UNIVERSITY OF CALIFORNIA, SAN DIEGO

Understanding Enhancer Role in Transcriptional Response

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Daria Merkurjev

Committee in charge:

Professor Michael Rosenfeld, Chair
Professor Wei Wang, Co-Chair
Professor Xiang-Dong Fu
Professor Bing Ren
Professor Gene Yeo

2015

The Dissertation of Daria Merkurjev is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego
2015

DEDICATION

to my family and friends

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Dr. Michael Rosenfeld for his support and guidance throughout my graduate school career. I would also like to thank Dr. Wei Wang, Dr. Bing Ren, Dr. Gene Yeo, Dr. Xiang-Dong Fu for their support and fruitful committee meetings and discussions.

I would also like thank my collaborators with whom I had the opportunity to work with during my graduate school career. Chapter 1, in full, is a reprint of the material as it appears in Proc Natl Acad Sci USA 2015. Enhancer-bound LDB1 regulates a corticotrope promoter-pausing repression program. Zhang F, Tanasa B, Merkurjev D, Lin C, Song X, Li W, Tan Y, Liu Z, Zhang J, Ohgi KA, Krones A, Skowronska-Krawczyk D, Rosenfeld MG., 2015. Chapter 2, in full, is a reprint of the material as it appears in Proc Natl Acad Sci USA 2015. Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens. Jin C, Yang L, Xie M, Lin C, Merkurjev D, Yang JC, Tanasa B, Oh S, Zhang J, Ohgi KA, Zhou H, Li W, Evans CP, Ding S, Rosenfeld MG., 2014. Chapter 3, in full, is a reprint of the material as it appears in Nature 2013. Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG., 2013. Chapter 4, in full, is a reprint of the material as it appears in Cell 2014. Enhancer activation requires trans-recruitment of a mega transcription factor complex. Liu Z, Merkurjev D, Yang F, Li W, Oh S,

Friedman MJ, Song X, Zhang F, Ma Q, Ohgi KA, Krones A, Rosenfeld MG., 2014. Chapter 5, in full, is a reprint of the material as it appears in Nature 2013. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, Merkurjev D, Ohgi KA, Meng D, Zhang J, Evans CP, Rosenfeld MG., 2013. Chapter 6 includes unpublished work with Amir Gamliel, Bogdan Tanasa, and Skowronska-Krawczyk. Chapter 7, in full, is a reprint of the material as it appears in Molecular Cell 2015. Enhancer activation requires trans-recruitment of a mega transcription factor complex. Li W, Hu Y, Oh S, Ma Q, Merkurjev D, Song X, Zhou X, Liu Z, Tanasa B, He X, Chen AY, Ohgi K, Zhang J, Liu W, Rosenfeld MG., 2015.

I would like to say a special thanks to Alexey, Michelle, Dinh, Andreea, Garrett, Josh, Natalie, and many others for the laughs and good times we have shared. Finally, I would like to thank my family, my father, Alexander Merkurjev, my mother, Olga Merkurjev, my sister, Ekaterina Merkurjev, and my grandparents, whose never-ending and unconditional love, wisdom, and support has guided me throughout my life.

VITA

OBJECTIVE Full-time Graduate School Student: Bioinformatics and Systems Biology (Ph.D.)

I study recent and novel technologies produce biological data sets of ever-increasing resolution that reveal not only genomic sequences but also RNA and protein abundances, their interactions with one another, their subcellular localization, and the identity and abundance of other biological molecules.

EDUCATION

University of California, San Diego Doctor of Philosophy in Bioinformatics and Systems Biology December 2015

University of California, Los Angeles Bachelor of Science in Mathematics June 2010

AWARDS/HONORS

- UCLA College Honors Program
- UCLA FWC Scholarship
- Smith College Book Award

RESEARCH EXPERIENCE

Michael Geoff Rosenfeld Laboratory, UCSD August 2011- Present

Advisor (Thesis Advisor): Prof. Michael Geoff Rosenfeld (Biomedical Sciences Department)

Suresh Subramani Laboratory, UCSD June 2011- August 2011

Advisor: Prof. Suresh Subramani (Molecular Biology Department)

Characterized cellular signaling processes and gene networks that are involved in pexophagy in mammalian cells

Globally assessed autophagy specificity

Trey Ideker Laboratory, UCSD Mar 2011- June 2011

Advisor: Prof. Trey Ideker (Bioengineering Department)

Developed tools for integrating the gene association data with protein network information to identify the pathways underlying a patient's genotype

Developed computational tools that use a network to identify paths and the subnetworks of interactions underlying the observed set of genome-wide associations

Giovanni Paternostro Laboratory, UCSD Dec 2010- Mar 2011

Advisor: Prof. Giovanni Paternostro (Sanford Burnham Medical Research Institute)

Analyzed RA-differentiated hESC derivatives and kinases

Created a more efficient search algorithm as a framework for the optimization of drug combinations therapy

Nuno Bandeira Laboratory, UCSD Dec 2010- Mar 2011

Advisor: Prof. Nuno Bandeira (Computer Science Department)

Clarified the significance of variation of the results of mass

spectrometry
Deduced the confidence interval of the ratio of the spectra of the mass spectrometry
Deduced whether repeated measurements increase the confidence interval of the mass spectrometry peaks
Resolved the peptide-to-protein linear programming problem, and find the ratio quality measure that determines the protein ratio measurements

Guarav Arya Laboratory, UCSD Sep 2010- Dec 2010
Advisor: Prof. Guarav Arya (Bioengineering Department)
Developed high-resolution, coarse-grained models of DNA and histone octamer and to integrate them to yield a coarse-grained model of the nucleosome
Studied molecular mechanisms associated with ATP-dependent nucleosome remodeling, particularly how nucleosome sliding enhances DNA accessibility within the chromatin

LONI Laboratory, UCLA Sep 2009 – August 2010
Advisor: Prof. Paul Thompson
Rated the factors leading to MCI, AD, and control patients
Determined factors controlling the transport the MCI patients to AD

REU, Mathematics Department, UCLA Jun 2009 – Aug 2009
Advisor: Prof. Andrea Bertozzi
Used imaging techniques to focus on the areas of the brain mostly affected by AD

Mathematics Department, UCLA Jan 2009 – Jun 2009
Advisor: Prof. Luminita Vese
Determined the best pan-sharpening method in terms of the highest resolution (Research Experience for Undergraduates) REU,

Mathematics Department, UCLA Jun 2008 – Aug 2008
Advisor: Prof. Andrea Bertozzi
Evaluated different pan-sharpening methods for enhancing the resolution of multispectral images
Developed a modification of the IHS method that produces images with higher spectral quality

Biomathematics Department, UCLA Jan 2008 – Mar 2008
Advisor: Prof. Carol Newton
Modeled tumor progression
Modeled angiogenesis

PUBLICATIONS

[1] Zhang, F. et al (2014). Enhancer-bound LDB1 regulates a corticotrope promoter pausing repression program. (accepted Proc Natl Acad Sci U S A).

Substantial evidence supports the hypothesis that enhancers are critical regulators of cell type determination, orchestrating both positive and negative transcriptional programs; however, the basic mechanisms by which enhancers

orchestrate interactions with cognate promoters during activation and repression events remain incompletely understood. Here we report the required actions of the LIM domain binding protein, LDB1/CLIM2/NLI, interacting with the enhancer binding protein, ASCL1, to mediate looping to coding target gene promoters and target gene regulation in corticotrope cells. LDB1-mediated enhancer:promoter looping appears to be required for both activation and repression of these target coding genes. While LDB1-dependend activated genes are regulated at the level of transcriptional initiation, the LDB1-dependent repressed transcription units appear to be regulated primarily at the level of promoter pausing, with LDB1 regulating recruitment of MTA2, a component of the NuRD complex, on these negative enhancers, required for the repressive enhancer function. These results indicate that LDB1-dependent looping events can deliver repressive cargo to cognate promoters to mediate promoter pausing events in a pituitary cell type.

[2] Basnet, H. et al (2014). Tyrosine phosphorylation of histone H2A by CK2 kinase regulates transcriptional elongation. Nature. 516(7530):267–271.

Post-translational histone modifications have a critical role in regulating transcription, the cell cycle, DNA replication and DNA damage repair. The identification of new histone modifications critical for transcriptional regulation at initiation, elongation or termination is of particular interest. Here we report a new layer of regulation in transcriptional elongation that is conserved from yeast to mammals. This regulation is based on the phosphorylation of a highly conserved tyrosine residue, Tyr 57, in histone H2A and is mediated by the unsuspected tyrosine kinase activity of casein kinase 2 (CK2). Mutation of Tyr 57 in H2A in yeast or inhibition of CK2 activity impairs transcriptional elongation in yeast as well as in mammalian cells. Genome-wide binding analysis reveals that CK2α, the catalytic subunit of CK2, binds across RNA-polymerase-II-transcribed coding genes and active enhancers. Mutation of Tyr 57 causes a loss of H2B mono-ubiquitination as well as H3K4me3 and H3K79me3, histone marks associated with active transcription. Mechanistically, both CK2 inhibition and the H2A(Y57F) mutation enhance H2B deubiquitination activity of the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex, suggesting a critical role of this phosphorylation in coordinating the activity of the SAGA complex during transcription. Together, these results identify a new component of regulation in transcriptional elongation based on CK2-dependent tyrosine phosphorylation of the globular domain of H2A.

[3] Liu, Z. et al (2014). Enhancer activation requires trans-recruitment of a mega transcription factor complex. Cell. Volume 159, Issue 2.

Enhancers provide critical information directing cell-type-specific transcriptional programs, regulated by binding of signal-dependent transcription factors and their associated cofactors. Here, we report that the most strongly activated estrogen (E2)-responsive enhancers are characterized by trans-

recruitment and in situ assembly of a large 1–2 MDa complex of diverse DNA-binding transcription factors by ERα at ERE-containing enhancers. We refer to enhancers recruiting these factors as mega transcription factor-bound in trans (MegaTrans) enhancers. The MegaTrans complex is a signature of the most potent functional enhancers and is required for activation of enhancer RNA transcription and recruitment of coactivators, including p300 and Med1. The MegaTrans complex functions, in part, by recruiting specific enzymatic machinery, exemplified by DNA-dependent protein kinase. Thus, MegaTrans-containing enhancers represent a cohort of functional enhancers that mediate a broad and important transcriptional program and provide a molecular explanation for transcription factor clustering and hotspots noted in the genome.

[4] Jin, C. et al (2014). Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens. Proc Natl Acad Sci U S A.: 111(25):9235-40.

Understanding the mechanisms by which compounds discovered using cell-based phenotypic screening strategies might exert their effects would be highly augmented by new approaches exploring their potential interactions with the genome. For example, altered androgen receptor (AR) transcriptional programs, including castration resistance and subsequent chromosomal translocations, play key roles in prostate cancer pathological progression, making the quest for identification of new therapeutic agents and an understanding of their actions a continued priority. Here we report an approach that has permitted us to uncover the sites and mechanisms of action of a drug, referred to as “SD70,” initially identified by phenotypic screening for inhibitors of ligand and genotoxic stress-induced translocations in prostate cancer cells. Based on synthesis of a derivatized form of SD70 that permits its application for a ChIP-sequencing–like approach, referred to as “Chem-seq,” we were next able to efficiently map the genome-wide binding locations of this small molecule, revealing that it largely colocalized with AR on regulatory enhancers. Based on these observations, we performed the appropriate global analyses to ascertain that SD70 inhibits the androgen-dependent AR program, and prostate cancer cell growth, acting, at least in part, by functionally inhibiting the Jumonji domain-containing demethylase, KDM4C. Global location of candidate drugs represents a powerful strategy for new drug development by mapping genome-wide location of small molecules, a powerful adjunct to contemporary drug development strategies.

[5] Yang, L. et al (2013). lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature. 500(7464):598-602.

Although recent studies have indicated roles of long non-coding RNAs (lncRNAs) in physiological aspects of cell-type determination and tissue homeostasis1, their potential involvement in regulated gene transcription programs remains rather poorly understood. The androgen receptor regulates a

large repertoire of genes central to the identity and behaviour of prostate cancer cells2, and functions in a ligand-independent fashion in many prostate cancers when they become hormone refractory after initial androgen deprivation therapy3. Here we report that two lncRNAs highly overexpressed in aggressive prostate cancer, PRNCR1 (also known as PCAT8) and PCGEM1, bind successively to the androgen receptor and strongly enhance both ligand-dependent and ligand-independent androgen-receptor-mediated gene activation programs and proliferation in prostate cancer cells. Binding of PRNCR1 to the carboxy-terminally acetylated androgen receptor on enhancers and its association with DOT1L appear to be required for recruitment of the second lncRNA, PCGEM1, to the androgen receptor amino terminus that is methylated by DOT1L. Unexpectedly, recognition of specific protein marks by PCGEM1-recruited pygopus 2 PHD domain enhances selective looping of androgen-receptor-bound enhancers to target gene promoters in these cells. In 'resistant' prostate cancer cells, these overexpressed lncRNAs can interact with, and are required for, the robust activation of both truncated and full-length androgen receptor, causing ligand-independent activation of the androgen receptor transcriptional program and cell proliferation. Conditionally expressed short hairpin RNA targeting these lncRNAs in castration-resistant prostate cancer cell lines strongly suppressed tumour xenograft growth in vivo. Together, these results indicate that these overexpressed lncRNAs can potentially serve as a required component of castration-resistance in prostatic tumours.

[6] Li, W. et al (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 498(7455):516-20.

The functional importance or relevance of gene enhancers in regulated gene expression is well established. In addition to widespread transcription of long non-coding RNAs (lncRNAs) in mammalian cells, bidirectional ncRNAs are transcribed on enhancers, and are thus referred to as enhancer RNAs (eRNAs). However, it has remained unclear whether these eRNAs are functional or merely a reflection of enhancer activation. Here we report that in human breast cancer cells 17β-oestradiol (E2)-bound oestrogen receptor α (ER-α) causes a global increase in eRNA transcription on enhancers adjacent to E2-upregulated coding genes. These induced eRNAs, as functional transcripts, seem to exert important roles for the observed ligand-dependent induction of target coding genes, increasing the strength of specific enhancer–promoter looping initiated by ER-α binding. Cohesin, present on many ER-α-regulated enhancers even before ligand treatment, apparently contributes to E2-dependent gene activation, at least in part by stabilizing E2/ER-α/eRNA-induced enhancer–promoter looping. Our data indicate that eRNAs are likely to have important functions in many regulated programs of gene transcription.

[7] Rahmani, S. et al (2010). An adaptive IHS pan-sharpening method. IEEE Geoscience And Remote Sensing Letters. Volume 7, Issue 4.

The goal of pan-sharpening is to combine a low spatial resolution multispectral image with a higher resolution panchromatic image to obtain an image with high spectral and spatial resolution. The Intensity-Hue-Saturation (IHS) method is a popular pan-sharpening method used for its efficiency and high spatial resolution. However, the final image produced experiences spectral distortion. In this letter, we introduce two new modifications to improve the spectral quality of the image. First, we propose image-adaptive coefficients for IHS to obtain more accurate spectral resolution. Second, an edge-adaptive IHS method was proposed to enforce spectral fidelity away from the edges. Experimental results show that these two modifications improve spectral resolution compared to the original IHS and we propose an adaptive IHS that incorporates these two techniques. The adaptive IHS method produces images with higher spectral resolution while maintaining the high-quality spatial resolution of the original IHS.

PUBLICATIONS IN PRESS/REVIEW

[1] Gamliel et al (2014). The Rett Syndrome protein MeCP2 establishes nuclear matrix-associated boundaries regulating transcription (submitted to Cell).

[2] Wang, J et al (2014). A novel neuron-specific histone H4K20 demethylase LSD1n promotes transcriptional elongation and is essential for learning and memory (submitted to Nature).

[3] Scully, K. et al (2014). Cell-nonautonomous expression of beta1 integrin is required for developmental angiogenesis in the mouse pituitary gland (submitted).

[4] Li, W et al (2014). Condensins as regulators of ligand-dependent transcriptional activation of regulatory enhancers and coding genes (submitted to Molecular Cell).

[5] Till, A. et al (2014). Evolutionary trends and functional anatomy of the human expanded autophagy network (accepted in Autophagy).

CONFERENCES

- NextGene GeneExpression Analysis at San Diego Salk Institute September 2011
- California Summer Research Training Program Conference at UCLA August 2009

  Talk Title: Improving on the Prognosis of Alzheimer's Disease
- California Summer Research Training Program Conference at UCLA August 2008

  Talk Title: Evaluation of Pansharpening Methods
- Pacific Coast Undergraduate Math Conference at Loyola Marymount University Spring 2008
- Pacific Coast Undergraduate Math Conference at Pepperdine University Spring 2007
- Professor/Student UMSA Conference at UCLA Spring 2007 (*helped organize and participated in)

  Talk Title: Exotic Differentiation (presentation by Ciprian Manolescu)

EMPLOYMENT

- CSE 202, Teaching Assistant, January 2012- March 2012
- Los Angeles Math Circle, Teaching Assistant, 2 hrs/week Sep 2006- June 2010
- Assisted the organizer in running the Math Circle; developed research problems and ways to explain them to students; discussed mathematical problems during meetings; led students through a variety of mathematical topics. Presented extracurricular mathematical topics to schoolchildren

UCLA, Project 101 Tutors, Tutor, 2hrs/week Dec 2007- June 2010

RELEVANT COURSEWORK

- Principles of Artificial Intelligence
- Graduate Statistical Physics
- Biological Data Representation and Analysis
- Introduction to Bioinformatics Algorithms
- Statistical Methods in Bioinformatics
- Genomics, Proteomics, and Network Biology
- Linear Algebra and Applications
- Differential Equations
- Deterministic Models in Biology (Graduate)
- Stochastic Models in Biology (Graduate) Mathematical & Statistical Phylogenetics (Graduate)
- Basic Human Biology for Biomedical Engineers (Graduate)

EXTRACURRICULAR ACTIVITIES/ VOLUNTEER ACTIVITIES

Undergraduate Math Students Association, UCLA Sept 2008- Present

- Assisted at monthly meetings

William Putnam Mathematics Competition 2007

- A nationwide mathematics competition

Los Angeles Math Circle, Volunteer, 2 hrs/week. Sept 2007- June 2008

- Assisted the organizer in running the Math Circle; found research problems and ways to explain them to students; discussed mathematical problems during meetings; led students through a variety of mathematical topics. Presented extracurricular mathematical topics to schoolchildren

Faculty Women's Club Assistant Sept 2006- Present

- Assisted in the Newcomers' meeting

SCHOLARHIPS

- FWC UCLA Scholarship ($1000)

SPECIAL INFORMATION

- Computer Skills: Matlab, C++, FORTRAN, LaTeX, MS Word, Excel, R
- Languages: English (fluent proficiency), Russian (fluent proficiency), French (limited working proficiency)

- Other Computer Programs/Software Used: C++, R, VisSim, Virtual Cell, BMDP, Berkeley Madonna, Microsoft PowerPoint, Microsoft Excel, LAB Fit, Video Edit Magic 4.1, Chemix, Windows Movie Maker, Adobe Reader 7.0, Stats, Genious v2.5.4, Cytoscape, BioGRID, STRING, DAVID

PERSONAL REFERENCES
available upon request

ABSTRACT OF THE DISSERTATION

Understanding Enhancer Role in Transcriptional Response

by
Daria Merkurjev

University of California, San Diego, 2015

Doctor of Philosophy in Bioinformatics and Systems Biology

Professor Michael Rosenfeld, Chair
Professor Wei Wang, Co-Chair

In genetics, an enhancer is a short (50-1500 bp) region of DNA that can be bound with proteins (activators) to activate transcription of a gene or transcription. Despite their discovery more than 35 years ago, the fundamental principles by which enhancers are activated and regulate their coding gene transcriptional targets in metazoans have remained poorly understood. The molecular mechanisms responsible for orchestrating and integrating genome-wide transcriptional responses to diverse signaling pathways critical for developmental, physiological, and pathological regulation are still widely unknown. The Rosenfeld lab has made an effort to study these molecular mechanisms, focused on previously unsuspected aspects of enhancer function, chromosomal structure, and subnuclear architectural interactions. These strategies, which underlie genome-wide transcriptional responses in the endocrine and central nervous systems and are critical for physiological and

behavioral processes in all vertebrates, are orchestrated by the network of genomic enhancers. Our recent findings have substantially altered concepts regarding the roles of noncoding RNAs (ncRNAs), mechanisms of enhancer activation and function, and nuclear architecture as critical aspects of regulated gene expression programs. This work has uncovered unexpected aspects of enhancer function, highlighting their functioning as regulated transcription units in dynamic alterations in nuclear architecture, by specific epigenomic strategies with therapeutic implications for many common diseases. We have studied the role of enhancers in a variety of biological systems: pituitary corticotrope, prostate cancer, breast cancer, and neuronal cells.

INTRODUCTION

In genetics, an enhancer is a short (50-1500 bp) region of DNA that can be bound with proteins (activators) to activate transcription of a gene or transcription. Enhancers work as cis-regulatory elements to mediate both spatial and temporal control of development by turning on transcription in specific cells and/or repressing it in other cells. In eukaryotic cells the structure of the chromatin complex of DNA is folded in a way that functionally mimics the supercoiled state characteristic of prokaryotic DNA, so that although the enhancer DNA is far from the gene in regard to the number of nucleotides, it is spatially close to the promoter and gene. This allows it to interact with the general transcription factors and RNA polymerase II. Thus, the particular combination of transcription factors and other DNA-binding proteins in a developing tissue controls which genes will be expressed in that tissue. Our focus is to utilize global genomic approaches to uncover and investigate the "enhancer code" controlled by new, previously unappreciated pathways that integrate the genome-wide response to permit proper development and homeostasis, and that also functions in disease, senescence, and in the brain. We have investigated these events in differentiated cells, neuronal development, stem cells, and cancer. Our biological focus is on molecular mechanisms of the "enhancer code" regulating learning and memory; aggressive prostate and breast cancer, and they underlying events of senescence/aging. Our laboratory is "Seq-ing" (RIP-seq, ChIP-seq, RNA-seq, GRO-seq, CLIP-seq, ChIRP-seq), and a new "FISH-seq," for open-ended discovery of long-distance genome interactions to uncover new

"rules" of regulated gene transcriptional programs. Coupling this with chemical library screens, we hope to introduce new types of therapies based on targeting specific gene enhancers, and histone protein readers/ writers for cancers and other diseases.

CHAPTER 1
Enhancer-bound LDB1 regulates a corticotrope promoter-pausing repression program.

The apparent importance of promoter:enhancer looping is well established; however, the molecular mechanisms of these interactions in gene activation vs. gene repression remain to be fully elucidated. For example, our work shows that LIM domain-binding protein 1 (LDB1) can function in transcriptional enhancer-mediated gene activation mainly at the level of transcription initiation by regulating promoter:enhancer looping, consequent to the recruitment to basic helix-loop-helix–bound enhancers in pituitary corticotrope cells. Intriguingly, LDB1 also mediates promoter:enhancer looping required for target gene repression, acting at the level of promoter pausing, by recruiting metastasis-associated 1 family, member 2 to these repressive enhancers. These findings shed light on a regulatory aspect of the molecular function of LDB1, providing a putative mechanism of enhancer-dependent transcriptional repression.

The pituitary gland is the one of the major endocrine organs and is responsible for the production and secretion of several hormones, regulating multiple physiological processes including body growth, metabolism, reproduction, and stress response. Adrenocorticotropic hormone, which regulates stress level, is a cleaved product of proopiomelanocortin (POMC) produced in the corticotrope. Numerous transcription factors have been reported to regulate POMC gene expression in AtT20 cells and to be involved in pituitary gland development.

Transcription factors in the basic helix-loop-helix (bHLH) family play

essential roles in many, if not all, studied systems and in multiple developmental stages. Although bHLH genes, such as neurogenic differentiation factor D1 (NeuroD1), have been reported to regulate POMC gene expression from in vitro studies, they play only a minor and transient role in the pituitary development in vivo. Achaete-scute homolog 1 (ASCL1), a homolog of the Drosophila achaete/scute (ac/sc) bHLH transcription factor, has been shown to be critically involved in the development of several lineages in central and peripheral nervous systems and other organs. Ascl1 also is expressed in the developing pituitary gland, and experiments in Zebrafish suggest its importance during early adenohypophysis development. Studies in Drosophila indicate that achaete recruits and functionally requires a cofactor, CHIP, to regulatein Drosophila indicate that achaete recruits and functionally requires a cofactor, CHIP, to regulate target genes. Lim domain-binding (LDB)/nuclear LIM interactor (NLI)/cofactor of LIM homeodomain protein (CLIM), the mammalian homolog of CHIP, has been demonstrated to be a critical factor for multiple developmental systems and to be a component of LIM homeodomain codes that determine neuronal subtype specification in the spinal cord. Recently, using an engineered transcription unit as a model, LIM domain-binding protein 1 (LDB1) has been reported to be capable of mediating promoter:enhancer looping through LDB1 homodimerization. Here, we report that ASCL1 is a major regulator of the corticotrope lineage in the pituitary gland, activating a large set of cell type-specific enhancers, marked by levels of H3K27ac and p300 occupancy, but also causing repression of another cohort of mRNAs. Both ASCL1-directed activation

and repression transcriptional programs are largely dependent on the recruitment on enhancers of LDB1, which is required for the looping of both activating and repressing enhancers to their cognate gene promoters. Although the effects of gene activation in this cell type occur primarily at the level of transcriptional initiation, the function of the LDB1-dependent repressive enhancer appears to result in promoter pausing of target genes and is largely the result of looping-induced delivery of negative cargo, including components of the nucleosome remodeling deacetylase (NuRD) complex, from these repressive enhancers. Together, our data reveal that enhancer-bound LDB1 has a role in mediating the actions of both activating and repressive enhancers in a cell type-specific gene-regulation program.

