amino acids (i.e., $^{13}C_6/^{15}N_2$ L-Lysine + $^{13}C_6/^{15}N_2$ L-Arginine) and cells in the control group in the light amino acid analogs. The use of double label (i.e., $^{13}C_6$ and $^{15}N_2$) and two different amino acids increases the differences between the masses of the peptides in the control and experimental groups, allowing for the best separation.

First, I have performed the MCF-7 cell proliferation assays (Fig. A2.2) to measure the time that is required for doubling 6 times. Typically, the most efficient incorporation of amino acid can be reached through approximately 6 doubling times. In case of MCF-7 cells, the amount of time that is required for this is approximately 8-9 days.

Second, I also titrated the amount of amino acids to maximize the incorporation efficiency of heavy amino acids in MCF-7 cells. Especially, it has been shown that excessive amount of Arginine in cells can be converted into Proline (Bendall et al., 2008; Bicho et al., 2010; Van Hoof et al., 2007), which can lower the incorporation efficiency for labeling Arginine. By titrating arginine from 25 mg/L to 100 mg as well as Proline from 0 to 200 mg/L, I found that the incorporation efficiency was maximal by supplementing 100 mg/L of proline to SILAC media in addition to 50 mg/L arginine (data not shown).

Third, I further explore the optimal condition for estrogen treatment. To investigate the membrane initiated actions of estrogen, I first performed short time course treatment of estrogen (0, 5, 10, 15, 30 min. of E2 treatment) under the grown