To investigate the putative enhancer program of the terminally differentiated pituitary lineage, we conducted a survey of histone markers in a corticotrope-derived mouse cell line, AtT20 cells. ChIP followed by sequencing (ChIP-seq) of monomethylated histone H3 lysine 4 (H3K4me1), H3K27ac histone markers, and p300 protein occupancy mapped 78,608, 81,366, and 42,614 peaks, respectively. Prediction of putative enhancers (71,225) was based on the presence of H3K4me1 histone markers and by excluding annotated transcriptional start sites (TSSs) in the mouse genome. Motif enrichment analysis of all putative enhancers revealed the enrichment of the binding sites of several reported regulators of the POMC gene, including the T-BOX, bHLH, ETS, bZIP, and NR4A subfamily of nuclear receptors. Further, we have defined the subset of the enhancers that have active chromatin markers (H3K27ac and p300). Motif

enrichment analysis of H3K27ac+ p300+ enhancers revealed enrichment of binding sites for the CCCTC-binding factor (CTCF), neurofibromin 1 (NF1), bHLH, FORKHEAD, X-BOX, and T-BOX families of transcription factors. Although members of the CTCF and NF1 family are general cellular transcription factors, the factors in the bHLH and other families may be critical for lineage development and enhancer programming in corticotropes. A predicted T-BOX factor, T-BOX 19 (TBX19), has been convincingly demonstrated to be essential for the development of the POMC lineage and the expression of the POMC gene. The high enrichment of bHLH-binding sites on active enhancers prompted us to search for the bHLH factors critical for the development of murine POMC lineage. Persistent expression of Ascl1 in the differentiated POMC lineage suggested that ASCL1 might be involved in POMC lineage development. Sequential ChIP performed using chromatin isolated from adult pituitary gland, first with anti-H3K27ac antibody to immunoprecipitate active regulatory regions and followed by precipitation using the anti-ASCL1 antibody, demonstrated the presence of ASCL1 on the active POMC promoter, supporting a potential functional role of ASCL1 in regulating genes of the corticotrope lineage.

To evaluate further the predicted role of Ascl1 in the development of the POMC lineage, we first examined Ascl1 expression in the developing pituitary. Ascl1 first was detected in the evaginating oral ectoderm at embryonic day (E)9.5 and intensified at the ventral part of the nascent gland after E12.5, subsequently declining in the anterior compartment on the initiation of terminal cell-type differentiation, and eventually becoming highly expressed in the

intermediate lobe and low in the anterior lobe of the mature pituitary. Examination of pituitary-specific terminal differentiation markers and upstream regulators revealed that several lineages were affected at different regulatory steps in pituitary glands harboring an Ascl1-null mutation. At E13.5 and E17.5, Ascl1−/−pituitaries exhibited a substantial reduction in POMC expression. Expression of Tbx19 was decreased in E13.5 and E17.5 Ascl1−/− embryos. A significant decrease in NeuroD1 expression at E13.5 and inNr4a1 expression at E17.5 was observed also. We next assessed the full gene-expression program affected in the Ascl1-null mutation on the development of the pituitary gland at E13.5, using a microarray approach. In addition to some pituitary-specific targets, expression of a broad variety of other regulatory genes, including transcriptional factors/cofactors, components of cell cycle, cell death, several developmental signaling pathway, membrane receptors, and channel proteins, also were altered significantly in the absence of Ascl1. Reduced expression of several targets, including Chromogranin B, Hes6, and Manic Fringe genes, was confirmed by in situ analysis, using E14.5 embryos. In contrast, Hes family bHLH transcription factor 1 (Hes1), which serves as an effector gene of the Delta/Notch signaling pathway and as a negative feedback target of Ascl1 in other systems, was not altered in theAscl1−/− pituitary. These results thus demonstrate a broad role of ASCL1 in mRNA expression during the development of the pituitary and POMC lineage. To investigate the function of ASCL1 further, we generated AtT20 cell lines stably expressing ASCL1 fused to the biotin ligase recognition peptide (BLRP) in a cell line already expressing high levels of the biotinylating enzyme

BirA. Streptavidin pulldown of chromatin followed by sequencing revealed that, when correlated with histone markers, 13,851 out of total 21,080 ASCL1 peaks were present on enhancers but only 700 (3.29%) peaks were present on promoters in those cells. ChIP-seq of TBX19 protein, also known to be crucial for corticotroph development, revealed that 7,326 TBX19 peaks were present on H3K4me1-marked enhancers. In contrast to ASCL1, TBX19 occupancy on promoters was very high (3,481 peaks, 23.66% of total peaks). We have noticed that, although the consensus binding site for ASCL1 is consistent with previous reports, TBX19 bound to a more relaxed TNNCA core site, suggesting that specificity of TBX19 binding may rely on homodimers or heterodimers with other proteins. Indeed, TBX19 was bound to POMC enhancer as a homodimers and to POMC promoter as a heterodimer with PTX1. Intersection of ASCL1- and TBX19-containing enhancers revealed 3,096 cobound peaks. More than 90% and more than 80% of the cobound enhancers were colocalized with H3K27ac and p300 peaks, respectively, in contrast to the lower association of those markers with other enhancers, supporting the active role of those factors in determining the enhancer program in those cells.

To investigate the effect of Ascl1 on gene transcription of their targets, global run-on sequencing (GRO-seq) experiments were performed, revealing down-regulation of 956 genes and up-regulation of 1,252 transcription units upon siAscl1 treatment. Among those dysregulated genes, □40% of ASCL1-activated genes had ASCL1 enhancers within a median distance of 21 kb from the regulated promoter, whereas 58% of the ASCL1-repressed genes had ASCL1-bound

enhancers at a median distance of 52 kb. These results suggest that repressive enhancers may locate farther from their cognate gene promoters than activating enhancers.

Additionally, knockdown of Ascl1 and Ldb1 using specific siRNAs repressed the expression of the POMC gene. Together, these data indicate that recruitment of LDB1 is required for POMC gene activation. To determine the global binding pattern of LDB1 in AtT20 cells, ChIP-seq was carried out in cell lines stably expressing the BLRP–LDB1 fusion protein. The analysis revealed 10,627 LDB1 peaks on enhancers with bHLH consensus binding sites highly enriched on those peaks. 5,846 (55%) LDB1 peaks on enhancers were co-occupied with ASCL1, consistent with the apparent physical interactions between the two factors. The percentage of LDB1 peaks on the promoter (5.61%) was comparable to the rate of the promoter binding for ASCL1. The cobinding of LDB1 and ASCL1, together with TBX19, p300, H3K27ac, and K3H4me1, was observed on both the POMC promoter and enhancer. Analysis of GRO-seq results for genes in the proximity (< 300 kb) of cobound LDB1 and ASCL1 enhancers revealed that the Ldb1 transcriptional program was highly correlated with that of Ascl1 (R = 0.80); in contrast, functional correlation between Ldb1 and Tbx19 was low (R = 0.27). LDB1 was postulated to play an important role in promoter:enhancer looping events and as a coactivator in several systems. To investigate its roles in gene transcription and in promoter:enhancer looping events in the POMC lineage, we performed GRO-seq after Ldb1 knockdown and observed that, of 3,378 genes down-regulated by

siLdb1, 927 transcription units are located in direct proximity to the LDB1-occupied enhancers. The tag density plot of GRO-seq data from Ldb1-knockdown cells revealed that nascent transcripts were highly concentrated at promoters of LDB1-activated genes in control cells and were massively reduced upon Ldb1 knockdown, suggesting that the primary transcriptional role of LDB1-bound activating enhancers was to promote transcriptional initiation. The GRO-seq signal also was greatly reduced across the gene body regions, presumably mainly because of a drastic lower transcriptional initiation rate. To investigate further how LDB1-bound enhancers activate their target genes, we examined the looping interactions of representative enhancers in AtT20 cells by a promoter:enhancer chromosome conformation capture-based DNA annealing selection and ligation assay, using oligonucleotide libraries synthesized to interrogate promoter and enhancer regions. We found that enhancer:promoter interactions between LDB1-bound enhancers and the adjacent promoters of POMC and lung carcinoma myc related oncogene 1 (Lmyc1) genes were greatly diminished following knockdown of Ldb1. The significant change in the efficiency of enhancer–promoter ligation upon Ldb1 knockdown also was observed when different restriction enzymes were used. Interestingly, LDB1-dependent interactions were not dependent on the presence of LDB1 on the promoters, as in the case of the Lmyc1 locus. To confirm these results further, we also performed a conventional chromosome conformation capture (3C) assay for the POMC locus and observed that the relative ligation efficiency was reduced by > 60% after knockdown of Ldb1, whereas internal controls showed comparable

genomic DNA present in both samples.

Tag density analysis of the GRO-seq data revealed that LDB1-repressed transcription units had higher nascent transcription signal in the gene body of Ldb1-knockdown cells than that of control cells; the reverse was observed for the promoter region, suggesting that knockdown of Ldb1 caused the release of transcriptional pausing of LDB1-repressed genes. Further analysis showed that 1,446 LDB1-repressed genes were located within 300 kb of LDB1-occupied enhancers. Interestingly, the PE3C-DSL assay showed that LDB1-dependent enhancer:promoter interactions also are required for repressed genes. A confirmation by the 3C assay showed that interactions between enhancers and promoters for EF-hand calcium-binding domain-containing protein 1 (Efcab1) and Synaptotagmin I (Syt1) genes were reduced by > 70% and > 90%, respectively. Taken together, these data indicated that LDB1 is required for enhancer:promoter interaction in both the LDB1-activated and -repressed transcriptional programs.

Mass spectrometry analysis of LDB1-interacting proteins detected several members of the NuRD complex, suggesting that the NuRD complex may have a possible role in LDB1 function. Coimmunoprecipitation assays were performed to confirm the interaction between LDB1 and chromodomain helicase DNA-binding protein 4 (CHD4) and between LDB1 and metastasis-associated 1 family, member 2 (MTA2). A significant interaction between LDB1 and MTA2 was observed, although only a minimal signal of CHD4 above beads control was observed. To investigate any possible functional relationship of Ldb1and Mtas,

GRO-seq was performed in AtT20 cells after specific knockdown of both Mta1 and Mta2. Interestingly, we observed that 586 genes repressed by LDB1 also were repressed by MTA1/2. To investigate MTA binding on these genes, we performed ChIP-seq using MTA2 antibody. A tag density plot of 369 cobound MTA2 and LDB1 enhancers in the proximity of genes repressed by both LDB1 and MTAs showed significant MTA2 binding. Interestingly, MTA2 occupancy on those enhancers was reduced significantly upon knockdown of LDB1, as exemplified by the protein LTV1 homolog (Ltv1), or DNA repair and recombination protein RAD54-like (Rad54l) loci. As a control, the binding of MTA2 on all LDB1 enhancers was not changed after Ldb1 knockdown. These two genes were repressed by LDB1 and MTA1/2, as demonstrated by the up-regulation of the prespliced nascent RNAs of both genes upon knockdown of Ldb1 and Mta1/2. Our data thus indicate that the genes corepressed by LDB1 and MTA1/2 are in the proximity of the LDB1-bound enhancers and that MTA2 is recruited to those enhancers in an LDB1-dependent fashion.

The role of LDB1 as a coactivator in development has been studied for more than a decade; however, the molecular mechanisms of its function remain poorly understood. Here, we report that LDB1 can function both for transcriptional enhancer-mediated gene activation in pituitary corticotrope cells, mainly at the level of transcription initiation, and for the actions of enhancers mediating target gene repression at the level of promoter pausing. LDB1 appears to mediate enhancer looping both for enhancers bringing activating cargo and for those bringing repressive cargo to the promoter. These findings shed light on the

molecular function of LDB1 and point to the putative mechanism of enhancer-dependent transcriptional repression based on the recruitment of the NuRD complex or components of the NuRD complex at repressive enhancers.

CHIP has been reported to interact with Drosophila proneuronal ASC-Da bHLH factor. Our proteomic studies demonstrate that LDB1 interacts strongly with pan bHLH factors, the E-proteins. This observation potentially expands the function of LDB1 to the large bHLH family. Because LDB1 also has been reported to interact with other transcription factors, it is tempting to suggest a potentially general role for LDB1 in regulating cell type-specific, enhancer-dependent gene-expression programs. LDB1-mediated promoter:enhancer looping has been reported in the β-globin locus based on the homodimerization of LDB1 occupying both enhancer and promoter sites. Our data show that, in corticotropes, LDB1 can mediate looping to promoters without detectable promoter LDB1 binding. Because we have detected by ChIP-seq that LDB1 is bound predominantly to enhancers, our data clearly suggest that the majority of LDB1-mediated looping events do not reflect enhancer:promoter-dependent LDB1 homodimerization. In accord with the report in the β-globin locus suggests that LDB1 regulates transcriptional elongation, our data also show reduced nascent RNA signal in gene body regions by half upon LDB1 knockdown. However, our data argue that the primary role of LDB1 is to regulate transcriptional initiation.

In addition to gene activation, LDB1 appears to be equally required for the function of enhancers putatively mediating target gene repression. We have

shown that LDB1 serves as a crucial component of looping mechanisms in enhancer-mediated gene repression. Furthermore, we show that LDB1 is involved in the recruitment of a NuRD complex component, MTA2, to the repressive enhancer, suggesting that this component is a part of the cargo that mediates the promoter-pausing events regulated by these LDB1-bound enhancers. These findings significantly expand our understanding of the role of LDB1 protein in gene regulation and also show, for the first time to our knowledge, that LDB1-dependent looping serves as the strategy for delivering repressive cargo to promoters exhibiting promoter pausing.

Materials and Methods.

Cell Culture and Transfection.

AtT-20 cells were obtained from ATCC, maintained in UltraCULTURE medium (Lonza) supple- mented with 5% (vol/vol) FBS and 1× GlutaMAX (Invitrogen) without antibiotics at 37 °C in a humidified atmosphere containing 5% (vol/vol) CO. Transfection of expression vectors was performed using Lipofectamine 2000 (Life Technologies), according to the manufacturer's protocol. To knock down endogenous pro- teins efficiently, the Neon Transfection System (Invitrogen) was used to electroporate gene-specific siRNAs to AtT20 cells at 1,400 V for 30 ms, following the manufacturer's instructions. Ten microliters of 50 μM total siRNA was used for 1×10^6 cells in a 100-μL tip. Specific mRNA and/or protein were monitored by qPCR or Western blot analysis for knockdown efficiencies.

BLRP-Tagged Cell Lines.

To make cell lines expressing tagged ASCL1 or LDB1, we modified an in vivo biotin-tagging system. AtT20 cell lines expressing a high level of the bacterial biotin ligase BirA were selected and tested for biotin ligation efficiency. Two lines were chosen as parental lines to make BLRP-tagged expression cell lines. Expression vectors for tagged-ASCL1 and LDB1 were engineered from a modified pCAGGS vector, which was inserted in a double-stranded oligonucleotide encoding the peptide MAGGLNDIFEAQKIEWHEDTGGSWRAPGGGGGSGGGGSGENLYFQSSDYKDHDGDYKDHDIDYKDDDDK. The BLRP (amino acids 3–20), in which the lysine residue (underlined) can be biotinylated by BirA, was followed by a glycine-rich region, a specific recognition site for TEV protease (ENLYFQS; also underlined), and a 3×FLAG sequence. The glycine-rich region serves as a spacer between the BLRP and the TEV site. To increase the specificity further, the TEV site is included for cleaving the tagged protein from the BLRP tag (Life Technologies). The cDNA of Ascl1 and Ldb1 was in-frame ligated to the 3′ end of the tag DNA sequence. The expression vectors were transfected into BirA-expressing cell lines. Four individual cell lines for each construct were chosen and pooled with a comparable number cells for ChIP-seq experiments.

Antibodies.

Specific antibodies used in this study are as follow: anti- P300 (sc585),

anti-MTA1 (sc-10813), and anti-MTA2 (sc-9447) from Santa Cruz Biotechnology; anti-H3K4me1 (07-436) from Millipore; anti-H3K27ac (ab4729) from Abcam; anti-TBX19 (HPA005800) from Sigma-Aldrich; anti-ASCL1 (556604) from BD Pharmingen; anti-CHD4 (A301-081A) from Bethyl Laboratories; and anti-LDB1 sera from Gordon N. Gill, University of California, San Diego.

Reporter Assays.

The POMC reporter was constructed by ligating the PCR product of the POMC 5′ region (−4757 to +74) into multiple cloning sites of the pGL2-Basic luciferase vector (Promega). Expression vectors were constructed by inserting cDNAs in pcDNA3. The reporter assay was performed 12 hr after transfection. The dual-luciferase reporter assay kit (Promega) was used for the luciferase assay; plates were read in a Veritas Microplate Lumin- ometer (Turner Biosystems).

qPCR and Data Analysis.

Total RNA was extracted with the RNeasy Mini Kit (Qiagen); then total RNA was reverse transcribed with an iScript cDNA synthesis kit (BioRad). qPCRs were performed in MX3000P (Stratagene) using 2× qPCR master mix from Affymetrix or Bio-Rad. Gene-specific primer pairs were used for quantitative analyses of relative transcript levels; Gapdh transcripts were used as an internal control. The comparative cycle threshold method was used to calculate relative mRNA levels. Statistics were performed using unpaired two- tailed

Student's t tests.

ChIP, in Vivo Sequential ChIP, and ChIP-Seq.

Briefly, cells were cross- linked with 1% formaldehyde in PBS at room temperature for 10 min. Double cross-linking was used for MTA2 ChIP, in which cells first were cross-linked with 2 mM disuccinimidyl glutarate at room temperature for 45 min, followed by 10 min with 1% formaldehyde. Fixation was quenched by glycine at a final concentration of 0.125 M for 5 min. Chromatin was fragmented to 300–500 bp using a tip sonicator and was centrifuged at 16,000 × g for 10 min. The supernatant was precleared with 20 μL of mag- netic beads. The precleared chromatin was incubated with 2–5 μg of antibodies at 4 °C overnight. Antibody–protein–DNA complexes were pulled down by incubation with 20 μL of Protein G Dynabeads and were washed. The eluted complex was decross-linked overnight at 65 °C. BLRP-tagged ChIP was performed using Nanolink Streptavidin Magnetic beads (Solulink) to pull down the BLRP-tagged protein–DNA complex directly. Additional brief washes with 1% SDS in Tris-EDTA (TE) and TE buffer were performed after normal washes. The washed streptavidin beads then were subjected to TEV protease digestion to release tagged protein and DNA complex before decross-linking at 65 °C overnight. For in vivo ChIP, adult murine pituitaries were dissected and immediately immersed in 1% formaldehyde fixation buffer at room temperature for 20 min. The cross-linked tissues were homogenized and sonicated. Sequential ChIP was conducted first with anti-H3K27ac antibody. After washing, the complex was eluted by

incubating beads with 25 μL of 10 mM DTT for 30 min at 37 °C. The supernatant was diluted 20 times with dilution buffer [20 mM Tris·HCl (pH7.4), 100 mM NaCl, 0.5% Triton X-100, 2 mM EDTA] and then was subjected to the second ChIP with anti-ASCL1 antibody. The decross-linked ChIP DNA was purified with the QIAquick PCR Purification Kit (Qiagen). The ChIP-seq library was constructed following the Illumina ChIP-seq Sample Prep kit. The library was amplified by 14 cycles of PCR.

PE3C-DSL Assay.

Briefly 5×10^6 AtT20 cells, treated with control or Ldb1 siRNAs for 2 d, were fixed in 1% formaldehyde at room temperature for 10 min. Fixation was terminated by adding 0.5 mL of 2.5-M glycine per 10-mL volume for 5 min at room temperature, fol- lowed by 15 min at 4 °C. Cells were incubated on ice in 500 μL of 10 mM Tris·HCl (pH 8.0), 10 mM NaCl, 0.2% IGEPAL CA-630 (Sigma-Aldrich), 1× protease inhibitor mixture for 15 min and were lysed with a Dounce homogenizer. The suspension was spun down for 5 min at 2,000 × g. The pellet was washed twice with 500 μL of 1× NEBuffer 2 (New England Biolabs) and re- suspended in 362 μL of 1× NEBuffer 2. Next, 38 μL of 1% SDS was added, and the mixture was incubated at 65 °C for 10 min. SDS then was quenched by 44 μL of 10% (vol/vol) Triton X-100. Chromatin subsequently was digested overnight with 400 units of the restriction enzymes indicated in the figures in a final volume of 500 μL. After enzymes were inactivated with 86 μL of 10% (wt/vol) SDS at 65 °C for 30 min, each digested chromatin mixture was

ligated by T4 DNA Ligase (800 units) in 7.5 mL ligation buffer for 4 hr at 16 °C. The chromatin subsequently was decross-linked overnight at 65 °C and was purified twice with phenol and once with phenol/chloroform. DNA was precipitated, and pellets were air-dried before being resuspended in 250 μL 1× TE buffer. RNA was removed by incubating DNA with 1 μL RNase A (1 mg/mL) at 37 °C for 15 min. DNA was extracted again with phenol/chloroform and precipitated. Equal amounts of 3C chromatin were biotinylated using the Photoprobe Kit (Vector Labs). After 10 min at 95° C to denature the 3C DNA, donor and acceptor probe pools (100 fmol per probe) were an- nealed to the biotinylated 3C samples at 45 °C for 2 hr. The biotinylated DNA was pulled down by streptavidin magnetic beads, and unbound oligonucleotides were removed by washing. The 5′- phosphate of acceptor probes and the 3′-OH of donor probes were ligated using Taq DNA ligase at 45 °C for 1 hr on the mag- netic beads. The ligated products (donor–acceptor pairs) were purified by alkaline elution from the biotinylated 3C DNA and further by size selection on denaturing PAGE. The purified ligation products were PCR amplified with limited cycle numbers using universal adaptor sequences and subjected to HiSeq 2000 sequencing.

In Situ Hybridization and Immunostaining.

In situ hybridization with ^{35}S-labeled anti-sense RNA probes was performed as previously described. For immunostaining, E14.5 embryos were fixed in 4% (wt/vol) paraformaldehyde, frozen in 1:1 OTC/Aqua- Mount,

cryosectioned at 12 μm, and mounted on Superfrost Plus slides (Fischer). Sections were incubated with primary antibody at a 1:500 dilution in PBS, 0.3% Triton X-100, overnight at 4° C. After three washings in PBS, slides were incubated with TRITC- coupled secondary antibodies for 2 hr at room temperature. Sections then were mounted using the Slow Fade antifade kit.

Coimmunoprecipitation.

A 150 mm plate of cells was washed twice with cold PBS and harvested at a confluence of 80–90% on ice. Cells were spun down and lysed in buffer A [10 mM Hepes-KOH (pH 8.0), 10 mM KCl, 1.5 mM MgCl2, 1 mM DTT, and 1× protease inhibitor mixture] for 15 min. Then 100 μL of 10% (vol/vol) Nonidet P-40 was added to cells, and cells were vortexed immediately for 10 s at maximum speed, followed by spinning at 10,000 × g for 3 min. The pellets were resuspended with brief sonication in buffer B [1% Triton X-100, 50 mM Tris·HCl (pH7.4), 150 mM NaCl, 0.2 mM EDTA, 1× proteinase inhibitors, and 0.1 mM DTT with 1 μL of Benzonase] for 1 h, followed by centrifugation at 16,000 × g for 10 min. Then 2 μg of antibodies were added to 0.5 mL supernatant and were incubated overnight at 4 °C. Then 30 μL of Dynabeads protein G was added for 2 hr at 4° C with rotation to pull down the antibody–protein complexes. After the supernatant was removed, the beads–protein complexes were washed three times with 150 mM NaCl, 25 μM Tris (pH 8.0), 0.1% Nonidet P-40, and 1 mM EDTA and were boiled in 10 mM DTT and 1×LDS sample buffer for 3 min. Western blotting was used to detect specific proteins in the complex.

Protein Complex Purification.

Cells (10^9) were harvested after washing in cold PBS, and 20 mL of nuclei extract was prepared as described above for immunoprecipitation. Then 100 μL of ANTI- FLAG M2 Magnetic Beads (Sigma) were added to the nuclear lysates and incubated for 4 hr at 4 °C. The beads–protein com- plexes were washed with 10 mL nuclear extraction buffer, followed by five washings in 1 mL of buffer B and two washings in 1 mL of Tris-buffered saline. Complexes were eluted with 0.5 mL 0.2 g/L 3X FLAG peptide (Sigma) for 30 min. Protein complexes were concentrated by Amicon Ultra 4 mL centrifugal filter devices (Millipore) and subjected to mass spectrometry analysis.

Mass Spectrometry Sample Preparation and Analysis.

Mass spectrometry sample preparation and analysis were performed as previous reported. Briefly, protein samples were boiled in 0.1% RapiGest SF reagent for 5 min and were incubated in 1 mM Tris (2-carboxyethyl) phosphine (TCEP) at 37 °C for 30 min. Subsequently, the samples were carboxy- methylated with 0.5 mg/mL of iodoacetamide for 30 min at 37 °C followed by neutralization with 2 mM TCEP. Protein samples then were digested with trypsin overnight at 37 °C. RapiGest was degraded and removed by treating the samples with 250 mM HCl at 37 °C for 1 h followed by centrifugation at 16,000 × g for 30 min at 4 °C. The peptides were extracted from the soluble fraction and desalted using C18 desalting columns. For LC-MS/MS analysis, Trypsin-digested peptides were

analyzed by ultra-high-pressure liquid chromatography (UPLC) coupled with LC-MS/MS using nano-spray ionization, which was performed by a Triple TOF 5600 hybrid mass spectrometer (AB SCIEX) interfaced with nano-scale reversed-phase UPLC (nanoACQUITY; Waters Corp.) using a 20 cm-75-μm i.d. glass capillary packed with 2.5 μm C18 (130) CSHTM beads (Waters Corp.). Peptides were eluted from the C18 column into the mass spectrometer using a linear gradient (5–80% vol/vol) of acetonitrile at a flow rate of 250 μL/min for 1 hr. MS/MS data were acquired in a data-dependent manner in which the MS1 data were acquired for 250 ms at an m/z of 400–1,250 Da and the MS/MS data were acquired from an m/z of 50–2,000 Da. The independent data-acquisition parameters were as follow; acquisition time of 250 ms, followed by 50 MS2 events with a 48-ms acquisition time for each event. The threshold to trigger an MS2 event was set to 150 counts when the ion had the charge state +2, +3, and +4. The ion exclusion time was set to 4 s. Finally, the collected data were analyzed using Protein Pilot 4.5 (AB SCIEX) for peptide identification.

GRO-Seq.

GRO-seq experiments were performed as reported previously. Briefly, 20 million AtT20 cells, treated with siRNAs for 2 d, were washed three times with cold PBS and then were incubated in swelling buffer [10 mM Tris·HCl (pH 7.5), 2 mM MgCl2, 3 mM CaCl2] on ice for 5 min. Cells were lysed in 1 mL of lysis buffer [swelling buffer plus 0.5% Nonidet P-40, 10% (vol/vol) glycerol, and 20 units of SUPERase-In] with a 15-s vortex at 800 rpm. The nuclei pellets were

washed twice with 10 mL lysis buffer and 1 mL of freezing buffer [50 mM Tris·HCl (pH8.3), 40% (vol/vol) glycerol, 5 mM MgCl2, 0.1 mM EDTA]. The isolated nuclei were suspended in 100 μL of freezing buffer. For the run-on assay, an equal volume of reaction buffer [10 mM Tris·HCl (pH 8.0), 5 mM MgCl2, 300 mM KCl, 1% sarkosyl, 500 μM ATP, GTP, and Br-UTP, 2 μM CTP, 1 mM DTT, 20 units of SUPERase-In] was added to the isolated nuclei and incubated for 5 min at 30° C. The nuclear-run-on RNA (NRO-RNA) then was extracted with TRIzol LS reagent (Life Technologies). NRO-RNA was hydrolyzed to □300–500 nucleotides by alkaline base on ice for 30 min followed by treatment with DNase I and Antarctic Phosphatase.

The fragmented NRO-RNA was pulled down with anti-BrdU agarose beads (Santa Cruz Biotechnology) in binding buffer (0.5× SSPE, 1 mM EDTA, 0.05% Tween) for 1–3 hr at 4 °C with rotation. Subsequently, T4 PNK was used to repair the ends of the precipitated bromouridine-triphosphate– labeled RNA at 37 °C for 1 hr. The RNA was extracted and pre- cipitated using acidic phenol-chloroform. The RNA fragments were subjected to poly-A tailing reaction by poly-A polymerase (New England Biolabs) for 30 min at 37 °C and were reverse transcribed using oNTI223 primer and the Superscript III RT kit (Life Technologies). The 100- to 500-bp cDNA products were isolated using 10% (vol/vol) polyacrylamide Tris-borate-EDTA (TBE)-urea gels. Next, the purified first-strand cDNA was circu- larized by CircLigase (Epicentre) and relinearized by APE1 (New England Biolabs). Approximately 120–320 bp of relinearized single-strand cDNA (sscDNA) was isolated with a 10% (vol/vol) polyacrylamide

TBE gel. Finally, the purified sscDNA template was amplified by PCR using the Phusion High-Fidelity enzyme (New England Biolabs). The oligonucleotide primers oNTI200 and oNTI201 were used to generate DNA for deep sequencing.

Deep Sequencing.

The DNA libraries were sequenced on Illumina GAII or HiSeq 2000 platforms according to the manufacturer's instructions. The sequences returned by the Illumina Pipeline were aligned to the mouse assembly mm8 or mm9 by using Bowtie or Bowtie2. The data were visualized by preparing custom tracks on the University of California, Santa Cruz genome browser using the HOMER software package (homer.salk.edu/homer/). For visualization, the total number of aligned reads was normal- ized to 10^7 for each experiment presented in this study.

ChIP-Seq Analysis, Heatmaps, and Tag Density Analysis.

ChIP-seq peaks were identified using HOMER with different parameters for transcription factors and histone marks. The called peaks were associated with genes by cross-referencing the RefSeq TSS database; for H3K4me1 and H3K27ac, the enhancer peaks were defined by the absence of nonoverlap with the TSSs in the Ref-Seq collection. For intersection analysis, the peaks were extended ±1 kb on each side; we used the intersectBED tool in the BEDTools package. Heatmap matrices were generated in HOMER by counting the tags in a 6 kb window (±3 kb of the peak centers) and were displayed in MeV. The tag density plots generated by HOMER by counting the tags in a 6 kb window (±3 kb of the

peak center) were plotted in Excel. The MTA2 ChIP- seq tag density plots were normalized using a 1 kb window up-stream and down-stream of the 1 kb peak regions of the LDB1/ MTA2 cobound enhancers. Motif enrichment was analyzed using a comparative algorithm as described in HOMER. The sequence logos were displayed by using WebLOGO (weblogo. berkeley.edu), the box plots were generated in R/BioC, and

1. Liu Z, et al. (2014) Enhancer activation requires trans-recruitment of a mega tran- scription factor complex. Cell 159(2):358–373.

2. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38(4):576–589.

3. Lieberman-Aiden E, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293.

4. Kwon YS, et al. (2007) Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. Proc Natl Acad Sci USA 104(12):4852–4857.

parametric or nonparametric tests were used for significance analysis.

GRO-Seq Analysis.

The sequencing reads were aligned to the mouse genome mm8 or mm9 by using Bowtie or Bowtie2, and the reads were counted over the gene bodies (excluding the promoter- proximal region) by using HOMER or BEDTools. For

siAscl1, siLdb1, and siTbx19 GRO-seq analysis, we used the same number (7 million) of randomly extracted reads to define the differentially expressed genes, excluding a 400 bp promoter-proximal region. edgeR was used to analyze the statistical significance of the differential expression (FDR < 0.01, FC > 1.2). We used the closestBED tool in the BEDTools package to assign the closest differentially expressed genes to ASCL1- or LDB1-bound enhancers. For siMta1/2 and siLdb1 GRO-seq analysis, the alignment was performed after extracting the same number of reads for each of the experiments. The common artifacts derived from clonal amplification were circumvented by considering a maxi- mum of three tags per unique sequence. To determine siLdb1- and siMta2-dependent changes in gene bodies, the sequencing reads were counted over the first 60 kb of the entire gene body, excluding the 2 kb promoter-proximal region on the sense strand with respect to the gene orientation, and a FC >1.2 was applied. The cobound LDB1/MTA2 enhancers were found within a 500 kb distance of double-activated or double-repressed genes by siLdb1 and siMta1/2.

Analysis of the PE3C-DSL Data.

After the adaptor sequences were removed, the distinct reads were counted by using scripts written in Perl and were aligned to a custom library that includes all the combinations of donors and acceptors. The alignment was performed with Bowtie or Novoalign, and the counts of all the interactions in distinct samples were normalized.

Chapter 1, in full, is a reprint of the material as it appears in Proc Natl

Acad Sci USA 2015. Enhancer-bound LDB1 regulates a corticotrope promoter-pausing repression program. Zhang F, Tanasa B, Merkurjev D, Lin C, Song X, Li W, Tan Y, Liu Z, Zhang J, Ohgi KA, Krones A, Skowronska-Krawczyk D, Rosenfeld MG., 2015.

CHAPTER 2
Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens.

The emergence of powerful new chemical library-screening approaches and the generation of new types of chemical structures makes novel methods available to link candidate chemicals to potential target genes, e.g., as in the interaction with and effects on chromatin-bound targets. Here we report a method that can provide the genome-wide location of a candidate drug. One such synthetic chemical, SD70—first identified in a screen for inhibitors of tumor translocation events—was resynthesized with a tag permitting a ChIP-sequencing–like analysis, referred to as "Chemical affinity capture and massively parallel DNA sequencing (Chem-seq)." As a consequence of finding its recruitment on androgen receptor-bound functional enhancers, we were able to demonstrate that SD70 could inhibit the prostate cancer cell transcriptional program, in part by inhibition of the demethylase KDM4C.

Understanding the mechanisms by which compounds discovered using cell-based phenotypic screening strategies might exert their effects would be highly augmented by new approaches exploring their potential interactions with the genome. For example, altered androgen receptor (AR) transcriptional programs, including castration resistance and subsequent chromosomal translocations, play key roles in prostate cancer pathological progression, making the quest for identification of new therapeutic agents and an understanding of their actions a continued priority. Here we report an approach that has permitted us to uncover the sites and mechanisms of action of a drug, referred to as "SD70,"

initially identified by phenotypic screening for inhibitors of ligand and genotoxic stress-induced translocations in prostate cancer cells. Based on synthesis of a derivatized form of SD70 that permits its application for a ChIP-sequencing–like approach, referred to as "Chem-seq," we were next able to efficiently map the genome-wide binding locations of this small molecule, revealing that it largely colocalized with AR on regulatory enhancers. Based on these observations, we performed the appropriate global analyses to ascertain that SD70 inhibits the androgen-dependent AR program, and prostate cancer cell growth, acting, at least in part, by functionally inhibiting the Jumonji domain-containing demethylase, KDM4C. Global location of candidate drugs represents a powerful strategy for new drug development by mapping genome-wide location of small molecules, a powerful adjunct to contemporary drug development strategies. Although numerous small molecules functioning as drugs have been discovered through various strategies, identifying drug targets and their mechanisms of action is a challenging process. In fact, a full understanding of mechanisms of action is available for only some of chemical entities that are approved by US Food and Drug Administration. Therefore, any approaches that would augment our identification of drug mechanism would be desirable for both basic and clinical applications. The application of high-throughput sequencing technology to global genomic approaches licenses our ability to approach drug mechanisms using strategies similar to those used to investigate transcriptional regulation. ChIP sequencing (ChIP-seq) had permitted to identify the location of many DNA binding transcription factors, cofactors, and other proteins in the genome, and

infers that if similar approaches could be applied to chemicals identified in library screens, particularly those involving readout of transcription or other genomic regulatory events, we could obtain clues about potential genomic actions of many of these newly identified chemicals that might bind directly or indirectly to chromatin. Thus, mapping the genome-wide binding profile for such small molecules would represent a significant advantage for drug development and formulating mechanistic insights. In fact, just such an approach, named "Chem-seq," was recently reported for location of bromodomain inhibitors. Here we report our parallel development of similar, small molecule-bound ChIP-seq–like strategy to explore the mechanism of action of an apparent anticancer drug, SD70, which we initially identified by screening for inhibitors of dihydrotestosterone (DHT) and genotoxic stress-induced gene translocation events, revealing colocalization at androgen receptor (AR)-occupied regulatory enhancers, and which provoked appropriate experiments to explore its mechanism of action and its potential efficacy in vivo.

Our first experiments centered on identifying new chemicals that might inhibit an AR/genotoxic stress-dependent chromosomal translocation events in LNCaP human prostate adenocarcinoma cell line. Chromosomal translocation and gene fusion events, which contain translocation of the 5′ untranslated region of the AR target gene TMPRSS2 to two members of the ETS family of genes, ERG and ETV1, placing specific members of the ETS gene family under the control of androgens; such events have been proposed to provide a driving force to the development or aggressiveness of prostate cancers. Androgens and genotoxic

stress rapidly and synergistically induce TMPRRSS2:ETS translocation events in LNCaP prostate cancer cells, in part due to the AR-mediating androgen-dependent spatial proximity of corresponding chromosomal translocating regions. Using an assay that measures TMPRSS2:ETS translocation frequency in LNCaP cells under the androgen stimulation and genotoxic stress, we screened a collection of small molecules that consists of epigenetic and nuclear receptor modulators. Remarkably, one heterocyclic small molecule, SD70, showed the strongest inhibition both for TMPRSS2:ERG andTMPRSS2:ETV1 fusion events in repeat experiments, highly suggesting its potential relevance to prostate cancer therapy. SD70 contains an 8-hydroxyquinoline moiety that is an Fe(II) chelator. A small molecule with a similar 8-hydroxyquinoline structure had been previously found to function as a competitive inhibitor of jumonji domain-containing histone demethylase JMJD2 (also known as KDM4) family for α-ketoglutarate, a cofactor for demethylase activity. Molecular-docking simulation suggested that KDM4C might be one of its target proteins. KDM4C was identified as a putative oncogene particularly for prostate cancer, and functioning in AR transcriptional program by histone H3K9me3/me2 demethylase activity; it also exhibits nonhistone substrate demethylase activity by demethylating Pc2 lysine191 dimethylation, and function in E2F1-mediated cell growth control. In the translocation assay, both TMPRSS2:ERG and TMPRSS2:ETV1 fusion rates were dramatically decreased upon KDM4C siRNA knockdown, phenocopying effects of SD70, and therefore further suggesting that KDM4C might represent as at least one target mediating SD70 effects.

On the basis of the effect of SD70 in the screening assay, it became of particular interest to investigate whether this chemical might exert effects by binding to specific genomic regulatory regions. This was approached by a linker with a biotin moiety to permit its use in a modified ChIP-seq assay referred to as "Chem-seq". LNCaP cell cultures were pretreated with 10 μM biotinylated drug or vehicle for 2 hr, followed by 100 nM DHT stimulation or vehicle control for 1 hr. The treated cells were cross-linked by formaldehyde and compound capture was performed using streptavidin beads. After elution and de–cross-linking, the precipitated DNA was subjected to high throughput sequencing. Using the Hypergeometric Optimization of Motif Enrichment (HOMER) peak-finding program (http://homer.salk.edu) with relative stringent parameters, we identified 2,123 and 2,128 robust binding sites in the absence and presence of DHT conditions, respectively, but interestingly, with only 129 peaks overlapping between the two conditions, indicating a dynamic pattern of altered SD70 binding in response to ligand. Following DHT treatment, ~91% of the gained peaks overlapped with AR-bound enhancers, as marked by H3K4me1 and H3K27ac. The vast majority of the binding sites lost following DHT treatment were not enhancers and did not harbor AR binding sites. Tag density profiling showed SD70 peaks upon DHT treatment and AR enhancers centered together. These data suggested that a significant gain of binding of SD70 occurred on AR-bound regulatory enhancers in a DHT-dependent manner, with androgen response elements as the top-predicted DNA-binding motif under the DHT-stimulated condition. In genome browser images, AR enhancer and SD70 colocalized, as

exemplified by the canonical AR target FKBP5 gene enhancer. The genome-wide binding profile, generated using SD70 with the linker-mediated biotin label, was therefore able to be successfully identified by Chem-seq, indicating the potential power of this strategy for drug target/mechanism evaluation.

To begin to test whether SD70 might in addition function as an inhibitor of DHT-regulated gene transcription, we tested expression of some AR target genes by quantitative real-time PCR (RT-qPCR), finding strong inhibition of DHT-induced expression in LNCaP cells, and of basal levels in castration-resistant prostate cancer cells CWR22Rv1 in which AR is constitutively activated, suggesting it also functions in gene transcriptional regulation. Based on the preferential localization of SD70 to AR-bound enhancers, it was of interest to investigate whether SD70 would have any effects on AR-regulated programs in the LNCaP prostate cancer cell model. To best determine any SD70 effects on transcription, we first performed Global Run-on followed by high-throughput sequencing (GRO-seq) to interrogate SD70's broader transcriptional role. LNCaP cells were stripped in hormone-deprived medium for 3 d before treatment with a drug or vehicle, and then stimulated by DHT, followed by GRO-seq. Overall, we found that of 2,445 DHT up-regulated genes, exhibiting a fold change (FC) > 1.5, the DHT induction was significantly reduced (median FC from 1.868 to 1.209, $P < 10-5$) following treatment with SD70, with essentially no effects on a randomly selected cohort of 1,000 non–DHT-regulated coding transcription units (median FC from 0.960 to 0.990). Forty-two percent of all SD70-repressed genes in the plus-DHT condition were DHT-induced genes. To investigate the

relationship of SD70 binding sites and gene transcription control effect, we assessed AR ChIP-seq, SD70 Chem-seq, and GRO-seq together, finding 1,132 of 2,445 genes had both AR enhancer and SD70-binding sites within 500 kb of a regulated coding gene promoter. This cohort of DHT-activated target genes exhibited strong inhibition of DHT induction in response to SD70 (median FC from 1.914 to 1.138), whereas for 1,078 of 2,445 genes having an AR enhancer but not exhibiting SD70 binding within 500 kb of a coding target gene, the DHT induction effect was not strongly decreased (median FC from 1.774 to 1.510). Although an enhancer could regulate more than a single target gene, long-distance interactions may account for the detectable, but minimal, effect of the down-regulation for coding genes located beyond 500 kb of the SD70 enhancer binding site. By hierarchical clustering of DHT up-regulated genes, we found □70% of the DHT program was altered by SD70. The molecular function Gene Ontology (GO) term analysis (http://david.abcc.ncifcrf.gov/) for SD70 repressed or activated genes in the DHT-stimulated condition indicated that “steroid hormone receptor activity” was among the top ten terms for repressed genes, consistent with a functional role of SD70 in repressing the AR transcriptional program. To rule out the possibility that the DHT-dependent transcriptional inhibition effect we observed might reflect loss/degradation of AR, we measured the AR protein level after SD70 treatment, finding by Western blot analysis no significant alteration in levels of the AR protein across a wide range of concentrations of SD70 treatment conditions. ChIP-seq of AR in SD70 treated/untreated LNCaP cells indicated that AR genome-wide occupancy was not affected upon SD70

treatment. In conclusion, our GRO-seq data suggested an inhibitory effect of SD70 on DHT-stimulated AR transcriptional program. Because KDM4 (JMJD2) family proteins have been reported to regulate AR-mediated gene transcriptional program through H3K9 demethylation, as well as its predicted ability to dock in the JMJ domain of this enzyme, we investigated whether SD70 might exert its function through the regulation of H3K9 methylation. Using in vitro demethylation assays, we found that SD70 indeed inhibited KDM4C demethylase activity, with substrate preference for H3K9me2. Quantitative demethylation assays showed the half inhibitory concentration (IC50) of SD70 to be □30 μM. Indeed, total H3K9me2 levels were significantly increased following SD70 treatment in 293T cells, whereas the methylated H3K36 (involved in gene transcriptional elongation) was also affected, being slightly decreased, consistent with a transcriptional inhibitory effect of SD70. SD70 enhancement of H3K9me2 levels was confirmed in both normal and KDM4C overexpression conditions.

To further investigate the potential role of KDM4C as one of the SD70 targets, in the context of an AR-dependent gene expression program, we performed conventional ChIP on four well-defined AR target gene enhancers, finding that KDM4C was located on these regulated enhancers and recruited in a DHT-dependent fashion. Based on these results, we knocked down KDM4C in LNCaP cells and evaluated AR target gene expression. With a reasonable knockdown efficiency (□70%), DHT-induced expression of classical AR target genes—including KLK3, KLK2, and TMPRSS2—was suppressed in a DHT-dependent fashion, consistent with a previous report that KDM4C promoted

expression of an AR-dependent gene expression by reporter assay. RNA-seq of KDM4C knockdown cells extended this observation to indicate a global effect of KDM4C on AR target gene regulation. To interrogate the genome-wide occupation of KDM4C, we performed a KDM4C ChIP-seq in LNCaP cells. As expected, a high percentage of KDM4C peaks were located on enhancers in intergenic regions. KDM4C tag density was increased on those AR enhancers that were cooccupied by SD70. H3K9me2 ChIP-seq revealed that SD70 reversed the DHT-induced decrease of H3K9me2 flanking transcriptional start sites (TSSs) and enhancer regions, which was also confirmed by qPCR on canonical AR target gene enhancers. These data indicated the SD70 and KDM4C-regulating AR programs reflect their actions at functional AR enhancers, reflected by the H3K9me2 mark, and support the notion that KDM4C is, at least, one of the important functional targets of SD70. However, as there are four members of the KDM4 family, it is quite possible that SD70 could exert effects on KDM4A and KDM4B, as well as KDM4D and the H3K9me2 demethylase, KDM3A. To extend the analysis to initially explore any potential role for SD70 predicted from these results, we performed a series of experiments to test the effects of SD70 on cell line and in vivo xenograft mouse models, focusing on castration-resistant prostate cancer cell models for testing the drug effect. First, we evaluated any growth-inhibitory effect of SD70 on hormone-independent CWR22Rv1 cells assessing titrated concentrations of SD70. This revealed a sharp transition at levels where the concentration is > 5 μM, exhibiting a striking growth-inhibitory phenotype.

To broaden our understanding of SD70 actions, we performed additional assays to assess other pathways that might be affected in cancer therapy. SD70 did not affect DNA-dependent RNA polymerase activity in vitro, and caused slight or no inhibition of DNA-dependent DNA polymerase activity in vitro, indicating it does not directly block DNA or RNA synthesis. Topoisomerases I and II have been identified as drug targets for multiple cancers. Because topoisomerase inhibitors have been widely used in cancer chemotherapy, we used an in vitro plasmid conformation capture assay used for detecting topoisomerase activity; we found that SD70 did not inhibit either topoisomerase I or II activity.

Next, we tested the potential efficacy of SD70 in vivo, assessing tumor growth in the CWR22Rv1 prostate cancer xenograft mice model. SD70 treatment (10 mg/kg drug daily) caused a dramatic inhibition of tumor growth in this model. In this in vivo trial, we did not observe any obvious toxicity, suggesting that future exploration of this drug for clinical cancer therapy might ultimately merit consideration. The increase in our mechanistic understanding of the origins of prostate cancer and the causes of castration-resistance has licensed different approaches to discovery of new types of chemical therapeutic. Although ever more comprehensive strategies for assessing effects in a high-throughput fashion have been introduced, a key requirement for newly discovered chemicals identified in such screens is to understand the molecular basis for their effects, particularly for drugs that affect transcription. Thus, here we have applied a strategy to take advantage of contemporary global deep sequencing techniques to assess specific patterns of genomic association of these new chemicals, in a

fashion analogous to the location of factor/cofactor recruitment to the genomic DNA, usually referred to as ChIP-seq. This technology, which we originally named “Drug-seq,” was designed to permit the assessment of region-specific, potentially regulated association of a potential new therapeutic agent in the genome. Indeed, while our manuscript was being prepared for submission, a similar concept piloted on BRD4/BET inhibitors, named “Chem-seq,” was published. Therefore, we have changed our terminology to conform with the already published term, Chem-seq. The basic approach is based on generating a modified chemical retaining the properties of the initial candidate compound, but permitting analysis by Chem-seq based on the presence of a biotin residue in the modified chemical to be investigated. In the example described in this paper, the analysis permitted a focused approach on transcriptional actions of the compound, even though the initial screen was based on effects on induced translocation events. We suggest that this approach will serve as a powerful adjunct to contemporary drug identification and optimization.

Use of a biotinylated drug probe permits the Chem-seq approach. To choose this method, we had considered the advantage of biotin–avidin affinity system, which has been successfully used for protein ChIP. Besides the tight and specific binding of biotin by avidin, biotin is a small molecule and is not known to affect the behavior of tagged protein, therefore we speculated it should give a reliable signal for the real drug. There are few naturally biotinylated proteins in mammalian cells, precluding a high background of Chem-seq, as most of these are cytoplasmic. Although SD70 represses much of the DHT-induced gene

expression program, it does not exclude that some SD70 up-regulated genes may also carry beneficial effects. For example, several genes that promote apoptosis could be induced by SD70.

A great number of histone demethylases have been known to play important roles in gene expression. Besides KDM4C, there are a few other H3K9 demethylases that have been reported to be involved in controlling AR target gene expression. KDM3A/JHDM2A, demethylates H3 mono- and dimethyl-K9 in vitro, binds to thePSA enhancer in a hormone-dependent manner and mediates H3K9 demethylation-dependent AR target gene activation. We do not know whether KDM3A is also located on SD70 target enhancers, because lack of ChIP-seq data may have been limited by antibody quality.

KDM4C is a putative oncogene that is involved in various cancers. KDM4C expression was significantly increased in prostate cancers relative to normal tissue, supporting the functional role of KDM4C in prostate cancer progression. Besides, KDM4C is a coactivator for HIF1 that required for breast cancer progression. KDM4C gene amplification was associated with a metastatic lung sarcomatoid carcinoma. Given that there are four members of the KDM4 subfamily, it is possible that SD70 may target other KDM4 family members in addition to KDM4C. KDM4A/JMJD2A and KDM4D/JMJD2D were also reported to interact with AR and work as AR coactivators. KDM4B/JMJD2B is found to be associated with estrogen receptor (ER) and is required for ER-regulated transcription, indicating the potential efficacy of SD70 in a variety of cancers.

Therefore, in the present study, we have identified by chemical library screening a molecule, SD70, and obtained data that characterize it as a potential drug for prostate cancer therapy, abetted by the Chem-seq data that revealed SD70 binding to AR-cobound enhancers, regulating DHT-induced gene transcriptional program. Establishing that SD70 functions as a JMJD2 inhibitor, we verified that it inhibited KDM4C demethylase activity. SD70 treatment, therefore, caused elevated H3K9me2 levels in enhancer and promoter regions, a probable component of the inhibitory effects on DHT target gene expression. Because both cell line and tumor mouse xenograft model experiments showed the remarkable inhibitory effects of SD70 on tumor cell growth, SD70 holds potential promise for ultimate use in prostate cancer therapy.

Materials and Methods.

Cell Culture and Transfection.

Prostate cancer LNCaP and CWR22Rv1 cells were obtained from American Type Culture Collection and cultured in RPMI-1640 containing 10% (vol/vol) FBS, and 293T cells were cultured in DMEM high glucose containing 10% (vol/vol) FBS. siRNA and plasmid transfections in 293T cells or siRNA transfection in LNCaP cells were carried out using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. The KDM4C construct was previous described. KDM4C siRNA (sc- 92765) was from Santa Cruz Biotechnology and negative control siRNA (1027281) was from Qiagen.

Antibodies.

Specific antibodies were purchased from the following commercial sources: anti-AR (N-20), anti-AR (441), and anti-jumonji domain-containing histone demethylase JMJD2C (S-15) were from Santa Cruz Biotechnology; anti–γ-Tubulin (T5326) and Monoclonal ANTI-FLAG M2 antibody (F1804) from Sigma-Aldrich Prestige Antibodies; anti-H3K9me2 (39375) from Active Motif; other histone mark antibody from Abcam; and anti-JMJD2C (A300-885A) from Bethyl Laboratories.

Cell Lysis and Immunoblotting.