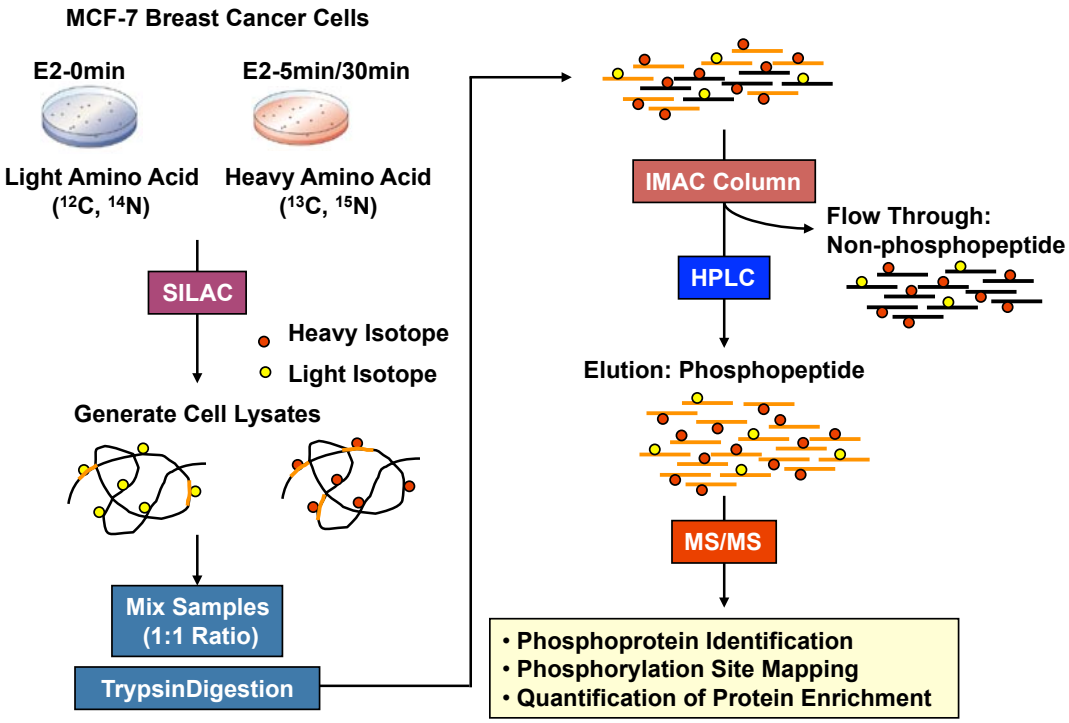


Figure A2.1. Flow chart of experimental set up

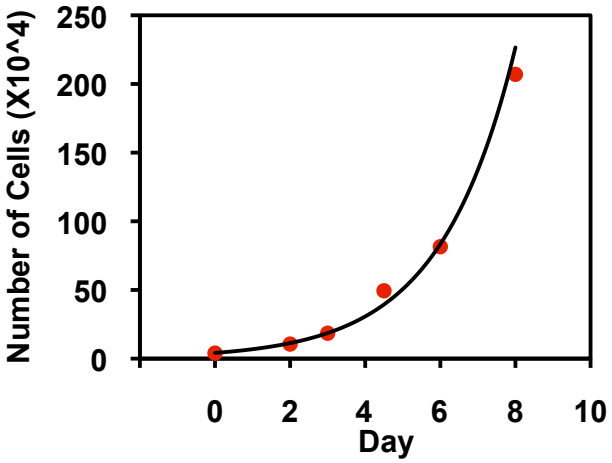


Figure A2.2. Optimization of MCF-7 cells growth condition. Measurement of MCF-7 doubling time by the rate of proliferation grown in SILAC medium with 5% dialyzed CDFBS. Cells were collected at the specified time point and counted.

conditions that I have optimized. Note that Erk1/2 phosphorylation is maximal at 5 min (Fig. A2.3), but other proteins might show a different time course of response.

A2.3.2. Phosphopeptide analysis by mass spectrometry

Whole cell extracts (WCEs) using 0.5 M NaCl were prepared from the treated cells. Volumes of WCE from cells in each treatment group were combined with a volume of extract from the corresponding control group containing equal amounts of total protein. Note that the different labels will allow the peptides derived from the control and experimental conditions to be distinguished. The combined mixture were now diluted to 150 mM NaCl and then digested with trypsin as described previously. Phosphopeptides were purified by IMAC and then fractionated by hydrophilic interaction chromatography (HILIC), as described previously. The HILIC fractions were analyzed using an LTQ-Orbitrap ion trap mass spectrometer. For data analysis, MS/MS spectra were searched by the SEQUEST algorithm (using the Sorcerer system) against a composite database containing the human protein database (downloaded from the International Protein Index) and its reversed complement. Xpress software was used for quantification, as described.

Interestingly, we found total 2262 phosphopeptides, out of which, only 745 that generated good statistics. For example, I found that nuclear receptor-interacting protein 1 (NRIP1; IPI00010196) showed an increased phosphorylation in the estrogen treated sample. NRIP1 is one of the first proteins to be identified as a hormone-recruited cofactor and shown to exhibits a strong transcriptional repressive activity which involves several inhibitory domains and different effectors. It has a phosphorylation at serine 671 that increases approximately 7-fold (R.LNSphosPLLSNK.T) (Fig A2.4). This result is very convincing, mainly because

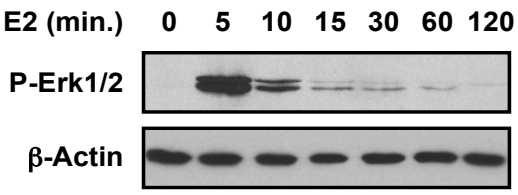


Figure A2.3. Optimization of E2 treatment condition for MCF-7 cells. Estrogen signaling stimulates Erk phosphorylation in MCF-7 human breast cancer cells. Western blotting for phospho-Erk1/2 showing the time course of E2 effects. β-actin is a loading control.