Cells were washed twice with cold PBS and lysed in lysis buffer (50 mM Tris·HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100). For histone mark Western blot, samples were briefly sonicated, followed by boiling in SDS sample buffer (Invitrogen), and separated by 4–12% (vol/vol) Bis-Tris NuPAGE gel (Invitrogen). Then proteins were transferred to nitro- cellulose membrane (Bio-Rad) and Western blotting was performed following standard protocols.

RNA Purification and RT-PCR.

Total RNA was isolated from cells using RNeasy Plus Mini Kit (Qiagen) following the manufacturer's protocol. First-strand cDNA synthesis from total RNA was carried out using iScript cDNA Synthesis Kit (Bio-Rad). Resulting cDNA was then analyzed by quantitative real-time PCR (qPCR) using the SsoAdvanced SYBR Green Supermix (Bio-Rad) on Stratagene Mx3000 or the

CFX Connect Real-Time PCR Detection System. For expression level detection, GAPDH or Actin was used as the endogenous control.

Drug Screen and Induced Gene Translocation Assay.

LNCaP cells were cultured in hormone-deficient media for 3 d before treatment with a vehicle or compounds (10-μM final concentration) for 2 h, followed with combined treatment of dihydrotestosterone (DHT) (100 nM) and irradiation (IR) (50 Gy) for 24 hr. For certain experiments, LNCaP cells were transfected with validated siRNAs or expression vectors before DHT and irradiation treatment. The gene fusion events of TMPRSS2:ERG and TMPRSS2:ETV1 were examined by RT-qPCR using SYBR Green Master mix on an Mx3000 Real Time PCR system (Stratagene). Relative quantities of TMPRSS2:ERG and TMPRSS2:ETV1 fusion transcript were normalized to Actin and to the expression level of TMPRSS2. The relative amount of each fusion transcript was then calibrated to DHT+IR-treated control samples as appropriate. Samples without detectable fusion transcript after 40 cycles of amplification were indicated by 0.

Biotinylated SD70 ChIP-Sequencing.

Synthesized biotinylated SD70 (Biotin-SD70) was used to treat LNCaP cells (10-μM final con- centration) for 2 hr, followed by DHT (100 nM) for 1 hr. The ChIP-sequencing (ChIP-seq) experiment was performed as previously described. Briefly, □10^7 treated cells were cross-linked with 1% formaldehyde at

room temperature for 15 min. After soni- cation, the soluble chromatin was incubated with 50 μL Streptavidin Dynabeads (Life Technologies) for 2 hr. Complexes were washed and DNA extracted and purified by QIAquick Spin columns (Qiagen). Each reaction of KDM4C, androgen receptor (AR), or H3K9me2 ChIP-seq was performed by using 3 μg anti- JMJD2C (A300-885A; Bethyl Laborataries), 3 μg AR (N-20) antibody, or 10 μL H3K9me2 (39375; Active Motif) antibody, re- spectively. Anti-KDM4C (S-15) was used for ChIP-qPCR to confirm the ChIP-seq result. For ChIP-seq, the extracted DNA was ligated to specific adaptors followed by HT-sequencing (HT-seq) on an Illumina HiSEq 2000 system according to the manu- facturer's instructions. The sequencing reads were aligned to the hg18 (www.1000genomes.org/category/assembly) assembly by using Bowtie2 (http://bowtie-bio.sourceforge.net/Bowtie2/index.shtml). The HT-seq data were visualized by preparing custom tracks for the University of California Santa Cruz genome browser by using Hypergeometric Optimization of Motif Enrichment (HOMER).

The total number of mappable reads was normalized to 10 million for each sample. The ChIP-seq peaks were identified by using HOMER with the default parameter settings. For SD70 ChIP-seq, the peaks were filtered by the peak score: Only the peaks with a peak score > 10 were kept for downstream analysis. The cobound peaks were defined by a 100 bp overlap window. The read density (RD) profiles were computed by counting the average within a 6 kb window by using HOMER, and were displayed in Microsoft Office Excel. The heatmap matrices

were computed using HOMER and displayed in Multiple Experiment Viewer. The statistical significance of the difference between ChIP-seq RD profiles was assessed by Kolmogorov–Smirnov test in R [ks.test()] and by t test [t.test()].

Global Run-On Followed by HT-Seq.

LNCaP cells (10^7) were treated with synthesized SD70 or vehicle DMSO for 2 h followed by DHT (100 nM) for 1 hr. The Global Run-on followed by HT-seq (GRO-seq) experiment was performed as described. The image analysis and base calling were performed by using Illumina's standard computational analysis pipeline. The sequencing reads were aligned to the hg18 assembly using Bowtie2. To estimate gene expression, the sequencing reads were counted on the first 60 kb of the gene bodies by using HOMER. A RD threshold (i.e., the normalized GRO-seq read counts per kilobase) was used to exclude genes with low counts. A fold change (FC) higher than 1.5 was used to call the differentially regulated genes. Multiple gene isoforms were considered, as provided by HOMER genome annotation files (http://homer.salk.edu). The Gene Ontology (GO) analyses were carried by using HOMER as well.

RNA-Seq.

KDM4C or negative control siRNA was transfected into LNCaP cells once every other day in stripping media, 100 nM DHT, or vehicle 0.1% EtOH was added to the cell culture for 6 h after 4 d knockdown of KDM4C. Total RNA was extracted by RNeasy Plus Mini Kit (Qiagen) following the manufacturer's

protocol. Residual DNA was removed by DNase I on column digestion (Qiagen). Knockdown efficiency was detected by qPCR. The ribosome RNA was removed by a Ribo-Zero Magnetic Kit (epicentre) and then subjected to library construction by ScriptSeq v2 RNA-Seq Library Preparation Kit (epicentre), followed by deep sequencing. The RNA-seq reads were aligned to the human genome hg18 using Bowtie2 with ultrasensitive parameters. The RNA-seq reads were counted over gene exons and normalized to 10 million. EdgeR (www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/ edgeRUsersGuide.pdf) was used for statistical analyses of siCTL and siKDM4C samples, and an FC of 1.5 was used to call the dif- ferentially expressed genes. For RNA- and GRO-seq computational analyses, multiple gene isoforms were included, as provided by HOMER genome annotation files (http://homer.salk.edu).

In Vitro Demethylation Assay.

Briefly, 100 ng recombinant KDM4C (1-460aa, ab167940; Abcam) was incubated with 2 μg histone from calf thymus (Sigma) at 37 °C for 3 hr in histone demethylation buffer: 20 mM Tris·HCl (pH 7.4), 150 mM NaCl, 50 μM Fe(SO4), 1 mM a-ketoglutarate, and 2 mM ascorbate. Substrate methylation levels were analyzed by immunoblotting with specific antibodies. The membrane was subjected Pouceau S staining to quantify the loading amount. For nondemethylation control, SDS loading buffer (Invitrogen) was added directly to stop the reaction at time 0. Quantitive demethylase assay was performed with an

EpiSeeker KDM4/ JMJD2 Activity Quantification Assay Kit (Fluorometric) (ab113462; Abcam) following manufacturers' instruction. The 530 nm excitation/590 nm emission was read by Infinite M1000 PRO (TECAN).

DNA-Dependent DNA Polymerase or RNA Polymerase Activity Assay.

For DNA polymerase activity, qPCR with a gradient concentration of SD70 was performed on □200-bp ActB transcript using LNCaP cDNA as template. qPCR was done by VeriQuest Fast Sybr qPCR Master Mix (Affymetrix). For RNA polymerase activity assay, 25 ng PCR product of a 548-bp lncRNA PRNCR1 fragment with 5′ additional T7 polymerase binding sequence was used as the template for in vitro transcription (Promega). Titrated SD70 at the in- dicated concentration was mixed with the reaction and the RNA product was denatured and followed by electrophoresis after reacting at 37 °C for 2 hr.

Topoisomerase Assay.

One unit of topoisomerase I (M0301S; New England Biolabs) or topoisomerase II (TH2000-1; TopoGEN) was incubated with 200 ng Supercoiled pHOT-1 plasmid (TopoGEN) at 37 °C for 1 hr according to manufacturers' instructions. The reaction was terminated by adding GelPilot Loading Dye (Qiagen). The reaction products were analyzed by ethidium bromide-free electro- phoresis on 1% agarose gel, then stained by ethidium bromide (1 μg/mL), and photographed using a short-wavelength UV lamp (Syngene Imaging System; Imgen Technologies).

Cell Proliferation Assay.

The cell proliferation assay was performed using CellTiter 96 AQueous One Solution Cell Proliferation Assay (MTS; Promega). Briefly, cells were distributed in a 96-well plate with six replicates for each sample. After removing media and having been rinsed once by PBS, cells were supplied with 100 μL PBS mixed with 4 μL MTS reagent, followed by a 1 hr incubation at 37 °C in a 5% CO incubator. After incubation, 490-nm absorbance on each well was measured by a light absorbance reader.

Xenografts and Animals.

Animal studies were conducted in accordance with the protocols approved by the Institutional Animal Care and Use Committee (IACUC) of the University of California, Davis. Three million of CWR22Rv1 cells were suspended with matrigel in the ratio of 2:1 for s.c. injection into male athymic nude mice. Tumor progression was monitored by caliper mea- surement twice a week and the tumor volume was calculated according to the equation, v = length $\times$ width^2 $\times$ 1/2. When the tumor size reached between 150 and 200 mm, animals were randomly distributed to groups receiving 10 mg/kg of SD70 via i.p. injection or vehicle once a day with continuous tumor monitoring until the tumor burden exceeded the limit indicated by the IACUC humane endpoints (less than 20 mm in one dimension) for 3–4 wk. For SD70 drug preparation, SD70 powder was first dissolved in DMSO at 50 mg/mL, then diluted into 75% PEG300:25% D5W to

arrive at 2.5 mg/mL; 100 μL per animal was used for around 25 g of body weight.

Chapter 2, in full, is a reprint of the material as it appears in Proc Natl Acad Sci USA 2015. Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens. Jin C, Yang L, Xie M, Lin C, Merkurjev D, Yang JC, Tanasa B, Oh S, Zhang J, Ohgi KA, Zhou H, Li W, Evans CP, Ding S, Rosenfeld MG., 2014.

CHAPTER 3
Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.

The functional importance of gene enhancers in regulated gene expression is well established. In addition to widespread transcription of long non-coding RNAs (ncRNA) in mammalian cells, bidirectional ncRNAs referred to as eRNAs are transcribed on enhancers. However, it has remained unclear whether these eRNAs are functional, or merely a reflection of enhancer activation. Here, we report that 17β-estradiol (E2)-bound estrogen receptor α (ERα) on enhancers causes a global increase in eRNA transcription on enhancers adjacent to E2-upregulated coding genes. These induced eRNAs, as functional transcripts, appear to exert important roles for the observed ligand-dependent induction of target coding genes, causing an increased strength of specific enhancer:promoter looping initiated by ERα binding. Cohesin, present on many ERα-regulated enhancers even prior to ligand treatment, apparently contributes to E2-dependent gene activation, at least in part, by stabilizing E2/ERα/eRNA-induced enhancer:promoter looping. Our data indicate that eRNAs are likely to exert important functions in many regulated programs of gene transcription. ERα ChIP-seq analysis using 1 hour E2-treated (100nM) MCF7 cells revealed 31,052 ERα binding sites genome-wide, including only 902 on promoters, analogous to reported analyses and 7,174 ERα-bound potential enhancers based on the presence of H3K4me1 and H3K27ac. GRO-seq analysis of MCF7 cells in similar conditions, identified 1,309 E2-upregulated coding genes, of which 1,145 exhibited an E2/ERα-binding enhancer within 200kb from their transcription start

sites (TSS), thus considered as direct estrogen target genes and hereafter referred to as UP-genes. Of these, only 112 exhibited ERα binding to their promoters, consistent with suggestions that ERα occupancy on enhancers is a key strategy underlying E2-induced gene expression. Most E2-regulated enhancers displayed a rapid bidirectional activation of eRNAs, exemplified by the FOXC1 locus, which is about ~1.5 kb long by GRO-seq, although ~10% exhibited an apparent unidirectional eRNA transcription. These data suggest that eRNA induction in response to ERα binding is a predictive mark of enhancer activity. Binding of ERα did not cause clear alterations in enhancer marks on ERα-bound enhancers, such as H3K27ac. In contrast to androgen receptor-regulated genes, there was often more than one ERα-bound enhancer adjacent to UP-genes, exemplified by the FOXC1 locus. 83% of enhancers with detectable GRO-seq signals adjacent to UP-genes exhibited E2-induced eRNA upregulation; for the remaining 17%, the tag count was not sufficient to assign upregulation bioinformatically. E2-induction of eRNA was not observed on non-ERα-bound H3K27ac-marked enhancers. The median distance between enhancers exhibiting E2-induced eRNAs (n=1248, referred as UP-enhancers) and their closest UP-genes was ~52 kb, compared with a median distance of ~270 kb between enhancers exhibiting no E2 induction of eRNAs with UP-genes. ChIP-seq analysis revealed that UP-enhancers exhibited significantly stronger binding of ERα than enhancers not exhibiting eRNA upregulation. The proximal (< 200 kb) UP-enhancers constituted a majority of all UP-enhancers and exhibited a higher affinity for ERα than distal UP-enhancers. The strength of ERα binding is much higher on UP-enhancers than on 112 ERα-

bound promoters of coding genes showing E2 induction, while the remaining 790 ERα-bound promoters of genes with no E2-upregulation exhibit the weakest ERα binding. Based on GRO-seq analyses, we selected ten highly upregulated transcription units for further experimentation, each associated with adjacent UP-enhancers exhibiting ~2.5-5-fold E2-induction of eRNAs. Despite increasing evidence for crucial nuclear functions of lncRNAs, it remains an unresolved question whether eRNAs are merely a byproduct of enhancer activation or whether they might serve as key regulators of coding gene transcription. To begin to investigate the potential roles of ligand-induced eRNAs on gene activation events, both specific siRNAs and locked nucleic acid antisense oligonucleotides (ASO-LNAs) directed against each eRNA transcript were designed based on the peaks of eRNA exhibited by GRO-seq. To exclude off-target effects, experiments were performed with two different LNAs or siRNAs targeting each eRNA. With a high efficiency of transfection, both siRNA and LNA-mediated knockdown of the TFF1, FOXC1 or CA12 eRNAs revealed that, for each transcription unit, the induction of both the eRNA and of the adjacent coding gene, as assessed by QPCR and GRO-seq, respectively, was significantly inhibited. In contrast, these siRNAs/LNAs did not affect housekeeping genes tested or E2-regulated or non-E2-regulated transcription units more distal to the regulated enhancers. Ligand-induced increase of ERα binding occurred even after eRNA knockdown. Similar eRNA requirements for coding gene induction by E2 were observed based on knockdown of eRNAs adjacent to the PGR, SIAH2, KCNK5, P2RY2, SMAD7, GREB1 and NRIP1 genes using either of two siRNAs/LNAs. GRO-seq data were

consistent with the notion that LNA against eRNA reduces the levels of eRNA transcript post-transcriptionally, but not its nascent transcription. Knockdown of an eRNA on ERα-bound distal enhancer (~222 kb from FOXC1 TSS) (FOXC1-e2) that did not exhibit E2-induced eRNA and with lower ERα binding affinity, did not affect neighboring FOXC1 gene induction, further indicating that eRNA induction potentially marks E2-regulated functional enhancers. While GRO-seq results already argue against any LNA-mediated transgene silencing, further assays, including methyl miner and enzyme digestion assay indicated no evidence of altered enhancer methylation on either of the FOXC1, P2RY2 orNRIP1 enhancers. Additional supporting evidence was provided by using an LNA targeting the sense transcript of GREB1-RR, a regulatory region near GREB1 gene, exhibiting overlapping-bidirectional transcription, where we observed no significant change in transcript levels from antisense strand by strand-specific QPCR. We also failed to observe any significant LNA effects on levels of total histone H3, H3K9me3 or H3K27me3 silencing marks on several targeted enhancers. Together, these data suggest that siRNA/LNA-mediated knockdown of eRNAs do not elicit transgene silencing of the interrogated enhancers. To independently validate that eRNAs per se are important for quantitative increases in target gene expression, we took advantage of a GAL4-Box-b tethering based reporter assay. For this we engineered a chimera RNA by fusing FOXC1 sense eRNA to five copies of Box-b RNA, permitting Box-b-FOXC1e RNA to be recruited by the RNA binding domain of λN protein fused with GAL4 DNA binding domain. Thus eRNA can be

artificially tethered to 5XUAS sites just downstream of the FOXC1 enhancer in the reporter plasmid where Luc is under control of the native FOXC1 promoter. We observed that the presence of full-length FOXC1 enhancer increased luciferase expression to ~2.5 fold when compared to random DNA in place of the enhancer. This effect was abolished when the sense eRNA sequence was deleted and substituted with 5XUAS sites, generating a non-functional "mis-sense" eRNA. Tethering of Box-b-FOXC1echimeric RNA, but not Box-b alone, could fully rescue the loss of activity of sense-eRNA deleted enhancer, while the antisense FOXC1 eRNA (-strand) could not. We confirmed the loss of plasmid-driven native FOXC1 eRNA expression from sense-eRNA deleted reporter construct and that GAL-4 tethering was not altered. These data further support the suggestion that the sequence-specific eRNA transcript per se, rather than merely the process of enhancer transcription, is required for the actions of the eRNA on enhancer-dependent coding gene activation events. This observation is consistent with a study of roles of ncRNAs in p53-dependent gene activation and regulation of the SNAI1 gene.

We next investigated whether enhancer:promoter looping events are induced in the E2-activation process, employing a strategy very analogous to 5C, the 3D-DSL, to study the spatial organization of genomes. We first examined two E2-regulated transcription units: P2RY2 and KCNK5. For P2RY2, E2 treatment significantly increased the specific promoter:enhancer interaction, also induced a new E2-dependent interaction between enhancer and gene terminus region. Similarly, for the KCNK5 locus, E2 treatment caused clear increase in

promoter:enhancer interactions, as well as loops to the terminator region of the gene and the other that might represent an enhancer-specific loop. These observations indicate that a major effect of ligand is to enhance specific promoter:enhancer interactions in parallel to induction of eRNA.

We next investigated whether E2-induced enhancer:promoter interactions are affected by eRNAs. The NRIP1locus exhibited specific enhancer:promoter and promoter:gene terminus loops, whereas treatment with LNA against its eRNA caused a marked inhibition of these interactions and E2-activation of the NRIP1 gene. siRNA-mediated GREB1 eRNA knockdown also coordinately inhibited GREB1 gene induction and the two specific enhancer:promoter interactions induced by E2 and two additional non-enhancer loops. Together these experiments indicate that estrogen causes quantitative, as well as some qualitative, alterations in the interactions between enhancers and coding gene promoters, and eRNAs are apparently of functional importance for robust enhancer:promoter interactions.

To address the possibility that eRNAs might also work in trans, we first estimated the absolute expression levels of the eRNAs, finding that most of the investigated eRNAs were present at levels of < 5®C15 copies/cell, although several, including TFF1e, present at ~70®C95 molecules/cell, suggesting that these eRNAs were likely to function primarily in cis. Furthermore, we employed the Chromatin Isolation by RNA Purification (ChIRP) to identify potential sites where FOXC1 eRNA localizes in the genome; despite robust detection of FOXC1eRNA from its transcribing site establishing the efficacy of the biotin-

labeled probes used, only 15 peaks could be confidently called, in none was the nearest genes an E2-regulated gene. In addition, QPCR analysis following knockdown of FOXC1 eRNA revealed no significant effects on E2 activation of NRIP1, TFF1, PGR or KCNK5 genes. GRO-seq following LNA transfection against FOXC1 eRNA revealed that a large majority (> 95%) of the E2-upregulated coding genes continued to exhibit clear E2-dependent upregulation. Therefore, any trans-effects of eRNAs are likely to be relatively infrequent or quantitatively small. Of course, there are inevitably indirect effects observed after knockdown of any eRNA that down-regulates a functional coding gene. However, at least for a few gene areas, there may be effects of enhancer-based long-range interactions. We identified at least one such example, between the NRIP1 and TFF1 loci, separated by ~27 Mb on chr 21, exhibiting an E2-induced increase of co-localization by FISH. Surprisingly, knockdown of NRIP1e eRNA by LNA caused a clear decrease in the interactions between these two genomic loci, suggesting that such E2-induced co-localization was eRNA-dependent.

Since several studies have strongly implicated a role for Cohesin in chromosomal interactions and enhancer:promoter looping events, we investigated whether Cohesin was involved in the observed eRNA functions. First, co-immunoprecipitation showed that ERα can interact with Cohesin subunits. Rad21 ChIP-seq revealed ~30®C40% of its binding sites overlap with putative H3K4me1/H3K27ac-marked enhancers in MCF7 cells. After E2 treatment, both ChIP-seq and ChIP-QPCR data revealed a reproducible, but modest (50®C200%), increased occupancy of the Cohesin subunits Rad21 and SMC3 on the

interrogated enhancers, as exemplified by FOXC1e, NRIP1e and TFF1e. By in vitrotranscribed (IVT) RNA pull-down, the investigated eRNAs could pull-down SMC3 and Rad21 from MCF7 nuclear extracts. RIP-QPCR confirmed the interaction between Cohesin and several eRNAs, but not with GAPDH or another nuclear RNA, TUG1. To test possible direct or indirect involvement of RNAs in Cohesin recruitment to enhancers, we found that RNase treatment caused some decrease of the Cohesin level in chromatin-bound fraction of cells. Knockdown of specific eRNAs by LNA or siRNA resulted in a decrease of Cohesin recruitment, to enhancers in response to E2 with essentially no significant alteration of the H3K4me1 mark, or in ligand-dependent increase of ERα recruitment. Expression levels of Cohesin subunits and ERα were not affected by knockdown of eRNAs. siRNA-mediated depletion of Rad21 caused loss of enhancer:promoter interactions, both basal and E2-induced, when assessed by 3C assay for the NRIP1 and GREB1 gene loci. When we tested the role of Cohesin in the estrogen transcription program by GRO-seq, we noted that siSMC3 caused a broad inhibition of coding gene activation by E2, with only ~34% of E2-upregulated genes remaining induced. Similarly, Rad21 knockdown inhibited E2-induction of genes, as revealed by the five targets evaluated. We excluded alterations in levels of ERα as the cause for these dramatic effects of Cohesin depletion. Based on these results, we are tempted to speculate that many regulatory genomic regions, such as enhancers, harbor the Cohesin complex, that "poises" the enhancer for the stable eRNA-induced looping necessary for gene activation events, but could not exclude the possibility that the role of Cohesin

could also reflect non-enhancer-based regulation.

Despite the discovery of enhancers more than 35 years ago, full understanding of the mechanisms by which they regulate gene expression has been a relatively resistant problem. Here, we have provided several lines of evidence that induced eRNA transcripts are functionally important for the actions of estrogen-regulated gene enhancers, at least in part by contributing to the dynamic generation/stabilization of enhancer:promoter looping between the regulated coding transcription units and these ERα-bound enhancers.

Materials and Methods.

Bioinformatics characterization of ERα enhancers.

The ERα-H3K27ac co-bound regions are defined as that the distance from the center of an ERα peak to the H3K27ac peak-occupied region is °‹ 1 kb. Overall, two methods were used to assign the ERα bound enhancers to E2 up-regulated genes: 1) identifying the E2-upregulated coding genes from GRO-seq first and coupling each of them to their closest ERα-H3K27ac co-bound enhancer within certain distance (200 kb) (a “gene-centric” view); 2) characterizing the ERα-H3K27ac co-bound enhancers first and then coupling each of them to their closest TSS that belongs to 1,309 E2 up-regulated coding genes (an “enhancer-centric” view). The comparison of ChIP-seq tag intensity, GRO-seq transcription levels or distances between different categories are presented as boxplots by using either log or normal scales. The p-values were determined by two-tailed Student's t-test.

3D-DSL.

The DSL ligation products were prepared. 3D-DSL was performed. Briefly, after 3C efficiency estimation, equal amount of 3C chromatin was biotinylated using the Photoprobe Kit (Vector Lab). Donor and acceptor probe pools (20fmol per probe) were annealed to the biotinylated 3C samples at 45 °C for 2 hours followed by 10 min at 95 °C. The biotinylated DNA was immunoprecipitated with magnetic beads conjugated to streptavidin, and during this process unbound oligonucleotides were removed by stringent washes. The 5Åå-phosphate of acceptor probes and 3Åå-OH of donor probes were ligated using Taq DNA ligase at 45 °C for 1 hr. These ligated products were washed and eluted from beads and then amplified by PCR using primers A and B-AD (or Primer B-BC1 and -BC2 if bar coding was used) for deep sequencing on the Illumina HiSeq 2000 using Primer A as sequencing primer.

3D-DSL data analysis.

After removing the adaptor sequences, the reads are aligned to a custom library that includes all the combinations donors-acceptors. The alignment was performed with Novoalign, and the reads were counted for every possible interaction by using custom Perl scripts (available upon request). The reads that were generated by donor-acceptor ligations on the same restriction site are removed: the remaining number of reads included both intra- and inter-chromosomal interactions. We used the median value (~6 millions) of the all the

samples from the same sequencing run for normalization; the reads accounting for ligation products in unligated controls were subtracted. In addition to standard tools such as my5C and HiTC we used an intensity-based method to characterize the set of interactions a related method was also in 3C-seq procedures. A p-value is assigned to an interaction based on Poisson probability distribution function, p(x) = eÅ|λ * λx/x!, where p(x) is the probability of an interaction, x is the interaction intensity, and λ is the average interaction intensity considering all the potential interactions in the library, i.e. the ratio (total number of usable reads)/(total number of all possible interactions given the set of acceptors and the set of donors). The p-values were corrected for multiple testing by using Bonferroni correction method. In addition, for each interaction, we define supplementary parameters, such as 1) fold enrichment over Poisson's λ, and 2) a fold enrichment over background (where the background represents the average intensity of the ligations between the probes on the neighboring restriction sites). We consider significant for downstream analyses the interactions that meet the following criteria: a) a (corrected) p-value < 0.01, and b) a fold enrichment over background > 2, although for display purposes, in the figure plots, we may also show the weak interactions. To generate 3D-DSL plots, Matlab was used; a 10 kb window was used to bundle the interactions, except for a 20 kb window for NRIP1. The interactions were plotted using a Bezier curve between the two positions with the third point in the middle of the positions with the y-axis corresponding to the log 10 intensity. For example, if the two x-axis positions are 1 and 2, and the intensity is 4, a Bezier curve is drawn between (1,0),(1.5,4), and

(2,0). The peak locations were then added on the bottom of the plot as stated in the legend. Interactions at distances generally < 10 kb were not plotted for the NRIP1 locus.

Chapter 3, in full, is a reprint of the material as it appears in Nature 2013. Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG., 2013.

CHAPTER 4
Enhancer activation requires trans-recruitment of a mega transcription factor complex.

Enhancers provide critical information directing cell-type-specific transcriptional programs, regulated by binding of signal-dependent transcription factors and their associated cofactors. Here, we report that the most strongly activated estrogen (E2)-responsive enhancers are characterized by trans-recruitment and in situ assembly of a large 1–2 MDa complex of diverse DNA-binding transcription factors by ERa at ERE-containing enhancers. We refer to enhancers recruiting these factors as mega transcription factor-bound in trans (MegaTrans) enhancers. The Mega-Trans complex is a signature of the most potent functional enhancers and is required for activation of enhancer RNA transcription and recruitment of co-activators, including p300 and Med1. The MegaTrans complex functions, in part, by recruiting specific enzymatic machinery, exemplified by DNA-dependent protein kinase. Thus, MegaTrans-containing enhancers represent a cohort of functional enhancers that mediate a broad and important transcriptional program and provide a molecular explanation for transcription factor clustering and hotspots noted in the genome.

Functional specialization and precise patterning of different cell and tissue types are vital for all metazoans, which also generate cell- or tissue-specific gene expression patterns. Enhancers, initially defined as DNA elements that act over a distance to positively regulate expression of protein-encoding target genes, are the principle regulatory components of the genome that enable such cell-type-specific and signal-dependent patterns of gene expression (Banerji et al., 1981; Shlyueva

et al., 2014). Each cell type harbors more than 100,000 candidate enhancers in 358 Cell 159, 358–373, October 9, 2014 a2014 Elsevier Inc. humans, vastly outnumbering protein-coding genes (Bernstein et al., 2012; Heintzman et al., 2009; Shlyueva et al., 2014). This makes it very important to be able to predict and understand which enhancers are actually functionally required for target coding gene transcriptional regulation.

Enhancer activation requires the presence of specific recognition sequences for the cooperative recruitment of DNA-binding transcription factors (TFs) and their cofactors that initially activate gene expression (Rosenfeld et al., 2006). While the role of a large number of coactivator complexes and their associated enzymatic activities is well established (Rosenfeld et al., 2006), the precise biochemical mechanisms by which so many coactivators are recruited and required for the different functional activities at specific enhancer sites remain incompletely understood. Global genomic technologies have uncovered characteristic markers of enhancers and have provided clues as to their activation. Features that have been used to predict enhancers that are likely to be functional include the levels of enhancer RNAs (eRNAs) transcribed from enhancer-like regions in the genome (Li et al., 2013), the presence of the histone acetyltransferase p300/CBP (Visel et al., 2009), the timing of RNA Pol II occupancy (Bonn et al., 2012), and levels of H3K4me2 and H3K27ac (Chepelev et al., 2012; Heintzman et al., 2009). However, because enhancers identified using these features are not equally functional, additional methods are needed to distinguish the enhancers with different activation potential. There are 2,600

DNA-binding TFs encoded by the human genome (Babu et al., 2004), with 200–300 TFs being expressed in each cell type (Vaquerizas et al., 2009). A long-standing question is how different TFs collaborate to regulate the enhancer network in a specific cell type. With the large expansion of genome-wide binding data, DNA-binding transcription factors were noted to co-bind to some so-called ‘‘hotspot’’ regions or to cooperatively cluster to some functional enhancers in various organisms or cell lines (Junion et al., 2012; Rada-Iglesias et al., 2012; Siersbæk et al., 2014a, 2014b; Wilson et al., 2010; Yan et al., 2013). However, the underlying mechanism(s) and functional significance of this phenomenon are not well understood.

Recently, the idea of clustered enhancers associated with critical developmental or cancERassociated transcription units has been proposed (Hnisz et al., 2013; Lovén et al., 2013; Whyte et al. 2013). The initial definition of this super-enhancer model was described as clusters of enhancers spanning > 8–10 kb, occupied by critical DNA-binding transcription factors at their cognate binding motifs (Lovén et al., 2013; Whyte et al., 2013). These clustered super-enhancers control key coding transcription units in stem cells or various disease states and exhibit high levels of co-activators, which are suggested to contribute to gene activation. Cancer cells were also noted to acquire super-enhancers regulating oncogene drivers (Hnisz et al., 2013; Lovén et al., 2013).

While the super-enhancer model can explain the higher expression levels for a small number of genes in some environments, it also highlights the need for exploring the functional activities of single enhancers in the regulation of coding

genes critical for development and disease and understanding the phenomenon of TF clustering in short-range genomic regions.

Here, we report a signature of the functionally active estrogen-regulated enhancers, particularly the 1,333 most active ERa enhancers linked to target coding gene activation. This signature is the selective recruitment in trans of an apparent complex of other DNA-binding TFs, including RARa/g, GATA3, AP2g, STAT1, AP1, and FoxA1. By gel filtration, we found these TFs migrated with ERa as a 1–2 MDa complex(es), referred to as the Mega-Trans complex. The MegaTrans complex is almost invariably recruited to functional ERa-bound enhancers, 22% of which fit the criteria of being components of super-enhancers. Furthermore, the MegaTrans complex is required for activation of the functional enhancers, apparently based in part on specific recruitment of enzymes. This is exemplified by the functionally important recruitment of the DNA-dependent protein kinase to ERa-regulated enhancers by RARs. The MegaTrans complex, in turn, is also required for activation of eRNA transcription and recruitment of coactivators, including p300 and Med1, and thus exerts critical biological functions, conceptually parallel to what has been proposed for super-enhancers. ERa functions as a central transcription factor for gene programs that mediate cell growth and proliferation, and it accomplishes this role primarily through enhancer regulation.

Among the total 7,174 ERa-bound enhancers, a subset of 1,333 enhancers that are located in proximity (< 200 kb) to their regulated coding transcription units have proved to be the most significantly activated upon

estrogen stimulation according to levels of H3K27ac and increased eRNA transcription and appear to constitute the most potent functional enhancer program (Li et al., 2013).

Our current study was initiated by investigating the possible functional mechanisms by which RARs on retinoic acid response element (RARE)-containing enhancers mediate RA-induced coding gene transcriptional programs, as well as the functional role(s) of RAR at enhancers that accommodate the effects of other signals, such as E2-induced coding gene transcriptional programs (Hua et al., 2009; Ross-Innes et al., 2010). To distinguish the possible binding in cis (the chromatin association of a transcription factor through direct DNA binding at its recognition sites) and in trans (the chromatin association of a transcription factor through protein-protein interaction) functional models of RAR, we engineered MCF7 to express a bacterial biotin ligase (BirA) that can biotinylate a biotin ligase recognition peptide (BLRP)-tagged protein in vivo. Under control of a Tet-On promoter, wild-type RAR and two DNA-binding domain mutants that cannot bind to RARE DNA sites were expressed at similar levels as the endogenous proteins upon doxycycline induction. Using these lines, we first performed biotin chromatin immunoprecipitation sequencing (ChIP-seq) for wild-type (WT) and mutant RARa/g (RARb is not expressed in MCF7 breast cancer cells) upon RA and E2 stimulation. Comparing wild-type and non-DNA-binding mutants, we found that 15,000/18,000 of WT RARa/g-bound sites required the intact RAR DNA-binding ability because binding was lost with mutant RARs, and none of these sites were bound by ERa. Among these 15,000

sites, 3,540 were enhancers that exhibited RA activation, exemplified by the 700 most active RAR cis-binding enhancers, which showed significant RA-induced eRNA and gene target activation by global run-on sequencing (GRO-seq). However, there were 3,000 RARa/g binding sites that did not depend on RAR DNA-binding ability. Remarkably, we found that both RARa and RARg were recruited to virtually all of the ERa-bound 1,333 active enhancers in response to E2. This observation is consistent with previous evidence that RAR can bind to ERa binding sites, although conflicting conclusions were reached regarding its activating or repressive effects (Hua et al., 2009; Ross-Innes et al., 2010). However, ERa did not exhibit colocalization with RARs on ERa non-active enhancers.

By comparing the binding patterns of wild-type and two non-DNA-binding mutants, we found the binding of RARs on the 1,333 ERa active enhancers was in trans. Knockdown of either RARa or RARg caused a significant decrease in both E2 -dependent induction of eRNAs and activation of target coding genes, while knockdown of both caused almost complete inhibition, as assessed by quantitative PCR (qPCR) of targets such as GREB1 and TFF1. The knockdown of RARa and RARg, which was confirmed for both RNA and protein levels, inhibited RA induction of the HoxA1 gene target as expected. Boxplot analysis of the GRO-seq experiments showed that the presence of RARs was required for effective induction of both eRNAs and target coding gene transcription units upon E2 treatment. RARa/g knockdown also inhibited classical RAR cis-bound enhancers and their target genes. Thus, while RAR binding in cis

activates a distinct RA-responsive transcriptional program, its recruitment in trans is also required for effective E2-dependent activation of ERa-bound functional enhancers. Next, we utilized wild-type and pBox mutant RARg to test their ability to rescue ERa-regulated enhancer function following endogenous RARg knockdown. Intriguingly, the non-DNA-binding mutant receptor continued to be effectively recruited to the ERa-bound regulatory enhancers at the GREB1 gene and was capable of restoring full E2-dependent GREB1 gene activation in rescue experiments.

However, as expected, it failed to activate the cis-bound, RAR-regulated HoxA1 gene. Administration of ICI 182780 to knockdown ERa caused a loss of RAR binding at the ERa-regulated enhancers, but did not alter the binding of RARa or RARg at activated enhancers harboring cis RAR binding sites. Knockdown of RAR did not cause downregulation of ERa RNA or protein levels and did not affect the ERa binding pattern on ERa active enhancers.

Collectively, our data indicate that ERa selectively recruits RARa and RARg in trans on the functional enhancers regulating the most robustly activated target coding genes and that this strong activation depends on the ERa-mediated trans-binding of RARs. These findings prompted us to examine the behavior of additional DNA-binding TFs associated with ERa, based on previously reported mass spectrometry analysis of proteins that coimmunoprecipitated with ERa (Mohammed et al., 2013) as well as our own confirmatory data. From these ERa complex data, we noted a number of DNA-binding transcription factors associated with ERa, including RARg, GATA3, AP2g, STAT1, and, intriguingly,

FoxA1. To complement these observations, we also examined the proteins associated with RAR following pull down from MCF7 cells stably expressing, at physiological levels, biotin-tagged RARa. In addition to RARa, RXRs, and many well-known cofactors for nuclear receptors, GATA3 was also detected along with other DNA-binding proteins including AP2g, STAT1, c-Fos, and FoxA1. We then performed gel filtration analysis on nuclear extracts prepared from MCF7 cells in the absence of DNase treatment and analyzed all fractions for ERa, RARa/g, GATA3, and the other DNA-binding transcription factors identified in the mass spectrometry analysis. This analysis revealed co-elution of ERa, RARa, RARg, GATA3, AP2g, FoxA1, STAT1, c-Fos, and other proteins in an estimated 1–2 MDa complex(es).

These components were all present in ERa-immunoprecipitates from nuclear extracts and their association was enhanced upon E2 treatment. Importantly, knockdown of nuclear ERa by administration of ICI 182780 caused a virtual loss of the entire complex associated with ERa by gel filtration analysis and recruitment of each factor to ERa-bound functional enhancers. Thus, the material co-migrating in the gel filtration represented proteins interacting as a complex with ERa rather than artifacts. This complex remained intact in the presence of 250 mM NaCl, but was lost under 600 mM NaCl high-salt conditions.

To further investigate the hypothesis that the ERa-dependent trans-recruitment/assembly of other DNA-binding transcription factors occurs only in situ at ERa active enhancers, we first confirmed that the interactions between ERa and the TFs were dependent on DNA. Using a non-DNA-binding ERa pBox

mutant, which is incapable of binding the estrogen response element (ERE) motif (Stender et al., 2010), we could show that this mutation abolished the interactions of ERa and these associated TFs. As a control, a comparable RARa pBox mutant did not affect its interaction with ERa and these TFs. These data suggest that RARa and other TFs are recruited by ERE-bound ERa to its activated enhancers; thus, the entire complex is assembled in situ on ERa-bound enhancers.

To further confirm that these factors were, indeed, co-recruited to the same transcription units, rather than the consequence of differential recruitment behavior in different cell populations, we performed serial pairwise two-step ChIP analyses to assess the co-recruitment of RARa with ERa, GATA3, FoxA1, AP2g, and STAT1 on the same ERa-bound enhancers. Using a BLRP-tagged RARa stable cell line, two-step ChIP was performed with biotin-streptavidin pull-down of RARa in the first round followed by immunoprecipitation with antibodies for RARa (as positive control), ERa, GATA3, FoxA1, AP2g, and STAT1. In each case, we found that these proteins were present on the interrogated active enhancers, including the GREB1 enhancer. In contrast, as a control, this was not the case for the RAR cis-bound enhancer regulating the HoxA1 transcription unit. Thus, the MegaTrans complex was co-recruited to ERa-bound active enhancers but not to functional enhancers that directly bind RARa in cis. RARa and the other TFs also were not present at ERa-bound, non-active enhancers. Double-ChIP experiments performed with a BLRP-tagged GATA3 stable line similarly demonstrated the co-binding of GATA3 with ERa and all of the other TFs at ERa active enhancers but not at either the HoxA1 enhancer or ERa non-active

enhancers.

To explore the possible functional consequences of the additional ERa-interacting transcription factors, we next explored the potential recruitment and function of GATA3 on ERa active and non-active enhancers. ChIP-seq experiments revealed, as in the case of RARa and RARg, that GATA3 was recruited in an E2-dependent fashion to active enhancers but not non-active enhancers. Because we found the presence of GATA3 on functional ERa-bound enhancers that did not harbor apparent GATA3 cis-binding elements by motif analysis, we again assessed the possibility that GATA3 was recruited in trans to these ERa-bound active enhancers. Knockdown of ERa by administration of ICI 182780 inhibited GATA3 recruitment to ERa active enhancers. Because direct or indirect ERa and GATA3 interactions were suggested by immunoprecipitation experiments, we investigated the consequences of disrupting the ability of GATA3 to bind to cognate DNA sites by two different mutations of the second zinc finger that is required for cis-binding of GATA3 (Nesbit et al., 2004). We generated inducible BLRP-tagged stable lines expressing wild-type and the two DNA-binding mutants at physiological levels, and biotin ChIP-seq revealed they were equivalently recruited, apparently in trans, to these ERa-bound active enhancers. By comparing the ChIP-seq data for wild-type and DNA-binding mutants, we found that among 18,000 wild-type GATA3 binding peaks 5,000 were retained in the two GATA3 mutants, and these trans-binding sites featured ERE as the top motif by Homer analysis. For the 13,000 cis-binding peaks, GATA motifs were enriched and a heatmap of the non-ERa enhancers

containing a GATA motif was used to confirm a total loss of binding of the two non-DNA-binding GATA3 mutants.

Using qPCRs or GRO-seq analysis, we explored the consequences of specific siRNA-mediated knockdown of GATA3 on E2-dependent induction of eRNAs. We found a dramatic inhibition of the eRNA activation events on active enhancers but no effect on ERa-bound non-active enhancers or non-ERa-bound enhancers. The same inhibition effects were also found for gene body expression of the targets of these 1,333 ERa active enhancers. Knockdown of GATA3 did not affect ERa gene expression at either the RNA or protein level or ERa binding at active enhancers. Thus, GATA3 and RARs, as components of a complex of DNA-binding TFs associated in trans with ERa on active enhancers, are required for E2-dependent enhancer activation.

We next investigated whether other DNA-binding transcription factors in the MegaTrans complex co-migrating with ERa were also recruited to E2 -actived enhancers even in the absence of their cognate DNA-binding elements. We reviewed our own and published ChIP-seq data from MCF7 cells for other DNA-binding TFs present in the MegaTrans complex (Joseph et al., 2010; Theodorou et al., 2013). E2 -regulated active enhancers were found to harbor AP2g, FoxA1, c-Jun, and c-Fos, along with RARa/g and GATA3, but these TFs were not present on non-active enhancers. Similar to RARa/g and GATA3, the recruitment of the other TFs was also increased by E2 and abolished by knockdown of nuclear ERa using ICI 182780.

In order to investigate whether, in fact, all DNA-binding transcription

factors present in the MegaTrans complex were recruited in trans to ERa functional enhancers, a series of DNA-binding domain mutations were generated for AP2g, c-Fos, c-Jun, and STAT1. ChIP-qPCR data on the GREB1 and TFF1 enhancers showed that the binding of the non-DNA-binding mutants at these two ERa active enhancers was comparable to that of the wildtype proteins, which confirms the trans-recruitment of these TFs by ERa.

Based on the roles of RARs and GATA3 on ERa active enhancers, we evaluated the functional effects of other recruited transcription factors. Beginning with AP2g, we found that, in addition to its recruitment in response to E2 on ERa regulatory enhancers, knockdown of AP2g caused a dramatic inhibition of eRNA and target coding gene expression, asassayed by both qPCR and GRO-seq. Similarly, as STAT1 was also recruited to ERa-bound enhancers, we evaluated its effect on two well-described ERa bound/regulated enhancers. Again, we found a functional contribution to the outcome of E 2 -induced activation of enhancer transcription and target coding gene expression. The same regulatory effects were also demonstrated upon knockdown of two AP1 components, c-Jun and c-Fos, that were present in the MegaTrans complex. To begin to assess the interdependency of the components of the MegaTrans complex on recruitment to ERa-bound functional enhancers, we tested the consequences of knockdown of RARa/g, GATA3, and AP2g on GREB1 and TFF1 enhancer occupancy. We found a marked inhibition of recruitment of other MegaTrans components upon knockdown of RARa/g and GATA3 but not by knockdown of AP2g.

These experiments raised a question about potential differences in DNA

sequence features between ERa active enhancers that bound the MegaTrans complex and ERa non-active enhancers that did not. Comparison of the EREs between these two groups revealed that its frequency and the primary consensus sequences were essentially identical. In contrast, and in accord with the well-known importance of FoxA1 as a pioneer factor for ERa binding (Hurtado et al., 2011), we noted that the functional, MegaTrans-bound 1,333 ERa active enhancers generally harbor a FoxA1 binding motif within 200 bp of the ERE sites, while the FoxA1 motif was virtually absent on the nonfunctional, non-MegaTrans-bound ERa enhancers. Indeed, the affinity for ERa is > 90% lower on the nonfunctional than functional ERa-bound enhancers. Consistent with FoxA1 functioning as a key determinant of ERa binding (Hurtado et al., 2011), our data showed greatly reduced binding of ERa at the 1,333 ERa-bound active enhancers upon FoxA1 knockdown.

Because FoxA1 appears to be a component of the MegaTrans complex based on gel filtration and coimmunoprecipitation (co-IP) data and also exhibits E2-induced binding at the 1,333 ERa active enhancers, we speculate that FoxA1 potentially plays dual roles in the binding of ERa to functional enhancers and in ERa-dependent recruitment of the MegaTrans complex. Indeed, knockdown of FoxA1 caused a dramatic impairment of ERa binding on the functional ERa enhancers, which was accompanied by a loss of recruitment of the MegaTrans complex on this functional enhancer cohort and inhibition of both eRNA and gene body activation. Thus, FoxA1 is distinct from the other DNA-binding TFs in the MegaTrans complex that apparently do not affect ERa binding upon knockdown.

A basic aspect of the mechanism by which MegaTrans components function is their requirement for effective activation of E2-induced eRNAs on the functional enhancers. Accordingly, we assessed the recruitment of the coactivator p300 by qPCR and ChIP-seq upon knockdown of RARa/g or GATA3. All of these knockdowns inhibited the E 2 -induced accumulation of p300 on activated enhancers, consistent with a previous report of a role for RARa in p300 recruitment (Ross-Innes et al., 2010). Based on the importance of Mediator complex for enhancer function, putatively due to its roles in enhancer:promoter looping events (Kagey et al., 2010), we also evaluated the effects of RARa/g and GATA3 knockdown on E2-dependent recruitment of Med1 to functional enhancers by qPCR, finding a dramatic inhibition following these knockdowns. This result was confirmed genome-wide by ChIP-seq.

Based on the criteria developed in the initial description of super-enhancers (Hnisz et al., 2013; Whyte et al., 2013), we assessed the number of super-enhancers in MCF7 cells by Med1 ChIP-seq under both –E2 and +E2 conditions. While there are only 122 super-enhancers under the –E2 condition, E2 treatment increases the total to 320 such enhancers, of which 212 contained at least one ERa-bound functional enhancer, including one at the c-Myc gene locus. Thus, only 300 of the 1,333 ERa-bound functional enhancers characterized by MegaTrans complex fulfill the current definition of being located in super-enhancers. The efficacy of this subset of 300 ERa-bound active enhancers was only slightly better than the other 1,033 ERa-bound active enhancers with respect to eRNA induction. Thus, the functional strength of the ERa-bound enhancers,

irrespective of their presence in a super-enhancer, is predicted by the presence of the MegaTrans complex. Actually, for the 212 super-enhancers that contain ERa active enhancers, their Med1 levels were also dependent on the E2 signal. Interestingly, we observed greatly reduced levels of Med1 at these 212 super-enhancers following knockdown of RARs, suggesting that MegaTrans enhancers are important constituents in the function of these clustered super-enhancers.

Based on the presence of specific non-transcription factor components in the mass spectrometry analysis of RARa-associated proteins, we evaluated the functional significance of these additional proteins. We elected to focus on DNA-dependent protein kinase (DNA-PK), comprising the catalytic subunit DNA-PKcs, Ku70, and Ku80, because all three DNA-PK subunits were present in the RARa pull down as revealed by mass spectrometry. We confirmed these associations by co-IP and western blot analysis. DNA-PKcs has previously been reported as a component of the ERa complex that directly phosphorylates S118 of ERa (Foulds et al., 2013), and we confirmed that knockdown of DNA-PKcs partially impacted phosphorylation of ERa S118 without affecting ERa binding at ERa active enhancers. Using a specific antibody against DNA-PKcs for ChIP analysis, we first evaluated the temporal kinetics of its potential recruitment on the GREB1 and TFF1 enhancers, finding recruitment at 10 min following E2 treatment of MCF7 cells In addition, a specific antibody against phosphorylated ERa S118 revealed strong enrichment that peaked at 20 min, slightly after the recruitment of ERa and DNA-PKcs. Based on these observations, we conducted ChIP-seq analysis of DNA-PKcs in MCF7 cells after 10 min of E2 treatment,

which revealed 12,629 peaks that mostly located in intergenic regions. Of the detected peaks, 971 were on the ERa-bound, MegaTrans-containing active enhancers but few were present on non-active enhancers. A second antibody for DNA-PKcs yielded similar ChIP-seq results, confirming the specificity of the signal.

In order to determine whether trans-bound RAR is required for the recruitment of DNA-PKcs at ERa active enhancers, we performed ChIPs for both DNA-PKcs and pERaS118 after knockdown of RARa/g. We found that RARa/g knockdown substantially reduced the levels of both DNA-PKcs and pE-RaS118 at ERa active enhancers, suggesting that trans-bound RARs may be required for the functionally relevant recruitment of DNA-PKcs at these ERa enhancers.

Knockdown of DNA-PKcs significantly inhibited E2-induced activation of ERa-bound functional enhancers and their target coding gene expression but did not affect RA-induced HoxA1 activation. Consistently, the treatment of MCF7 cells with the DNA-PK kinase inhibitor NU7441 also inhibited ERa-dependent target activation. Thus, at least one role of RARs that are recruited to ERa-bound functional enhancers may be to facilitate the concomitant recruitment of a specific protein kinase. It is possible that, analogous to this role of RARs in recruitment of DNA-PK, other DNA-binding TFs components in the MegaTrans complex also contribute to the recruitment of additional enzymatic factors that are required for functional enhancer activation.

Here, we suggest that, in addition to the critical recruitment of an ever-increasing number of well-characterized coactivator complexes, many with

specific enzymatic functions, activation of the most robust subset of ERa enhancers by E2 is dependent upon, and can be predicted by, their ability to recruit a complex of established DNA-binding transcription factors, referred to as the MegaTrans complex. This complex appears to be recruited/assembled in trans on ERa-bound functional enhancers and requires the presence of ERa. In addition to the requirement for ERa, certain other components of the complex appear to be necessary for its assembly on functional enhancers; for example, knockdown of RARa/g and GATA3 abolishes recruitment of other components of the complex and inhibits enhancer/target coding gene activation. Although the precise biochemical interactions that underlie the formation of the Mega-Trans complex remain incompletely defined, our data on the effects of DNase I treatment and DNA-binding domain mutation suggest that the MegaTrans complex assembles in situ at ERa-bound, ERE-containing enhancers, which also typically harbor nearby FoxA1 cis-binding sites.

While the idea that DNA-binding transcription factors can be recruited in trans to either activate or repress specific target coding genes is well established (Langlais et al., 2012; Pascual et al., 2005; Reichardt et al., 1998), this study provides an initial description of a ligand-dependent recruitment in trans of a complex of DNA-binding transcription factors that proves important for ERa function. Using the published criteria for defining super-enhancers (Hnisz et al., 2013; Whyte et al., 2013), only 22% of the functional MegaTrans enhancers can be classified as components of super-enhancers, and we note that there is only a very slight distinction in the levels of eRNA induction in response to E2 on the

functional MegaTrans enhancers associated with super-enhancers compared to those not associated with the super-enhancers. Thus, recruitment of the MegaTrans complex serves as a mark that distinguishes the most active enhancers of the estrogen-regulated transcriptional program.