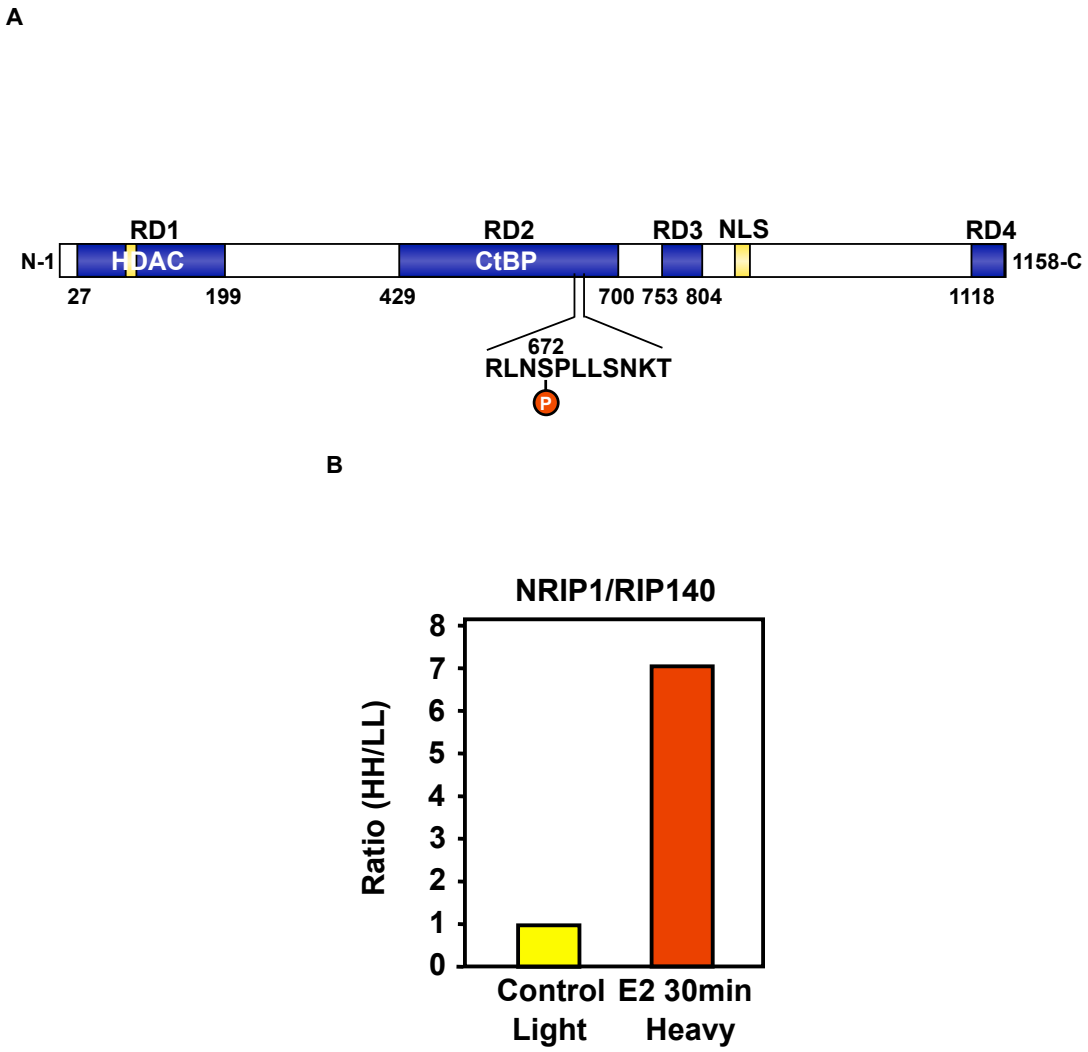


Figure A2.4. Increased phophorylation of NRIP1 at specific site. (A) Schematics of NRIP1 peptides showing a specific phosphorylation site at serine 671. (B) Approximately 7-fold increases in phosphorylation of NRIP1 in estrogen-treated sample.

another phosphorylation site in the same peptide does not change (AGSphosPINLSQHSLVIK), indicating that detected phophorylation is estrogen-specific modification which may be involved in a specific protein activity. Also the control shows that the fold change at serine 671 is not due to abundance change. Also, I found several CDKs that showed increased Y phosphorylation, which possibly due to the mitogenic effects of estrogen.

A2.4. Future Directions

Growing evidence has linked both the nuclear and membrane-initiated estrogen signaling pathways to kinase-dependent cellular signaling pathways. This interplay is likely to lead to extensive changes in the phosphoproteome upon estrogen treatment. Therefore, it is important to have a better understanding how estrogen affects kinase-dependent singaling pathways.

Here, I have showed the potential for identification of phosphoproteins in a very quantitative manner using SILAC and IMAC in combination with Mass spectrometry. However, in order to obtain high confident phosphoproteins as well as map the estrogen-regulated phosphoproteome, extensive analyses and further confirmations are necessary. First, to enhance the specificity of phosphoproteins, it will be more informative to separate the cellular proteins into nuclear and cytoplasmic fractions. Also, it will be interesting to include treatments with tamoxifen or an estrogen-dendrimer conjugate (EDC) that promotes membrane-initiated estrogen signaling as well as pathway-specific inhibitors. Second, to utilize the data set maximally, it is important to develop the bioinformatics tools that will enable to map and predict the phosphoproteome networks. Third, obtained phosphoproteins need to be confirmed by follow-up experiments such as Western

blotting with protein- and site-specific antibodies and cell-based assays to investigate biological relevance.