These observations raise several corollary questions. First, does this MegaTrans complex serve on all active or activated enhancers, irrespective of the DNA-binding transcription factors bound in cis to those enhancers? It appears that the RARE-containing functional enhancers, which recruit RARa/g in cis, do not recruit this complex or ERa. Therefore, we speculate that there may be a number of distinct MegaTrans complexes that are recruited only by certain regulatory DNA-binding factors, and these complexes, analogous to events for ERa-regulated enhancers, serve to mark and initiate other specific enhancer activation events. Second, how is the MegaTrans complex selectively recruited only to the functional ERa-bound enhancers? Based on our initial data, we suggest that the answer likely involves the apparent dual roles of the ‘‘pioneer factor’’ FoxA1, which is selectively recruited to the functional, MegaTrans-dependent enhancers at < 200 bp from the ERE but is also required for the binding of ERa to these enhancers.

In addition to its established pioneering role, FoxA1 may also make an important contribution to the recruitment/stabilization of the MegaTrans complex. We are tempted to speculate that, in addition to promoting cooperative binding of ERa to enhancers, FoxA1 may cause a conformational alteration in the ERa receptor, either directly or via altered enhancer DNA architecture, that facilitates

the recruitment of the MegaTrans complex; however, it is formally possible that the increased affinity of ERa for the enhancer alone determines binding of the MegaTrans complex. In light of the already large number of important coactivator complexes, why would these additional DNA-binding transcription factors, most of which are recruited to the active enhancers by the ERE-bound ERa, be required? First, we have found that they play important ‘‘early’’ roles in enhancer function as they are important for eRNA induction and the ligand-dependent increase of p300 and Med1 occupancy on the enhancers. Thus, components of the MegaTrans complex are required to license the recruitment of well-known, important coactivators, as exemplified by p300 and Mediator subunits. In this regard, the DNA-binding transcription factors summoned to bind in trans through ERa are subserving functions that are quite analogous to those of the recognized coactivator complexes, many of which feature associated/intrinsic enzymatic activities.

Similarly, we note that RARs are capable of interacting with many known or potential coactivators, and we have focused on one such potential regulator. The enzyme DNA-PKcs binds to RARs and is recruited with rapid temporal kinetics to ERa-bound functional enhancers. Additionally, the knockdown of DNA-PKcs partially phenocopies the functional consequences of RARa/g knockdown in MCF7 cells. Therefore, we are tempted to speculate that components of the MegaTrans complex individually.

DNA-PK is a kinase with multiple targets, including ERa on Ser118 (Foulds et al., 2013), which we find occurs on the active ERa-bound enhancers,

dependent on the presence of RARs on these functional enhancers. It is particularly intriguing that DNA-PKcs is associated with the Ku80 complex, classically considered to be involved in DNA damage repair (Hartley et al., 1995; Jin and Weaver, 1997), which may in fact be pertinent to its functions in transcriptional control events. The rapid appearance of DNA-PKcs on the ligand-regulated enhancers is analogous to other examples of recruited protein kinases in gene regulation events (Perissi et al., 2008; Tee et al., 2014). Thus, investigation of the ERa-regulated enhancers has revealed that an additional and critical component of the most active enhancers is the ERa-dependent recruitment of the MegaTrans complex, which promotes combinatorial recruitment of additional coactivators/enzymes that increase enhancer activation and target coding gene transcription. Analogous to the hypothesis that super-enhancers regulate critical developmental or disease-associated coding gene transcriptional programs, MegaTrans complex recruitment appears to serve as a mechanism of marking/empowering enhancers to control key aspects of the regulatory transcriptional programs in a specific cell type. The super-enhancer model defines the combinatorial effects of multiple, clustered enhancers spanning > 8–10 kb, while the MegaTrans enhancer model explains the differential functional activity of single enhancers.

The uncovering of another layer of machinery involved in the effective activation of ERa-regulated enhancers raises the possibility that distinct MegaTrans enhancers exist for other classes of DNA-binding TFs that are responsible for activation of unique transcriptional programs. We note that ChIP-

seq analyses for many established DNA-binding TFs have revealed their binding on enhancers that do not harbor any known cognate binding sequences. This raises the possibility that these TFs might exert roles, in trans, on other transcription programs analogous to the effects of the MegaTrans complex on the ERa-regulated functional enhancers. The ‘‘hotspot’’ or ‘‘clustering’’ phenomenon of DNA-binding TFs has recently been reported in several different organisms (Junion et al., 2012; Rada-Iglesias et al., 2012; Siersbæk et al., 2014a; 2014b; Wilson et al., 2010; Yan et al., 2013). However, the underlying molecular mechanism(s) and functional significance are not well understood. Our results provide a functional model to explain at least many cases of the clustering phenomena. Specifically, our data suggest that the DNA-dependent binding of ERa and FoxA1 at ERa functional enhancers establishes a platform for recruiting a MegaTrans complex of other DNA-binding TFs by protein-protein interactions (in trans). MegaTrans complex-bound enhancers function as more robust enhancers by recruiting certain unique factors and enzymes, such as DNA-PK. Thus, our study provides insights into understanding the phenomenon of TF clustering. Our data also simultaneously help to explain why ChIP-seq analyses reveal 50% of the regions occupied by many of the DNA-binding TFs assayed in the ENCODE project do not harbor cognate DNA-binding motifs.

Materials and Methods.

Cell Culture and BLRP-Tagged Stable Cell Lines MCF7 cells, initially obtained from ATCC, were maintained in culture and treated as described (Li et

al., 2013). To study binding patterns for wild-type and non-DNA-binding mutants of RARa/g, GATA3, ERa, and other TFs, we first established a parental MCF7 stable cell line that expressed BirA enzyme and Tet-Repressor. We then used this parental cell line to make doxycycline- inducible stable cell lines expressing BLRP-tagged proteins at close to endogenous levels. BLRP-tagged proteins were biotinylated in vivo by BirA enzyme, allowing for pull downs to be performed with NanoLink streptavidin magnetic beads (Solulink) under very stringent washing conditions.

Chromatin Immunoprecipitation and Global Run-on Sequencing.

The qPCR-validated DNA samples were used to make libraries for deep sequencing.

Deep Sequencing.

For all ChIP-seqs and GRO-seqs, the extracted DNA libraries were sequenced with Illumina's HiSeq 2000 system according to the manufacturer's instructions. The first 48 bp for each sequence tag returned by the Illumina Pipeline were aligned to the human genome (hg18) assembly using Bowtie2. The uniquely mapped tags were selected for further ChIP-seq data analysis and a maximum of 3 tags per genomic position were collected for further GRO-seq data analysis. The data were visualized by preparing custom tracks on the University of California, Santa Cruz (UCSC) genome browser using HOMER software package (http://homer.salk. edu/homer/). The total number of mappable reads was

normalized to 10^7 for each experiment presented in this study.

Identification of ChIP-Seq Peaks, Heatmap, and Tag Density Analyses.

The identification of ChIP-seq peaks was performed using HOMER similar to our previously published methods (Li et al., 2013; Wang et al., 2011). Briefly, for transcription factors and histone marks, different parameters were utilized due to different patterns of tag distribution. For transcription factor/cofactor binding, high read enrichment regions were searched within a 200 bp sliding window. Regions of maximal density exceeding a given threshold were called as peaks. Only one tag from each unique position was considered to avoid clonal artifacts from sequencing. The tag threshold to call a valid peak was set by a cutoff with a false discovery rate (FDR) of 0.001, determined by peak finding using randomized tag positions in a genome with an effective size of $2.3*10^9$ bp.

We also required peaks to have at least 4-fold more tags (normalized to total count) than input control samples. In addition, the defined peaks must have at least four-fold more tags relative to the local background region (10 kb window) to avoid identifying regions with genomic duplications or non-localized binding. All called peaks meeting the criteria were then associated with genes by cross-referencing the RefSeq TSS database. Peaks from individual experiments were considered equivalent/overlapping if their peak centers were located within 200 bp. For all ChIP-seq analysis, the peaks within ± 400 bp of RefSeq gene TSSs were considered as promoter-bound peaks. Annotated positions for promoters, exons, introns and other features were based on RefSeq transcripts and repeat

annotations from the University of California, Santa Cruz. Heatmap matrices were created by counting tags using a 6 kb window (±3 kb of the peak center) and 100bp bin size. Heatmaps were later displayed using MultiExperiment Viewer (MeV). For regular enhancer analysis, tag density plots were created by counting tags using a 6 kb window (±3 kb of the peak center) and 10bp bin size and were plotted using Excel. For the super-enhancers, tag density plots were created by counting tags using a 10 kb window (±5 kb of the peak center) and 20 bp bin size and were plotted using Excel.

ChIP-Seq Data from Published Sources.

Published ChIP-seq data from the Gene Expression Omnibus database for FoxA1 under accession number GSE23852 (Tan et al., 2011), GATA3 under accession number GSE40129 (Theodorou et al., 2013), and c-Fos and c-Jun under accession number GSE26831 (Joseph et al., 2010) were used.

Motif Analysis.

Motif discovery was performed using a comparative algorithm as described in HOMER (Heinz et al., 2010). Motif finding for transcription factors was performed on sequence from ± 200 bp relative to the peak center. Briefly, sequences were divided into target and background sets for each application of the algorithm. Background sequences were then selectively weighted to equalize the distributions of G+C content in target and background sequences to avoid comparing sequences of different general sequence content. Motifs of length 8 to

12 were identified separately by first exhaustively screening all oligonucleotides for enrichment in the target set compared with the background set using the cumulative hypergeometric distribution to score enrichment. One mismatch was allowed in each oligonucleotide sequence to increase the sensitivity of the method. The top 200 oligonucleotides of each length with the lowest p values were then converted into probability matrices and heuristically optimized to maximize hypergeometric enrichment of each motif in the given data set. As optimized motifs were found, they were removed from the data set to facilitate the identification of additional motifs in subsequent rounds. HOMER also screens for the enrichment of known motifs previously identified through the analysis of published ChIP-ChIP and ChIP-seq data sets by calculating the known motifs' hypergeometric enrichment in the same set of G+C normalized sequences used for de novo analysis. Sequence logos were generated using WebLOGO (http://weblogo.berkeley.edu). Student's two-tailed t test was used to assess the significance of the ERE and FOXA1 motifs between the ERa active and non-active enhancers.

Bioinformatic Characterization of ERa Active and Non-active Enhancers.

ERa active enhancers and non-active enhancers were defined as in the previously published method (Li et al., 2013). The ERa and H3K27ac co-bound regions are defined as those in which the distance from the center of an ERa peak to the H3K27ac peak-occupied region is % 1 kb. Overall, two methods were used to assign the ERa-bound enhancers to E2-upregulated genes: (1) identifying the

E-upregulated coding genes from GRO-seq first and then coupling each of them to their closest ERa/H3K27ac co-bound enhancer within a certain distance (200 kb) (a 'gene-centric' view); and (2) characterizing the ERa/H3K27ac co-bound enhancers first and then coupling each of them to the closest TSS that belong to E2-upregulated coding genes (an 'enhancer-centric' view). eRNA levels within a 2kb window centered on ERa enhancers was computed for each enhancer. 1,333 enhancers were found to be with eRNA fold change (FC) > 1.5 and more than 8 tags/kb. They were considered to be the ERa most active enhancers. 1,110 ERa enhancers with very limited eRNA induction (FC < 1.2) were defined as ERa non-active enhancers. In addition, there were 519 ERa enhancers with no eRNA induction at all (FC < 1.1 and % 8tags/kb) upon E2 treatment. They were defined as the most non-active enhancers, and these were used for most analyses as control group. The ERa binding strength for these two enhancer groups was compared using boxplots of ERa tag numbers. The comparison of ChIP-seq tag intensity, GRO-seq transcription levels or distances between different categories are presented as boxplots. The p-values were determined by two-tailed Student's t tests.

Bioinformatic Characterization of RAR/GATA3 cis-and trans-Binding Sites.

ChIP-seq peaks for RARa/g and GATA3 were identified by HOMER as described above. Cis-binding and trans-binding sites for these three genes were identified by counting ChIP-seq tags (for wild-type and two DNA-binding mutants) within a 2kb window centered on each peak for the wild-type. A wild-

type/mutant ratio was calculated for each peak by dividing the ChIP-seq tags corresponding to wild-type and mutant. 3,000 potential trans-binding sites for RAR that showed < 1.2 ratio between the wild-type and mutant were identified. For the remaining RAR cis-binding sites, the tag density ratio between the wild-type and mutant was much greater than 1.2. Similarly, GATA3 had about 5,000 potential trans-binding sites that showed < 1.2 ratio between the wild-type and mutant.

RAR cis Active Enhancer Identification.

3,540 enhancers were identified from roughly 15,000 RAR cis sites using H3K27ac and H3K4me1 markers in MCF7. RAR cis enhancers were ranked according to eRNA level by counting GRO-seq tags using a 2 kb window centered on each RAR cis enhancer. 700 RAR cis-binding enhancers that showed > 1.5 FC with RA treatment in addition to > 8 tags/kb tag density were found, which were treated as RAR cis active enhancers.

Super-Enhancer Definition in MCF7.

To define super-enhancers, the pipeline from a previous publication (Whyte et al., 2013) was followed. Constituent enhancers that occurred within 12.5 kb into a single larger enhancer domain were captured and combined. The 12.5 kb stitching window is optimized and chosen by maximization of signal within each domain. To identify super-enhancers, all enhancers in MCF7 were ranked by increasing total background-subtracted ChIP-seq occupancy of Med1

and plotted the total ChIP-seq occupancy of Med1 in units of total rpm/bp (reads per 10 million per base pair). These plots revealed a clear point in the distribution of enhancers where the occupancy signal began increasing rapidly. To geometrically define this point, the data were scaled such that the x and y axis were from 0-1. Then the x axis point for which a line with a slope of 1 was tangent to the curve was found. The enhancers above this point to be super-enhancers and the enhancers below that point to be typical enhancers.

GRO-seq Analysis.

GRO-seq data analyses were performed as previously reported (Li et al., 2013; Wang et al., 2011). The sequencing reads were aligned to hg18 using Bowtie2 using very sensitive parameters. The common artifacts derived from clonal amplification were circumvented by considering maximal three tags from each unique genomic position as determined from the mapping data. To determine E2- or RA-dependent changes in gene body, the sequencing reads for RefSeq genes were counted over the first 60 kb (a conservative estimate considering 1 hr of E2 treatment may affect a limited distance of polymerase movement) of the entire gene body, excluding the 400 bp promoter-proximal region on the sense strand with respect to the gene orientation by using BED Tools or HOMER. To determine E2- or RA-dependent changes in eRNA expression, de novo transcript initiation sites were identified as regions where the GRO-seq read density increased threefold relative to the preceding 1 kb region. Transcript termination sites were defined by either a reduction in reads below 10% as compare to that of

TSS or when another transcript's start was identified on the same strand. Transcripts were defined as putative eRNAs if their de novo called start sites were located distal to RefSeq TSS (R3kb) and were asso- ciated with ERa or RAR and H3K27ac co-bound enhancer regions. EdgeR (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) was used to detect expression changes for either gene body or eRNA between the vehicle control and the E2 or RA treatment samples. P-values returned by EdgeR were adjusted for multiple hypotheses. Transcripts with FDR < 0.001 and fold change FC > 1.5 were selected for further analysis. The boxplots of ERa active enhancers, ERa non-active enhancers, and also a group of non-ERa enhancers were generated using software R by plotting FC for gene body or eRNA level obtained from EdgeR. The p-values were determined by two-tailed Student's t tests.

Chapter 4, in full, is a reprint of the material as it appears in Cell 2014. Enhancer activation requires trans-recruitment of a mega transcription factor complex. Liu Z, Merkurjev D, Yang F, Li W, Oh S, Friedman MJ, Song X, Zhang F, Ma Q, Ohgi KA, Krones A, Rosenfeld MG., 2014.

CHAPTER 5
lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs.

While recent studies indicated roles of long non-coding RNAs (lncRNAs) in physiologic aspects of cell-type determination and tissue homeostasis yet their potential involvement in regulated gene transcription programs remain rather poorly understood. Androgen receptor (AR) regulates a large repertoire of genes central to the identity and behavior of prostate cancer cells, and functions in a ligand-independent fashion in many prostate cancers when they become hormone refractory after initial androgen deprivation therapy. Here, we report that two lncRNAs highly overexpressed in aggressive prostate cancer, PRNCR1 and PCGEM1, bind successively to the AR and strongly enhance both ligand-dependent and ligand-independent AR-mediated gene activation programs and proliferation in prostate cancer cells. Binding of PRNCR1 to the C-terminally acetylated AR on enhancers and its association with DOT1L appear to be required for recruitment of the second lncRNA, PCGEM1, to the DOT1L-mediated methylated AR N-terminus. Unexpectedly, recognition of specific protein marks by PCGEM1-recruited Pygopus2 PHD domain proves to enhance selective looping of AR-bound enhancers to target gene promoters in these cells. In "resistant" prostate cancer cells, these overexpressed lncRNAs can interact with, and are required for, the robust activation of both truncated and full length AR, causing ligand-independent activation of the AR transcriptional program and cell proliferation. Conditionally-expressed short hairpin RNA (shRNA) targeting of these lncRNAs in castration-

resistant prostate cancer (CRPC) cell lines strongly suppressed tumor xenograft growth in vivo. Together, these results suggest that these overexpressed lncRNAs can potentially serve as a required component of castration-resistance in prostatic tumors.

One of the overexpressed lncRNAs in prostate cancer, PCGEM1, is tissue-specific and correlated with high-risk prostate cancer patients, including African-American men, while a second highly expressed lncRNA, PRNCR1(PCNCR1), is pervasively transcribed from the 8q24 "gene desert" region in strong association with susceptibility of prostate cancer. Paired benign prostatic hyperplasia (BPT) and aggressive tumor specimens (Gleason scores 3+3) derived from three individual prostate cancer patients exhibited > 100-fold upregulation of PRNCR1 andPCGEM1 expression. Native RNA-Immunoprecipitation (RIP) experiments in paired prostate tumor and BPH tissues (Gleason scores 2+3 to 4+3), revealed a specific association of full-length AR with both PRNCR1 and PCGEM1 in prostate tumor tissues compared to minimal interactions with glucocorticoid receptor (GR). In LNCaP cells, DHT treatment induced association of AR with both PRNCR1 and PCGEM1, but not withNEAT2. Antisense oligonucleotides (ASO)-based knockdown of PRNCR1 abolished both AR-PRNCR1 and AR-PCGEM1 interactions while knockdown of PCGEM1 did not affect the AR-PRNCR1interaction, suggesting the PRNCR1-dependent recruitment of PCGEM1. Knockdown of either PRNCR1 or PCGEM1 resulted in a significantly decrease in DHT-induced activation of canonical AR target genes while not affecting AR levels. Global

run-on sequencing (GRO-Seq) confirmed that knockdown of either PRNCR1 or PCGEM1 significantly decreased the induction of 617 DHT-upregulated genes (n=617, edgeR FDR < 0.01, and read density RD > 2) with AR-bound enhancers within 200 kb, but had no effect on DHT-unresponsive genes located > 200kb away from AR-bound enhancers, verified by randomly extracting sets of 1,000 genes.

Using Chromatin Isolation by RNA Purification (ChIRP) with biotin-labeled DNA probes (40-60 nt) tilingPRNCR1 and PCGEM1 RNAs, we identified 2,142 high-confidence PCGEM1occupancy sites genome-wide and motif analysis revealed a very significantly enriched AR response element (ARE) DNA motif, revealing that ~82% of PGCEM1 co-localized with AR-bound sites (−3 kb/+3 kb relative to the center of PCGEM1 peak), of which ~70% corresponded to AR bound, H3K4me1-marked enhancers, independently confirmed by qPCR analyses and ChIRP-Seq using even-numbered and odd-numbered probe sets and data not shown). These data suggest a stoichiometry of PCGEM1 sufficient to account for the recruitment to AR DNA regulatory binding sites on enhancers. Levels of PRNCR1 in LNCaP cells are estimated as ~400-600 copies/cell. The ability of these lncRNAs to read enhancERassociated histone marks might account for their preferential presence at AR-bound enhancers (vide infra).

By imposing a high stringency wash condition, we identified that DOT1L, CARM1, GADD45α, and AR specifically associated with in vitro-transcribed biotinylated PRNCR1 by Mass-spectrometry analysis, while AR, β-Catenin, and

Pygopus2 (Pygo2) preferentially associated with in vitro-transcribed biotinylated PCGEM1. β-Catenin, CARM1 and GADD45α have been suggested to play important roles in AR signaling. Knockdown of AR, Dot1L, β-catenin, and Pygo2 by specific siRNAs impaired DHT-induced activation of AR-target genes, TMPRSS2, PSA, and FKBP5. Mass-spectrometry analysis revealed that the lncRNA-bound AR contains several post-translational modifications, including K631/634 acetylation and K349 methylation. Consistent with the proposed importance of acetylation of AR in activation of an AR target gene, a K631/634R mutation on AR inhibited its interaction with PRNCR1 and PCGEM1 and DHT-induced expression of AR target genes, while overexpression of a AR K631/634Q mutant resulted in enhanced DHT-dependent interactions with PRNCR1 and PCGEM. These data suggest that PRNCR1 and PCGEM1 interact with AR in a K631/634 acetylation- and K349 methylation-dependent manner, respectively.

Because effective AR-PCGEM1 interaction requires the methylation of AR at K349, we confirmed DOT1L-mediated AR methylation at K349 using in vitro methylation assays, finding that a K349R point mutation significantly inhibited AR methylation. Specific DOT1L knockdown impaired the interaction between AR and PCGEM1, but not that with PRNCR1, suggesting that AR K349 methylation, mediated by PRNCR1-bound DOT1L, is critical for the recruitment of PCGEM1 to AR. Indeed, overexpression of an AR K349R mutant significantly reduced DHT-induced gene activation in LNCaP cells.

In vitro binding studies demonstrated that PRNCR1 bound to the region aa

549-623 of AR and PCGEM1 bound to AR N terminal region when methylated at K349 by overexpressing DOT1L. By incubating in vitro transcribed PCGEM1 with nuclear lysate from cells overexpressing Myc-tagged Pygo2 proteins, including full-length, N- or C-terminally truncated proteins, we identified strong interactions between PCGEM1 and the Pygo2 C-terminus.

To map the sequence motif of PCGEM1 responsible for AR or Pygo2 binding, we performed modified in vitroRNA pulldown followed by dot-blot assay, using two regions of NEAT2-bound by unmethylated Pc2 as a control for the CLIP assay. Methylated AR bound/protected PCGEM1 sequence was identified to encompass 421GAT…TCC480 with unmethylated AR or the unrelated protein, His-tagged MURF1, not showing specific binding to any region of PCGEM1. A sequence motif of PCGEM1 encompassing 1201TGT…ATT1260, distinct from the AR binding region, was identified as the Pygo2 binding motif, with deletion of this motif (Δ1191-1270) abolishing Pygo2 binding with no effect on AR binding. Similarly, deletion of AR binding site of PCGEM1 (Δ411-490) abolished AR-PCGEM1 interaction, with minimal effect on Pygo2-PCGEM1interaction.

MODified Histone Peptide Array experiments using in vitro transcribed biotinylated PCGEM1 or PRNCR1revealed that they selectively recognize H3K4me1 and H4K16ac histone marks indicative of enhancers, respectively. Therefore, it is likely that these histone tail associations of PRNCR1 andPCGEM1 serve as a functional component of their preferred

recruitment to enhancers of AR-regulated transcription units. PCGEM1 and PRNCR1 were highly upregulated in the LNCaP-cds2, and CWR22Rv1 castration-resistant prostate cancer cell line models compared with immortalized "normal" prostate epithelial cell lines, WPE and RWPE, or even LNCaP cells. The AR antagonist, bicalutamide (Casodex), reduced the DHT-induced PSA expression in LNCaP cells but failed to act as an antagonist in LNCaP-cds2 cells. Transduction of LNCaP-cds2 cells with lentivirus expressing shRNAs against PRNCR1or PCGEM1, but not a non-specific (LacZ) shRNA, significantly reduced the activation of several canonical AR target genes while having no effect on AR expression levels. Because truncated forms of AR that exhibit ligand-independent transcriptional activation are frequently detected in castration-resistant prostate cancer and may often reflect alterations in AR gene structure, we investigated the potential roles of PRNCR1 and PCGEM1 in AR-mediated basal transcription activity in CRPC cells. RT-PCR using primers specific for one AR "splicing" variant, AR-V712 confirmed the presence of this variant in LNCaP-cds2 cells. Western blot analysis using N-terminal AR-specific antibody (441), revealed that the AR-V7 variant (~75kDa) represents ~1-2% of total AR in LNCaP-cds2 cells, although it is the predominant form in CWR22Rv1 cells. Because overexpression of truncated AR can constitutively activate androgen-responsive genes in the absence of ligand, we transfected LNCaP cells with AR Q641X mutant, with activation of canonical androgen-responsive genes, including TMPRSS2, PSA, KLK2,FKBP5, and NKX3-1 in the absence of added androgen. This constitutive effect of AR Q641X was highly reduced upon

knockdown of either PRNCR1 or PCGEM1. RIP assay in CWR22Rv1 cells demonstrated that both PRNCR1 and PCGEM1 associated constitutively with truncated AR (AR-V7) with or without ligand. By immunoblotting the AR-V7 immunoprecipitates with N-terminal AR-specific antibody (441), we did not observe any residual pull-down indicative of interaction between full length and truncated AR, arguing against an indirect association of PRNCR1 and PCGEM1 with truncated AR consequent to hetrodimerization with full length AR. Using an antibody specific for the C-terminal ligand binding domain of AR (C-19) to selectively recognize full length AR, we observed interactions between these lncRNAs and full length AR in the absence of added ligand, possibly due to the relative higher level of basal acetylation and methylation of full-length AR in CWR22Rv1 cells. To study the biological roles of PRNCR1 or PCGEM1, we generated stable cell lines derived from CWR22Rv1 harboring Doxycycline (Dox)-induced shRNA against LacZ, PRNCR1 or PCGEM1. Dox-induced either PCGEM1 or PRNCR1 knockdown significantly reduced the expression of canonical AR target genes, with no noticeable effect on AR expression level. Dox-induced knockdown of either PRNCR1 or PCGEM1 also inhibited the growth of CWR22Rv1 cells comparable to the effect of AR knockdown, without affecting AR expression levels. Remarkably, conditional shRNA-mediated inhibition of either PRNCR1 or PCGEM1 robustly inhibited in vivotumor growth in a CWR22Rv1 prostate cancer xenograft mice model, indicative of an lncRNA-dependent regulatory network that critically regulates growth of castration-

resistant prostate cancer cells in vivo.

While knockdown of either PRNCR1 or PCGEM1 did not affect the recruitment of AR on PSA and KLK2enhancers, knockdown of PCGEM1 inhibited SMC1 recruitment onPSA and KLK2 promoters, with only minimal effects on SMC1 levels on enhancers, consistent with proposed Cohesin-dependent formation of chromatin loops between enhancers and promoters. We further demonstrated ligand-induced enhancer: promoter interactions in the PSA transcription unit by ChIP-3C assay and found that these interactions were impaired by depletion of either PRNCR1 orPCGEM1.

The ability of Pygo2, associated with PCGEM1, to recognize a canonical promoter histone mark16, H3K4me3, raised the possibility that Pygo2 might be involved, at least quantitatively, in enhancer: promoter looping. ChIP assays revealed that Pygo2 was efficiently recruited to enhancer and promoter regions of the PSA, KLK2 and TMPRSS2 transcription units in a DHT-dependent manner, exhibiting relatively higher and earlier association with the enhancer regions. Knockdown of either PRNCR1 or PCGEM1 in LNCaP cells inhibited Pygo2 recruitment to AR-dependent enhancers/promoters. On knockdown of Pygo2, AR and SMC1 failed to effectively associate with the PSA, KLK2 or TMPRSS2 promoters despite unperturbed DHT-dependent AR or SMC1 recruitment to their enhancers and DHT-induced enhancer: promoter looping in the PSA transcription unit was inhibited. GRO-Seq analysis revealed a broad inhibition of AR-dependent transcriptional program under the condition of Pygo2 knockdown (n=290, edgeR FDR < 0.01), which did not affect AR expression. Depleting

PRNCR1 or PCGEM1 from LNCaP-cds2 cells also inhibited enhancer: promoter looping in FASN and NDRG1, previously shown to be activated following FoxA1 knockdown in LNCaP cells17. Again, knockdown of either PRNCR1 or PCGEM1 had no effect on recruitment of AR to enhancer regions of the FASN and NDRG1 transcription units or on Pygo2 expression levels.

To address whether the Pygo2 PHD domain might itself be instrumental for its function in mediating chromatin looping, we first depleted Pygo2 by shRNA followed by overexpression in LNCaP-cds2 cells of either shRNA-resistant wild-type Pygo218 or a W352A mutant defective for H3K4me3 recognition16. In 3C assays, knockdown of Pygo2 reduced FASN enhancer: promoter interactions, which could be rescued by overexpression of wild-type, but not W352A, Pygo2, even though there was equal recruitment of wild-type Pygo2 or the W352A mutant to enhancers, and no altered promoter H3K4me3 levels. Knockdown of Pygo2 curtailed the expression of canonical of AR target genes TMPRSS2, KLK2, PSA, FKBP5, and NKX3-1, and overexpression of wild-type, but not W352A, Pygo2 was able to robustly rescue the induction of these genes. These data suggest that Pygo2 exerts a quantitatively-important role in DHT-dependent enhancer: promoter interactions and coding target gene activation. For 220 AR-regulated coding gene promoters under regulation of an enhancer exhibiting ligand-dependent recruitment of Pygo2 also exhibiting recruitment to the coding gene promoter by ChIP-seq, we did not observe ligand-induced increase in the next adjacent, non-AR-regulated transcription unit (~204 promoters), indicating that the H3K4me3 mark cannot alone be sufficient to

effectively recruit Pygo2, suggesting a role of other similarly-modified proteins in prostate cancer cell gene activation events.

In the present study, we have found a mechanistic link between prostate cancer-upregulated lncRNAs and AR transcriptional activity, revealing the biological importance of the lncRNAs, PRNCR1 and PCGEM1, in licensing C-terminally truncated, as well as full length, AR-dependent gene activation events in prostate cancer cells. Considering the regulatory potential of enhancer RNAs (eRNAs) identified in recent studies lncRNAs may also be part of a broad transcription regulatory network.

Materials and Methods.

Tissue Samples and Processing.

Experiments using paired benign prostatic hyperplasia (BPT) and tumor (T) were performed from unidentified individual prostate cancer patients, which were obtained from Dr. Ralph W. deVere White, UC Davis Comprehensive Cancer Center. The informed consent was obtained from all subjects. The tissue samples used in manuscript were received as the de-identified samples without any PHI attached. The Gleason score or tumor/BPH status was considered pathological information, not patient information. We did not know the names or birth dates. The tissues were homogenized by Precellys®24 tissue homogenizer followed by downstream assays.

Cell Culture and Transfection.

Prostate cancer LNCaP cells were obtained from ATCC and cultured in RPMI1640 containing 10% (vol/vol) FBS. The benign immortalized prostate cell line RWPE, WPE and the castration-resistant prostate cancer cell lines LNCaP-cds1, LNCaP-cds2, CWR22Rv1 were kindly provided by Dr. Christopher Evans (Department of Urology, University of California Davis). RWPE and WEP cells were grown in KGM media and bulletkit from Lonza supplemented with L-Glutamine. CWR22Rv1, LNCaP-cds1 and LNCaP-cds2 cells were grown in RPMI1640 media containing 5% final volume of Charcoal Stripped Serum. LNCaP cells were grown to 30-50% confluence and siRNA/ASO transfections were carried out using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Transfection of LNCaP cells with DNA plasmids was performed using Amaxa™ Nucleofector™ kit R from Lonza. shRNAs specific for LacZ, PRNCR1, PCGEM1 or AR were delivered, by lentiviral transduction, to LNCaP-cds2 and CWR22Rv1 cells.

Cloning Procedures.

The full-length AR expression vector has been previously described17. Human PCGEM1 (14-1556) and PRNCR1 fragments (1-3240, 3156-6428, 6331-9670 and 9531-12710) were amplified from cDNAs generated from LNCaP cells and cloned into pSTBlue-1 vector (Novagen) for in vitro transcription assay. PCGEM1 gene sequence with 411-490 and 1191-1270 deletion were synthesized by GeneScript Inc. and cloned into pcDNA3.1 vector (Invitrogen). Lentiviral vector pLKO.1 containing the shRNA against scrambled sequence, Pygo2 and

pHIV vector containing RNAi-resistant Pygo2 cDNA were obtained from Dr. Xing Dai (Department of Biological Chemistry, University of California at Irvine). A 4.8 kb genomic sequence upstream of PSA promoter was amplified from LNCaP cells and subcloned into pSTBlue-1 vector (Novagen). Bacterial expression vectors for AR and K349R mutant were constructed by subcloning the gene sequences into pET-28a backbone (Novagen). Nuclear expression vectors for AR2-920, AR2-548, AR549-920, AR2-623, AR549-623, AR624-666, AR667-920 and Pygo22-406, Pygo22-266, Pygo2267-406 were constructed by subcloning the cDNA sequences into pCMV/myc/nuc backbone (Invitrogen). The expression vector of FLAG-DOT1L was kindly provided by Dr. Yi Zhang's laboratory. All mutants were generated using QuikChange™ Lightning Site-Directed Mutagenesis Kit (Agilent Technologies).

Antisense DNA Oligonucleotide, siRNA and Lentiviral shRNA.

2′-O-methyl phosphor-othioate oligonucleotides were designed and synthesized by Integrated DNA Technologies, Inc. Commercially available FlexiTube siRNA targeting AR (SI02757258) and CARM1 (SI02663815) from Qiagen, ON-TARGETplus SMARTpool siRNA targeting DOT1L (L-014900-01-0005) and GADD45α (L-003893-00-0005) from Dharmacon, MISSION® siRNA targeting β-Catenin (SASI_Hs01_00117958), Pygopus2 (SASI_Hs01_00059018) from Sigma-Aldrich were used in this study. The knockdown efficiency and specificity of all siRNAs were validated either by vendors or ourselves. Oligonucleotides for shRNA targeting PRNCR1, PCGEM1 and AR were

designed at http://biosettia.com/support/shrna-designer and cloned into pLV-H1TetO-GFP-Puro vector according to manufacturer's instructions (Biosettia). We tested the efficacy and specificity of 3 and 2 shRNAs in LNCaP and LNCa-cds2 cells, respectively, for both PRNCR1 and PCGEM1. For functional assays, the specific ASO/shRNA giving the best knockdown efficiency was used.

Antibodies.

Specific antibodies were purchased from the following commercial sources: anti-AR (N-20), anti-AR (C-19), anti-AR (441), anti-GR (E-20), anti-β-Catenin (D-10), and anti-GAPDH (6C5) from Santa Cruz Biotechnology; anti-CARM1 (4438), anti-GADD45α (3518) and anti-pan acetylated-lysine (#9441) from Cell Signaling Technology; anti-FLAG® M2, anti-γ-Tubulin (T5326) and anti-Pygo2 from Sigma-Aldrich Prestige Antibodies®; anti-Pygo2 (GTX119726) from GeneTex Inc.; Anti-DOT1L (39954) and anti-Myc tag (clone 4E12) from Active Motif; anti-SMC1 (A300-055A) from Bethyl Laboratories Inc.; anti-pan methylated lysine (7315) from Abcam and anti-AR-V7 from Precision Antibody™.

Protein Recombination and Purification.

Recombinant His-AR proteins were expressed in E.coli strain BL21-CodonPlus® (DE3)-RIPL (Agilent Technologies) and purified using TALON® Metal Affinity Resin (Clontech). Recombinant Histone H3 was purchased from Active Motif. Human DOT1L (amino acids 2-416) was obtained from Sigma-

Aldrich. Human Pygo2 was purchased from BioClone Inc. His-tagged MURF1 was purchased from BostonBiochem®.

RNA Immunoprecipitation (RIP).

RIP was performed in native conditions as described. Briefly, 1×107 LNCaP cell nuclei were pelleted and lysed in 1mL ice-cold Polysomal Lysis Buffer (100mM KCl, 5mM MgCl2, 10mM HEPES [pH 7.0], 0.5% NP-40, 1mM DTT) supplemented with Anti-RNase, Protease Inhibitor Cocktail, Phosphatase Inhibitor Cocktail, Panobinostat, and Methylstat. The lysate were passed through a 27.5 gauge needle 4 times to promote nuclear lysis. Turbo™ DNase (400 U) was then added to the lysate and incubated on ice for 30 min. The cell lysate was diluted in the NT2 buffer (50mM Tris-HCl [pH 7.4], 150mM NaCl, 1mM MgCl2, 0.05% NP-40) and 50 ul of the supernatant was saved as input for PCR analysis. 500 ul of the supernatant was incubated with 4 μg of AR (441) antibodies-IgG magnetic beads (pre-blocked by 1×PBS+5 mg/ml BSA) at 4 °C overnight. The RNA/antibody complex was washed four times (1 ml wash, 5 minutes each) by NT2 buffer supplemented with Anti-RNase, Protease Inhibitor Cocktail, Phosphatase Inhibitor Cocktail, Panobinostat and Methylstat. The RNA was extracted using Acid-Phenol: Chloroform, pH 4.5 (with IAA, 125:24:1) (Invitrogen) according to the manufacturer's protocol and subjected to RT-qPCR analysis.

Chromatin Isolation by RNA Purification (ChIRP).

ChIRP was performed as described with minor modifications. Briefly, 60-mer antisense DNA probes targeting PRNCR1 and PCGEM1 full-length sequences were designed at http://www.singlemoleculefish.com/designer.html. A set of probes targeting LacZ RNA was also designed as the negative control. All probes were biotinylated using Label I® Nucleic Acid Biotin Labeling Kit from Mirus Biotechnology. LNCaP cells were fixed with 1% formaldehyde for 10 min at room temperature. Crosslinking was then quenched with 125 mM glycine for 5 min. The chromatin preparation, hybridization/elution, deep sequencing steps were essentially performed as described except that wash was conducted at 50 °C and 65 °C. The image analysis and base calling were performed using Illumina's computational analysis pipeline. The sequencing reads were aligned to hg18 human genome by using Bowtie2 and only 1 read/position was kept for downstream analyses. Peak finding was performed by using HOMER suite and the peaks within 1 kb from each other were merged. Peak intersection was computed by using intersectBed in BedTools, after extending the peaks with 1 kb. In order to call reliable peaks, we have excluded from analysis the peaks that overlapped the satellite repeats or LacZ ChIRP peaks. The annotation of the peaks on the human genome and the tag density profiles were computed in HOMER, and the display of the heatmap were carried in MeV. Sequenced motif enrichment analysis was carried in HOMER. For ChIP-Seq data (AR and H3K4me1), peak finding was performed by using HOMER or SICER.

Data Processing Procedure for Overlapping ChIP-Seq and ChIRP-Seq.

We used the standard ChIP-Seq peak finding tools in the processing of ChIRP-Seq data. We chose HOMER software suite (http://biowhat.ucsd.edu) for the analysis of both AR ChIP-Seq and PCGEM1 ChIRP-Seq data, using the same program routine (i.e. findPeaks command) and the same default parameters that calls only the robust peaks (these parameters are outlined below): 1) fold enrichment over input tag count, default: 4.0; 2) poisson p-value threshold relative to input tag count, default: 0.0001; 3) fold enrichment over local tag count, default: 4.0; 4) poisson p-value threshold relative to local tag count, default: 0.0001; 5) fdr=false discovery rate, default = 0.001; 6) size of region used for local filtering = 10000; 7) fold over local region required = 4.00. Peak finding procedure: 1) typically, the tag distribution along the genome could be modeled by a Poisson distribution and the peak finding algorithm slides windows of fixed size across the genome to find candidate peaks with a significant tag enrichment (Poisson distribution p-value default 10-4 to 10-5); 2) we use very strict parameters in HOMER to call a peak: a very low FDR (i.e. 0.001), and a high fold enrichment over input tag count (i.e. 4). By default, HOMER also requires the tag density at peaks to be 4-fold greater than in the surrounding 10 kb region. Therefore, we ensure that only the sharp peaks with low local background are called and considered for downstream analyses. Both ChIP-seq and ChIRP-seq data were processed precisely in the same way using the same default settings of HOMER.

For the analysis of overlapping between ChIP-Seq and ChIRP-Seq data, we used the following samples: PCGEM1ChIRP-Seq (−DHT), PCGEM1 ChIRP-

Seq (+DHT), and AR ChIP-Seq (at a higher sequencing depth). The heatmaps were generated in 2 steps: 1) we used HOMER and the command [(annotatePeaks.pl <peak file> <genome> −size <#> −hist <#> −ghist) that specifies the list of peaks (<peak file>), and the tag density in a region that covers <size> kb around the peak center; the tag density is specified in a bin size that is specified by the parameter <hist>)] to generate a matrix of tag densities across the samples, for each peak. Typically the tag densities are normalized to 10 mil reads for each sample; 2) we used MeV package (http://www.tm4.org/mev/) to display the heatmap and to color code it on a scale from 0 to 2: a difference in the colors from 2 (red) to 0 (green) may reflect a fold enrichment over local tag count higher than 4 (4 is typically the default value, used for peak finding, as we described above).

GRO-Seq.

The image analysis and base calling were performed by using Illumina's standard computational analysis pipeline. Bowtie2 was used to align the sequencing reads to hg18 human genome; when multiple reads aligned to the same genomic position, only 1 read/position was kept for downstream analyses. We have used RefSeq annotations in order to estimate the gene expression level by counting the sequencing reads over the gene body (from 400 bp downstream of TSS to TTS) on the sense strand with respect to the gene transcription, by using intersectBed in BedTools. In order to call statistically significant differentially expressed genes, we have used edgeR and a FDR < 0.01. Additional

criteria, including read density (normalized number of reads/kb) were imposed in order to filter the differentially expressed genes. Wilcoxon tests for sample comparisons were computed in R.

Chromatin immunoprecipitation (ChIP) assay, ChIP-Seq and ChIP-3C assay.

ChIP assays and ChIP-Seq was performed as described. ChIP-3C assays were performed as described with minor modifications. Briefly, LNCaP cells were cross-linked with 1% formaldehyde for 10 min followed by incubation with 125 mM glycine for 5 min. The cross-linked chromatin was sonicated and digested with restriction enzyme BstYI (New England Lab) followed by immunoprecipitated with specific antibody coupled to Protein A Dynabeads® (Invitrogen). The beads bound chromatin was ligated with T4 DNA ligase (Promega), eluted and reverse-cross-linked. The ChIP-3C material was purified by QIAquick Gel Extraction Kit (Qiagen) and subjected to PCR analysis with sets of primers as described.

3C Assay.

LNCaP-cds2 cells were cross-linked with 1% formaldehyde for 10 min at room temperature and processed according to the procedures as described with minor modification. Briefly, BAC clones (Empire Genomics) for FASN (RP11-1033I, RP11-1033J9) and NDRG1 (RP11-671M3, RP11-125I19) loci were used to generate control templates for the positive controls. 15 μg of the BAC clone for each locus were mixed and digested with 300 units of corresponding restriction

enzyme overnight. DNA fragments were extracted and ligated with T4 DNA ligase at a DNA concentration of 300 ng/μl. The primers for the fragments on the FASN and NDRG1 loci were used as previously described.

In Vitro RNA pulldown Coupled with Dot-Blot Assay.

In vitro transcribed biotinylated RNA was formed secondary structure as described and incubated with recombinant His-tagged or GST-tagged proteins in gel shift protein binding buffer (Promega) on ice for 1 hr. The reactions were ultraviolet (UV) irradiated (150 mJ/cm2) to crosslink RNA-protein complexes. After UV irradiation, the RNA was partially digested by RNase I (Ambion) at 1:50 and 1: 500 dilutions for 5 min, allowing a small fragment to remain attached to protein. RNA-protein complexes of interest were then partially purified by His tag or GST tag magnetic beads and the purified RNA-protein complexes were treated with proteinase K, which removes protein but leaves intact RNAs. The recovered RNAs were hybridized to BrightStar®-Plus Positively Charged Nylon Membrane spotted with 60-mer antisense DNA oligonucleotides tiling along the indicated lncRNA sequence at 37 °C overnight. The anti-sense oligonucleotides corresponding to the indicated lncRNAs were spotted on membrane as following order (left to right in each row): A1 is corresponding to the oligonucleotide sequence nt 1-60 of indicated lncRNA, A2 is corresponding to oligonucleotide sequence nt 61-120 of indicated lncRNA and so on till the end of the lncRNA sequences. The hybridized membrane was washed as described at sequential 37 °C, 50 °C and 65 °C. The protein-bound RNA sequence was visualized by

detection of Streptavidin-HRP signals. The positions and sequences of antisense DNA oligonucleotides tiling along PCGEM1 is available upon request.

In Vitro Methylation Assay.

In vitro methylation assays were conducted in a total volume of 30 μl using 2 μg of substrate, and 1 μg of recombinant GST-DOT1L in methylation buffer (50 mM HEPES [pH 8.0], 0.01% (v/v) NP-40, 10 mM NaCl, 1 mM DTT and 1 mM PMSF) supplemented with either 40 μM S-Adenosyl-methionine (SAM, Sigma-Aldrich; for nonradioactive methylation) or 300 nM S-Adenosyl-L-[methyl-[3H]methionine ([3H]-SAM, Perkin Elmer; 0.55 μCi per reaction for radioactive methylation). Reactions were carried out at 37 °C for 3 hours and the reactions were separation on a 4-12% SDS-PAGE gel. The resulting protein bands were visualized by Coomassie blue staining, immunoblotting or autoradiography using EN3HANCE™ spray (Pelkin Elmer). Purification of methylated protein was carried out as described.

Cell Lysis, Protein Immunoprecipitation, and Immunoblotting.

Cells were homogenized in 1×RIPA buffer with protease inhibitor and Anti-RNase (Invitrogen). Lysates were cleared by centrifugation at 13,000 rpm for 15 min at 4 °C. The protein concentration of the extracts was determined by Bradford assay (Bio-Rad). Immunoprecipitation experiments and immunoblotting analyses were performed as described in previous studies. The blotting signals were detected using SuperSignal West Dura Extended Duration Substrate

(Pierce).

Lentivirus Packaging and Transduction.

Production of Lentiviral shRNA stocks were performed in 293LTV cell line according to manufacturer's instruction (Cell Biolabs). The lentivirus was further purified and concentrated by ViraBind™ Lentivirus Concentration and Purification Kit (Cell Biolabs). LNCaP-cds2 and CWR22Rv1 cells were transduced using ViraDuctin™ Lentivirus Transduction Kit (Cell Biolabs) and purified with 0.6 and 0.4 μg/ml puromycin, respectively. To establish stable cell line expressing shRNA against LacZ, PRNCR1, PCGEM1 or AR, CWR22Rv1 cells were transduced as described above and stable transduced clones were generated following selection with 0.4 μg/ml puromycin. Single cell clones were then isolated by three rounds of single cell dilution, and the efficiency of Doxycycline-induced target gene knockdown was screened by RT- qPCR with 24 individual clones for each target gene.

RNA Pulldown Assay and Mass Spectrometric Analysis.

RNA pulldown assay was performed as previously described with minor modifications. Biotin-labeled PCGEM1 RNA (14-1556) and PRNCR1 RNA fragments (1-3240, 3156-6428, 6331-9670 and 9531-12710) were in vitro transcribed with the Biotin RNA Labeling Mix (Roche) and T7 or SP6 RNA polymerase (Promega) and purified with RNeasy® Mini Kit (QIAGEN). RNAs were incubated with nuclear extract from LNCaP cells in the presence of anti-

RNase, Protease Inhibitor Cocktail, Phosphatase Inhibitor Cocktail, Panobinostat, and Methylstat. Proteins pulled down by biotinylated PRNCR1 and PCGEM1 were first eluted, denatured, reduced, alkylated and then digested with immobilized trypsin (Promega). The digested peptides were batch purified on a reversed-phase ZipTip® (Millipore), and resulting peptide pools were then subjected to mass spectrometric analysis at UCSD Biomolecular and Proteomics Mass Spectrometry Facility. Data were analyzed using ProteinPilot™ software (Applied Biosystems).

Modified Histone Peptide Array.

The interaction between in vitro transcribed PRNCR1 and PCGEM1 with Histone was analyzed by MODified™ Histone Peptide Array (Active Motif) according to manufacturer's instruction. The specificity of interactions was quantified by Array. Analyses Software (Active Motif).

Histone Peptide Pulldown Assay.

Histone peptide pulldown assays were performed using SensoLyte® Methylated Histone H3 kit from AnaSpec. Briefly, 1 μg of recombinant Pygo2 protein was incubated with 1μg of biotinylated histone H3 peptides (unmodified, K4me1, K4me2, or K4me3) in the presence of yeast tRNA or in vitro transcribed PCGEM1 in binding buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% NP-40) for 2 hrs at 4 °C with rotation. After 1 hr incubation with streptavidin magnetic beads and extensive washing, the bound protein was analyzed by SDS-

gel electrophoresis and western blotting with anti-Pygo2 antibodies. For this assay, PCGEM1 was in vitro transcribed using unlabeled UTP.

Cell Proliferation Assay.

Cell proliferation assay was performed using CellTiter 96® AQueous One Solution Cell Proliferation Assay (MTS) (Promega). Briefly, cells were distributed in 96-well plate with 100 μl media each. After removing media and rinsed once by PBS, cells were supplied with 100 μl PBS mixed by 4 μl MTS reagent, followed by 1 hr incubation at 37 °C 5% CO2 incubator. After incubation, 490nm absorption of each well was measured by light absorbance reader. Every group has 6 replicates.

Xenografts and Animals.

CWR22Rv1 stable cell lines with Doxycycline-inducible shRNA against LacZ, PRNCR1 or PCGEM1 were injected into male athymic Nu/Nu mice, 4-5 weeks old. Mice arrived in our facility were randomly put into cages with four mice each. They were implanted with respective tumor cells in the unit of cages. Three million of cells were suspended with matrigel in the ratio of 2:1 for subcutaneous injection. The experiment was set up to use eight animals per group to detect a 2-fold increase in tumor size with power of 80% and at the significance level of 0.05 by a two-sided test for significant studies. In these experiments, the tumor take rate was often about 50%, so we often ended up with about 4 animals per group, with or without doxycycline treatment. Tumor

progression was monitored by caliper measurement twice a week and the tumor volume was calculated according to the equation, v=length × width2 ×1/2. When the tumor size reached between 150 – 200 mm3, animals were randomly distributed to groups receiving 1 mg/ml of doxycycline water or regular water with continuous tumor monitor until tumor burden exceeded the limit of IACUC humane endpoints (less than 20 mm in one dimension), for 3 to 4 weeks. The investigators were blinded to the group allocation during the experiment and/or when assessing the outcome. This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health.

RNA isolation and qRT–PCR.

Total RNA was isolated from cells using RNeasy Plus Mini Kit (Qiagen) following the manufacturer's protocol. First-strand cDNA synthesis from total RNA was carried out using iScript™ cDNA Synthesis Kit (Bio-Rad). Resulting cDNA was then analyzed by qPCR using SsoAdvanced™ SYBR® Green Supermix (Bio-Rad) on Stratagene Mx3000 or CFX Connect™ Real-Time PCR Detection System. All qPCRs were repeated at least three times.

Chapter 5, in full, is a reprint of the material as it appears in Nature 2013. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, Merkurjev D, Ohgi KA, Meng D, Zhang J, Evans CP, Rosenfeld MG., 2013.

CHAPTER 6
The Rett Syndrome protein MeCP2 establishes nuclear matrix-associated boundaries regulating transcription.