For the two decades, the ERα status of breast cancer cells has become an important biomarker of cancers that will respond well to endocrine therapies. ERα levels, however, are not necessarily an accurate representation of the levels of estrogen signaling activity in cells. Using the estrogen-regulated phosphoproteome as an indicator of estrogen signaling activity will be very useful in this regard.

A2.5. Methods

Cell Culture. MCF-7 human breast adenocarcinoma cells were kindly provided by Dr. Benita Katzenellenbogen (University of Illinois, Urbana-Champaign). The cells were maintained in RPMI1640 with 5% FBS, penicillin/streptomycin and gentamicin. For the SILAC analyses, the cells are maintained in RPMI1640 medium that lacks arginine, leucine, lysine, and phenol red (Sigma) supplemented with 5% dialyzed FBS (sigma), penicillin/streptomycin, 50 mg/L of L-proline, 19.3 mg/L of L-leucine, and either (1) 400 μM of "light" L-lysine and 225 μM of L-arginine or (2) heavy isotope-labeled amino acids, 400 μM of L-lysine-$^{13}C_6$, $^{15}N_2$-HCl (Sigma-Aldrich, Inc./Isotec) and 225 μM of L-arginine-$^{13}C_6$, $^{15}N_2$-HCl (Sigma-Aldrich, Inc./Isotec) for three days. Then, the cells were passaged into 15 cm-plates under the same conditions with 5% charcoal dextran-stripped dialyzed FBS for three days prior to E2 treatment.

Generation of whole cell extracts. The cells were collected, washed three times with ice-cold PBS, lysed with 1 % NP40 containing lysis buffer [1 % NP40, 20 mM Tris-HCl (pH 7.9), 10% (v/v) glycerol, 150 mM NaCl, 1 mM EDTA, 1 mM DTT,

0.2 mM phenylmethylsulfonylfluoride (PMSF), 1 mM benzamidine, 2 μg/ml leupeptin, 2 μg/ml aprotinin, 2 μg/ml pepstatin, phophoprotease inhibitor (Roche)] and incubated on ice for 15 minutes. The cells were further disrupted by Dounce homogenization ~15 stokes with a tight pestle, then supernatants were collected by centrifugation.

Proteomic Analyses

Trypsinization of Peptide. Heavy and light amino acid incorporated whole cell extracts were combined with 1:1 ratio (1 mg each). Then, samples were incubated for 5 minutes in the presence of 1 % SDS, 5 mM DTT, 50 mM Tris-Cl (pH8.0), then 15mM iodoacetamide were further added to samples and incubated for 15min at room temperature. The samples were precipitated by adding 3 volumes of mixture of [50 % Acetone: 50 % EtOH: 0.1 % HAc] on ice for 15 minutes. The samples were spun at 4,700 rpm for 5min and remove supernatant. The pellets were resuspended in 750 μl of dH_2O and 7.5 μl of urea resuspension buffer [8 M Urea and 100 mM Tris-Cl (pH8.0)] and resuspended pellet, then, cetrifugated for 10 min at 13,000 rpm at RT. The pellets are again resuspended in urea resuspension buffer with in the presence of 150 mM NaCl, followed by douncing. Samples were then treated with 10 μg of TPCK-treated trypsin per 1 mg of proteins and incubated . overnight at 37°C. In order to stop the reaction, 0.2% Formic Acid and 0.2% of TFA were added.

Sample clean-up. Trypsinized samples were spud down at 4500rpm for 5min and loaded onto C18-SepPack column that are pre-conditioned with C18-buffer D [80 % CAN, 0.1% HAc] and and C18-buffer A [0.1% TFA], follows by washinig the column C18-buffer C [0.1% HAc]. The bound peptides were eluted from each C18 column in a glass tube with C18-buffer D in a glass vial. The eluted peptides

were dried in a speed-vac at 45 °C. The dried peptimdes were resuspend with 1 % acetic acid to be subjected onto IMAC.

Purification of phosphopeptides by IMAC. The supernatants were loaded IMAC column, followed by washing once with 0.6% HAc, twice with 0.6% Acetonitrile, 0.1M NaCl, 0.1% HAc, and once with deionized water. Then, the peptides were eluted, dried in speed-vac, and resuspended with C18 buffer A with 0.1 pmol/μL of phosphorylated angiotensin II.

HILIC fractionation and LC-MS/MS. The samples were further supjected into HILIC fractionation and analyzed by LC-MS/MS as described previsouly.

REFERENCES

Aksamitiene, E., Kholodenko, B.N., Kolch, W., Hoek, J.B., and Kiyatkin, A. (2010). PI3K/Akt-sensitive MEK-independent compensatory circuit of ERK activation in ER-positive PI3K-mutant T47D breast cancer cells. Cell Signal *22*, 1369-1378.

Bendall, S.C., Hughes, C., Stewart, M.H., Doble, B., Bhatia, M., and Lajoie, G.A. (2008). Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. Mol Cell Proteomics *7*, 1587-1597.

Bicho, C.C., de Lima Alves, F., Chen, Z.A., Rappsilber, J., and Sawin, K.E. (2010). A genetic engineering solution to the "arginine conversion problem" in stable isotope labeling by amino acids in cell culture (SILAC). Mol Cell Proteomics *9*, 1567-1577.

Cheung, E., Acevedo, M.L., Cole, P.A., and Kraus, W.L. (2005). Altered pharmacology and distinct coactivator usage for estrogen receptor-dependent transcription through activating protein-1. Proc Natl Acad Sci U S A *102*, 559-564.

Clark, D.E., Poteet-Smith, C.E., Smith, J.A., and Lannigan, D.A. (2001). Rsk2 allosterically activates estrogen receptor alpha by docking to the hormone-binding domain. EMBO J *20*, 3484-3494.

Feng, Q., Yi, P., Wong, J., and O'Malley, B.W. (2006). Signaling within a coactivator complex: methylation of SRC-3/AIB1 is a molecular switch for complex disassembly. Mol Cell Biol *26*, 7846-7857.

Harrington, W.R., Kim, S.H., Funk, C.C., Madak-Erdogan, Z., Schiff, R., Katzenellenbogen, J.A., and Katzenellenbogen, B.S. (2006). Estrogen dendrimer conjugates that preferentially activate extranuclear, nongenomic versus genomic pathways of estrogen action. Mol Endocrinol *20*, 491-502.

Klinge, C.M. (1999). Role of estrogen receptor ligand and estrogen response element sequence on interaction with chicken ovalbumin upstream promoter transcription factor (COUP-TF). J Steroid Biochem Mol Biol *71*, 1-19.

Levin, E.R. (2005). Integration of the extranuclear and nuclear actions of estrogen. Mol Endocrinol *19*, 1951-1959.

Madak-Erdogan, Z., Kieser, K.J., Kim, S.H., Komm, B., Katzenellenbogen, J.A., and Katzenellenbogen, B.S. (2008). Nuclear and extranuclear pathway inputs in the regulation of global gene expression by estrogen receptors. Mol Endocrinol *22*, 2116-2127.

Madak-Erdogan, Z., Lupien, M., Stossi, F., Brown, M., and Katzenellenbogen, B.S. (2011). Genomic collaboration of estrogen receptor alpha and extracellular signal-regulated kinase 2 in regulating gene and proliferation programs. Mol Cell Biol *31*, 226-236.

Song, R.X. (2007). Membrane-initiated steroid signaling action of estrogen and breast cancer. Semin Reprod Med *25*, 187-197.

Van Hoof, D., Pinkse, M.W., Oostwaard, D.W., Mummery, C.L., Heck, A.J., and Krijgsveld, J. (2007). An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. Nat Methods *4*, 677-678.

Wong, C.W., McNally, C., Nickbarg, E., Komm, B.S., and Cheskis, B.J. (2002). Estrogen receptor-interacting protein that modulates its nongenomic activity-crosstalk with Src/Erk phosphorylation cascade. Proc Natl Acad Sci U S A *99*, 14783-14788.

Zheng, F.F., Wu, R.C., Smith, C.L., and O'Malley, B.W. (2005). Rapid estrogen-induced phosphorylation of the SRC-3 coactivator occurs in an extranuclear complex containing estrogen receptor. Mol Cell Biol *25*, 8273-8284.