Mutations in MeCP2 are the predominant cause of the Rett syndrome neurological disorder, but it remains unclear how loss of MeCP2 function leads to the causative, aberrant neuronal gene expression programs. Here, we report that MeCP2 is required for the establishment of a specific subset (~20%) of the ~2000 chromosomal topologically associating domain (TAD) boundaries and that loss of MeCP2 induces a disruption of nuclear architecture by destabilizing this subset of TAD boundaries dependent on binding of MeCP2, with CTCF, and their association with the nuclear matrix network. Unexpectedly, the vast majority of the > 1500 neuronal coding genes dysregulated in Rett syndrome models prove to be located in the chromosomal territories adjacent to these affected boundaries. These results reveal a general mechanism by which MeCP2 enforces matrix-dependent boundary transcriptional control in territories adjacent to specific boundaries and serves as a gatekeeper to prevent aberrant, pathological transcriptional programs. Rett syndrome therefore apparently provides an initial example of a chromosomal boundary disease.

Rett syndrome is caused by mutations of the methyl CG binding protein, MeCP2 . Because Rett syndrome is potentially a reversible disorder , it is clearly important to accelerate the rate of discovery to achieve predicted new therapies, however, in order to assess such therapies we must understand the basic functions of MeCP2, and how their deregulation causes disease. Loss of MeCP2 function is a salient feature of MeCP2 mutations in Rett syndrome patients and is most likely

required for both neuronal and glial function. Since MeCP2 is a transcriptional regulator, aberrant gene regulation is thought to play a major role in the disease. However, how MeCP2 actually regulates gene transcription still remains under intense debate. Most previous work has focused on the mechanisms by which MeCP2 regulates transcription directly by recruitment of cofactors to DNA (eg. HDACs, Sin3a, NCoR, CREB etc.), while other work has also hinted that MeCP2 could function as a regulator of chromatin structure and nuclear architecture, especially as MeCP2 is abundant in mature neurons and is suggested to function as a linker "histone" . For example, it can condense chromatin in vitro , and may be required for maintaining specific chromatin looping structures , required for the proper localization of alpha thalassemia/mental retardation syndrome X-linked (ATRX) protein in the organization of pericentric heterochromatin. Recent work has suggested that the three dimensional organization of the nucleus is critical for proper regulation of gene transcription . Specifically, using methods such as HiC and 5C, it has been found that chromosomes are composed of discrete topologically associating domains (TADs) , which are ~1 Mb regions of high local contact frequencies, conserved between cell types and even between organisms, and may define transcriptional co-regulation of genes within the same TAD . TADs are defined by boundary elements that are enriched in CTCF; however, the mechanisms that establish TAD boundaries are still unclear. Because TAD boundaries define the borders and limit interactions between neighboring TADs, it is suggested that boundaries limit the potential of inter-TAD enhancers to regulate genes within a defined TAD, and thus help define the

transcriptional potential of TADs . Here, we find that loss of MeCP2 induces a disruption of nuclear architecture, by destabilizing a specific subset of ~20% of all TAD boundaries and their association with the nuclear matrix network, leading to aberrant transcription of > 1500 neuronal coding genes in boundary associated domains (BADs) adjacent to these affected boundaries. Thus, subTAD chromosomal territories within several hundred kb from the MeCP2-enriched boundaries, due, in part, to reduced binding of the structural protein CTCF and reduced association with the nuclear matrix in these boundaries, specifically affects a Rett-associated MeCP2-dependent transcriptional program. Because MeCP2 may have roles in structural organization of chromatin, we hypothesized that some of the transcriptional changes induced by MeCP2 loss might be due to changes in chromatin architecture. To test this, we performed tethered conformation capture (TCC), a variant of HiC with improved signal to noise ratio , under control or siRNA knockdown of MeCP2 conditions, and performed paired-end deep sequencing of the interactions obtained. We chose to initially use neuronal cell culture models of mouse Neuro2a neuroblastoma cells in order to take advantage of the homogenous cell population and thus circumvent the heterogeneity of mixed neuronal cell types. We produced a total of 215 million paired-end aligned reads and computed interactions between loci with a window of 10 kb. From the filtered interaction data, we generated interaction matrices for each chromosome. The large-scale structural organization of chromosomes is unchanged, where the predominant interactions identified within each chromosome are preserved. In order to directly examine whether loss of

MeCP2 can affect TAD or boundary integrity, we defined TADs and the boundaries between them using the directionality index as previously described . While the TAD and boundary organization is essentially preserved for the most part, closer examination reveals striking effects of MeCP2 loss on structural organization of the genome. For example, knockdown of MeCP2 induced a significantly location shift in ~20% of boundaries (366/1666) Approximately a third of the TADs that are bordered by shifted boundaries have shifted boundaries on both sides, while the rest exhibit shifting of only one of their boundaries. This "shifting" is consistent with the definition of boundaries by directional bias in the interactions at the periphery of TADs towards the center of the TAD. Indeed altered interactions can be observed when plotting the most robust interactions using Circos plots and matrix plots. More specifically, interactions occurring across boundaries (interTAD) are seen to be increased specifically across shifted boundaries, while interaction within TADs (intraTAD) are slightly decreased, following knock down of MeCP2. When plotting the distance of paired interTAD and intraTAD interaction, we observed that most gained interTAD interactions were short-range interactions (< 400 kb), suggesting they occurred relatively close to the boundary. Within the intraTAD interactions, most interactions were indeed slightly reduced following MeCP2 knockdown, irrespective of the distance of the interaction. However, we noticed a slight increase in intraTAD short-range interactions (20-100 kb). These short-range interactions seemed to be distributed across the TAD, and therefore distinct from the effects on the boundary. However, since it is known that TADs are most likely divided into

"subTADs" which are enriched in CTCF sites , we tested whether these increased interaction occurred close to CTCF sites. Indeed, they were much closer to CTCF sites than the long-range interactions which suggests that increased interactions occurred across "subTAD boundaries". Indeed, within TADs, we observed more interactions crossing MeCP2/CTCF co-bound sites. We next investigated how loss of MeCP2 leads to altered TAD boundaries, and whether MeCP2 was specifically localized to these areas in our system. First, we took advantage of a powerful biotin tagging strategy for genome-wide MeCP2 localization (ChIP-seq). These cells express biotin tagged MeCP2 at near endogenous levels and exhibit the typical heterochromatic localization stain pattern typical of MeCP2 and allowed for specific MeCP2 location in the genome. Analysis of MeCP2 distribution identified 68,880 peaks and revealed that MeCP2 is localized predominantly to intergenic and intronic regions, as well as being enriched over exons. This is consistent with recent findings showing that MeCP2 binding on gene bodies is due to 5 hmc binding in neurons . Importantly, we found that, indeed, MeCP2 was particularly enriched on "shifted" TAD boundaries, suggesting a potential unique function at these locations.

Because nuclear organization has been shown to be critical for proper regulation of gene transcription , we tested how TAD boundary disorganization affected gene transcription. We performed global run-on sequencing (GRO-seq) in siRNA control or siMeCP2 conditions, and identified > 2000 nascent RNA transcripts regulated by more than 1.5-fold (reads per kilobase per million reads, RPKM > 0.5, P-value < 0.05). Gene ontology analysis of MeCP2 regulated genes

identified GO terms highly enriched in neuronal function, suggesting that binding of MeCP2 may be crucial in regulating multiple genes critical for neuronal function. To determine whether the effects of MeCP2-dependent alterations of boundaries might be correlated with the transcriptional changes, we compared the location of up- and down- regulated genes in MeCP2 knock down neuronal cells, and asked whether each gene is closer to a shifted or non-shifted boundary. We found that the coding genes exhibiting altered expression were, remarkably, located near shifted boundaries for > 90% of all up regulated coding genes (1,645 vs. 118 genes near shifted vs. non-shifted boundaries, respectively) relative to the non-regulated gene distribution which occur as expected at approximately 20% of the shifted boundaries. Down-regulated genes were also relatively enriched, but to a lesser extent, with 54% located near shifted boundaries (530 and 437 genes near shifted and non-shifted boundaries, respectively). Importantly, the genes that were deregulated after MeCP2 knockdown were much closer to boundaries (median ~300 kb), relative to unchanged genes (median ~600 kb) predominantly for altered genes that were in TADs with "shifted" boundaries, whereas the much smaller number of changed genes adjacent to non-shifted boundaries did not show this trend. Furthermore, when comparing the number of genes on or spanning boundaries that are changed after MeCP2 knockdown relative to the proportional size of boundaries in the genome (10%), we found that boundaries in general had a relatively much higher abundance of changed genes relative to the random or actual gene distribution overlapping the boundaries. These data suggest that boundaries that are altered ("shifted") following loss of MeCP2 tend to be

associated with the vast majority of deregulated coding gene targets and that the misregulated coding transcription units in these TADs are much closer to than those adjacent to non-altered boundaries.

Because many of the MeCP2-dependent changed genes, which are close to boundaries, were enriched in gene ontology terms related to neuronal function, this suggests these genes could be of functional significance in the pathogenesis of Rett syndrome. Further informatic analysis of readily available published transcriptome data sets from Rett syndrome mouse hippocampus , cerebellum and striatum also show that more than 70% of all changed genes are associated with "shifted" rather than non-"shifted" boundaries as defined in our current datasets, and that they are significantly closer to boundaries relative to non-changed genes. These results strongly suggest that the predominant mechanism of MeCP2-dependent transcriptional deregulation in Rett syndrome is through destabilization of specific TAD boundaries, as > 80% of these reported up-regulated coding genes in the murine Rett model are adjacent to the very same "shifted" boundaries that we have identified in the Neuro2A neuronal cell model. Taken together, our data suggest that loss of MeCP2 destabilizes TAD boundary function, leading to altered transcription of important neuronal function-related genes in a defined ~250 kb region adjacent to the boundary, which we term as Boundary Associated Domains (BADs), thus suggesting that MeCP2 is preferentially required for the maintenance of BADs.

In order to gain insight into the possible mechanism of action of MeCP2, we performed motif analysis of MeCP2 bound peaks. The highest motif enriched

was found to be CTCF, a well established regulator of nuclear architecture and function . ChIP-seq of CTCF in Neuro2a cells revealed that approximately 60% of all CTCF peaks were co-bound with MeCP2. Moreover, knock down of MeCP2 and subsequent CTCF ChIP revealed that loss of MeCP2 caused concomitant loss of CTCF binding on interrogated CTCF binding sites, and this was not due to effects on CTCF levels as CTCF mRNA levels remained unchanged. In parallel, ChIP on hippocampal tissue from MeCP2 knockout mice confirmed that loss of MeCP2 caused a marked reduction in CTCF binding on select sites. This effect was also confirmed genome-wide by performing CTCF ChIP-seq after MeCP2 knockdown, revealing that a large fraction of CTCF peaks were reduced or lost. Importantly, CTCF peak number, enriched on boundaries, was diminished following MeCP2 knockdown, predominantly on the “shifted” boundaries. Moreover, CTCF and MeCP2 can directly interact, suggesting that MeCP2 may play a role in stabilizing a subset of CTCF binding events. In this regard, we also found that the effects of MeCP2-dependent CTCF peak loss are predominantly on boundaries that exhibit gene expression changes after MeCP2 knockdown. This suggests that loss of CTCF binding, at least in part, serves as a contributor to the observed altered interactions and transcriptional changes in adjacent coding target genes. MeCP2 is known to bind methylated DNA , and it has been shown that CTCF binding may be regulated by the methylation status of its cognate binding site . We therefore assayed whether any correlation was observed between DNA methylation and CTCF binding using the MethylMiner assay and we found no clear correlation between loss of CTCF binding and

methylation. Moreover, CTCF could be clearly bound on areas with high levels of methylation, consistent with observations that CTCF binding is sensitive to methylation of only select nucleotides within its binding motif, which would not be selectively detected using the MethylMiner method . In addition, analysis of the motifs found at CTCF peaks lost after MeCP2 knock down detected primarily the core canonical CTCF motifs in about 50-60% of the sites with additional up and down-stream motifs for only a small subset of the lost peaks, suggesting that CTCF is recruited to its binding site directly for the majority of cases (not shown). Taken together, these results demonstrate that MeCP2 is required for the stabilization of a significant portion of CTCF binding events, especially at the subset of TAD boundaries that are most affected by loss of MeCP2. While it remains unclear precisely how TAD structure is established, enhancers, which engage in long-range enhancer®Cpromoter and enhancer-enhancer interactions for transcriptional regulation , may constitute the foundations for TAD structure , as well as help to define the transcriptional potential of TADs . Our location data found a significant overlap of MeCP2 with enhancers, as characterized by the H3K4me1 mark as well as H3K4me2 and H3K27ac marks, suggesting that by binding enhancers and boundaries, MeCP2 enforces TAD structure. GRO-seq data allowed us to identify active enhancers based on their H3K4me1 mark and the presence of eRNAs. We found that the eRNA levels on enhancers proximal to MeCP2 regulated genes were higher than on enhancers of non-regulated genes, and that their eRNA level were induced upon MeCP2 knockdown predominantly near shifted boundaries. These results suggest BADs contain genes with active

enhancers that may play a role in the altered interactions observed upon MeCP2 loss. Recent data from our laboratory identified a functional role for the nuclear matrin3-network, a distinct organizer of nuclear structure and function, in the activation of homeodomain-bound regulatory enhancers underlying developmental regulation of gene transcription in the pituitary gland , providing an instructive example of the importance of specific location in a sub-nuclear architectural structure for effective enhancer activation events. MeCP2 is partially homologous to a matrix attachment binding protein ARBP , has been shown to interact with ATRX , a protein tightly associated with the nuclear matrix , required for spreading of X-inactivation . MeCP2 is required for ATRX recruitment to imprinted loci and to regulate its nuclear localization . Therefore we examined whether MeCP2 deficiency might lead to defects in nuclear matrix integrity and/or patterns of interactions with specific chromosomal territories. Integrity of the nuclear matrix was assessed by both biochemical fractionation and by sequencing methods. Biochemical purification of the matrix fraction revealed that approximately a third of MeCP2 is associated with the nuclear matrix, and that CTCF, a known component of the nuclear matrix , is specifically lost from this fraction upon MeCP2 deficiency, mirroring the loss of CTCF observed by ChIP-seq following MeCP2 knockdown. Moreover, we extracted the matrix-associated DNA, and performed DNA sequencing, an approach designed to begin to determine the integrity of DNA sequences associated with the nuclear matrix upon MeCP2 knockdown, and define matrix associated DNA (MAD) sequences. MAD sequences occurred as distinct peaks highly enriched in

repetitive sequences and we observed a significant loss of MAD association upon MeCP2 knockdown. In addition, CTCF tag density was much higher on MADs with MeCP2 and CTCF co-binding, as opposed to that of CTCF co-bound with MeCP2 on non-MADs sequences. These matrix-binding sites also exhibited a greater loss of CTCF binding upon MeCP2 knockdown than other sites, suggesting a preferred effect of MeCP2 loss on MADs. Moreover, MeCP2 regulated genes, which, as we previously showed, are most commonly adjacent to "shifted" boundaries, were much closer to these CTCF/MeCP2/MAD binding sites relative to non-regulated genes. Because of the discovered MeCP2 role on "shifted" boundaries, we explored the connection between matrix attachment in "shifted" vs. "non-shifted" TAD boundaries. TADs exhibiting "shifted" boundaries had much higher matrix DNA association adjacent to the boundary regions, which was preferentially lost upon MeCP2 knockdown, relative to TADs with "non-shifted" boundaries. Moreover, we preformed matrix fractionation and DNA sequencing analysis in murine cortex from three independent MeCP2 knockout and three wild type littermate controls and revealed a virtually identical pattern of matrix association with shifted boundaries, an association lost in the MeCP2 knockout mice. In concert with this observation, heat map analysis revealed strikingly similar patterns of loss of matrix association in Neuro2A cells and murine cortex. Together these data suggest the intriguing possibility that MeCP2-dependent matrix attachment plays a role in boundary function, which is in concert with previous suggestions that CTCF can be a component of nuclear matrix. In concert, our results indicate that a major effect of loss of MeCP2 is the

location-specific disruption of nuclear architecture, destabilizing a cohort of TAD boundaries harboring the highest levels of MeCP2, based on loss of CTCF and change in nuclear matrix association of these 20% of most affected boundary regions. Importantly, > 80% of all coding transcription units altered upon knockdown of MeCP2 in the Neuro2a cell model proved to be close to the shifted boundaries with a median ~300 kb. Surprisingly, our model was sufficient to predict that the same boundaries were affected in vivo in Rett syndrome model mice since ~75% of genes changed were close to MeCP2-dependend boundaries. Indeed the concept that the MeCP2-bound shifted boundaries regulate the adjacent TAD territories could explain why despite in multiple efforts, a consistent MeCP2 target gene list has not been reported, because it implies that aberrant but stochastic interactions would occur in the absence of functional MeCP2 at boundaries, leading to altered neuronal gene programs, that could nonetheless still be in a territory proximal to these boundaries (BADs) in individuals with Rett syndrome. The transcriptional effects due to alterations in these specific boundaries in response to loss of MeCP2 suggest two non-exclusive mechanisms. The first is that the positive effects of enhancer interactions from the adjacent TAD, normally blocked by a MeCP2-dependant boundary, now license an increase in transcription of the affected coding genes adjacent to the boundary. Alternatively, it is equally plausible that the boundary itself normally inhibits expression via interactions with regulatory enhancers adjacent to the boundary, an inhibitory effect diminished upon loss of MeCP2 binding. Either or both mechanisms may lead to improper gene transcription via influence of aberrant

enhancer interactions both within TADs and between neighboring TADs. These results do not exclude to possibility that MeCP2 also has local effect on gene transcription, by interacting with repressive machinery such as histone deactylases, co-repressors and co-activators, suggesting that MeCP2 loss of function phenotype may be contributed by multiple mechanisms.

Our MeCP2 ChIP-seq data reveals that while MeCP2 is broadly distributed as has been suggested before, it does exhibit preferential enrichment on some areas such as gene bodies and enhancers. Moreover, we find MeCP2 to be enriched on boundaries that become "shifted" upon loss of MeCP2. Indeed, examination of data obtained in Purkinje cells reveals that shifted boundaries appear to exhibit enhanced 5-hmC levels relative to least-shifted boundaries, possibly explaining MeCP2 enrichment on shifted boundaries.

Our work suggests that at MeCP2 function on boundaries is dependent, at least in part, through it's effect on CTCF binding. Indeed, CTCF is a well known structural protein responsible for higher order chromatin structure that is known have barrier functions and is enriched in TAD; thus, we hypothesize that the observed MeCP2-dependent loss of CTCF at boundaries could partially explain the changes in nuclear architecture and loss of boundary function following MeCP2 knock down. Moreover, recent work has suggested that CTCF alone is only partially sufficient for maintaining TAD boundaries and that loss of CTCF does not lead to complete TAD disarray, but to a rather modest gain in interTAD interactions, suggesting increased aberrant interactions occurring across boundaries , suggesting that MeCP2 is likely to have additional roles in

maintaining TAD boundaries. Therefore it is not surprising that the effects of MeCP2 knockdown are not identical to CTCF knockdown , since MeCP2 knockdown only abrogates a subset of CTCF binding events, while CTCF probably has additional, broader roles in genome organization. However, our results in MeCP2 knockdown suggest that the observed 20% of shifted boundaries represent a subset that are more susceptible to alteration, as defined by the number of interactions crossing these boundaries in the basal state. This is line with work in showing that boundary strength can be variable and is dependent on the cohort of bound factors at each boundary . This suggests that non-“shifted” boundaries employ different mechanisms to maintain boundary function that are not dependent on MeCP2 function, and have yet to be defined. Moreover, our results fit well with recent work that has shown that MeCP2 is required for proper chromatin looping interactions, as well as the binding of CTCF and ATRX on select imprinted gene loci , suggesting that while MeCP2 may have a broad distribution in the genome , it specific recruitment to boundaries and other specific regions may have a predominant effect on the transcriptional alterations observed in Rett syndrome. While our data suggests a direct role for MeCP2 in CTCF binding, it is possible that loss of CTCF occurs via effects on ATRX recruitment, since it has been suggested that MeCP2-dependent ATRX binding repositions nucleosomes to allow for CTCF binding on select imprinted loci . It is also interesting to note that CTCF conditional knockout in post-mitotic projections neurons showed phenotypes stingingly similar to Rett syndrome MeCP2 knockout mice, including postnatal growth retardation and lethality,

defects in dendritic arborization and neuronal spine density, as well as behavioral defects and abnormal hind limb clasping . These results in concert with our data shown here suggest that perhaps some of Rett syndrome phenotypes are mediated through MeCP2-dependent effects on CTCF at the shifted boundaries.

Importantly, our results suggest a novel role for the nuclear matrix in boundary function. The importance of the nuclear matrix in functional organization of chromatin into higher-order loops and “matrix attached” elements has been proposed in 1970s but has been highly debated ever since . Recent data from our laboratory identified a functional role for the nuclear matrix components in the activation of homeodomain-specific transcriptional program , providing an example of the importance of specific interactions in a sub-nuclear space for proper gene regulation. Moreover, our data suggested that the association of sub-nuclear structures such as the matrin-3-enriched nuclear network/nuclear matrix with regulatory DNA elements is dynamic and dependent on cell-type specific transcription factors. Here we found that the MeCP2 enriched boundaries specifically interact with Matrix Attached Domains (MADs) and CTCF and that this interaction is required for proper regulation of genes within the adjacent BADs. Our work not only links the MeCP2 as a regulatory component mediating this specific interaction, but also sheds new light on the role of internal nuclear network in regulation of gene expression by influencing patterns of genomic interactions. Finally, our findings suggest that multiple genes will be deregulated either stochastically, or depending on the specific TAD and boundary architecture of each cell type, supporting the notion that multiple pathways may be affected in

Rett syndrome, perhaps dependent on the cell type and its specific boundary/TAD interactions. It is also interesting to speculate that the reduced nuclear volume observed in Rett brains and model animals may be due to a loss of structural stability that would be conferred by MeCP2-boundary-matrix interactions. Therefore, these results suggest that potential therapies that can address the structural defects induced by loss of MeCP2 may have potential benefit for Rett syndrome therapy in the future.

Materials and Methods.

ChIP-seq.

ChIP was performed as previously described . Briefly, for ChIP-seq, approximately 10^7 cells were cross-linked with 1% formaldehyde at room temperature for 10 min and neutralized with 0.125 M glycine. After sonication, 20®C30 μg of soluble chromatin was incubated with 1®C5 μg of antibody at 4 °C overnight. Immunoprecipitated complexes were collected using G-protein Dynabeads (Invitrogen). Subsequently, immuno-complexes were washed, and DNA was extracted and purified by QIAquick Spin columns (Qiagen). For biotinilated MeCP2 Chip-seq, sonicated chromatin was incubated with BSA-blocked streptavidin T1 Dynabeads (Invitrogen) overnight at 4° C with rotation. The captured biotin®Cstreptavidin complexes were washed, and samples were equilibrated in TEV buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.1% NP-40 (CA630), 0.5 mM EDTA) for 5 min at room temperature (25 °C), followed by AcTEV protease (Invitrogen) digestion (5®C10 U) for 1

hr at room temperature in 40 μl TEV buffer. The streptavidin-conjugated beads were eluted again with TEV buffer with 10 min incubation at room temperature. Eluted samples were reverse cross-linked and RNaseA- and proteinase-K-treated following a standard ChIP protocol. DNA libraries were sequenced on Illumina's GAII or HiSeq2000 platforms according to the manufacturer's instructions. The sequences returned by the Illumina Pipeline were aligned to the mouse assembly mm8 by using Bowtie or Bowtie2. The extracted DNA was ligated to specific adaptors followed by deep sequencing with the Illumina's HiSeq 2000 system according to the manufacturer's instructions. The first 48bp for each sequence tag returned by the Illumina Pipeline were aligned to the mouse genome (mm8) assembly using Bowtie2. The uniquely mapped tags were selected for further ChIP-seq data analysis and a maximum of 3 tags per genomic position were collected for further data analysis. The data were visualized by preparing custom tracks on the University of California, Santa Cruz (UCSC) genome browser using HOMER software package (http://homer.salk.edu/homer/). The total number of mappable reads was normalized to 10^7 for each experiment presented in this study.

GRO-seq.

GRO-seq experiments were performed as previously reported.

MethylMiner assay.

MethylMiner¢, Methylated DNA Enrichment Kit was obtained from

Invitrogen (Carlsbad, CA) and performed according to manufacturer instructions.

Tethered conformation Capture (TCC) .

Tethered conformation capture was essentially performed as described , with minor modifications. Briefly, for each experiment, 25 x 10^6 cells, fixed for 15 minutes with 1% formaldehyde/PBS and washed twice with PBS were thawed, permeabilized for 15 minutes on ice in 550 µl lysis buffer (10 mM HEPES pH=8.0, 10 mM NaCl, 0.2% IGEPAL CA-630, and 1X protease inhibitors solution (Roche) and pelleted at 2500x g, room temperature for 5 minutes. Cells were washed twice with 2 ml ice-cold wash buffer (50 mM Tris HCl pH=8.0, 50 mM NaCl, 1 mM EDTA) and resuspended in 500 µl wash buffer. Nuclei were isolated by adding 190 µl 2% SDS and incubating the cells for 10 minutes at 65°C and collected by centrifugation for 1 minute at 1000 x g. Nuclei were resuspended in 245 µl wash buffer and cysteine-biotinylated with 105 µl 25 mM EZlink Iodoacetyl-PEG2-Biotin (Pierce) dissolved in water, and rocking at RT for 75'. Samples were diluted with 650 µl 1x NEB2, incubated on ice, then 150 µl 10% Triton X-100 were added, followed by incubation on ice for 5', then at 37°C for an additional 10'. Chromatin was digested overnight at 37 °C after adding 85 µl 10x NEB2, 3 µl 1 M DTT, 227 µl water and 35 µl 25 U/µl MboI (NEB). The sample was dialyzed against 1 l TE pH 8.0 for 3 h at RT, then another 1 h with fresh 1 l TE, in a G2 Slide-A-Lyzer cassette with a (Pierce). In the meantime, 400 µl T1 Streptavidin Dynabeads (Invitrogen) were washed 3x with 2 ml 0.01% Tween 20/PBS (0.01TPBS) each and resuspended in 1.5 ml 0.01% Tween

20/PBS. The sample was divided into 5 aliquots in 1.5 ml Eppies, 300 µl Dynabeads each were added and the protein-DNA complexed collected by rocking at RT. Beads were blocked by adding 5 µl of 20 mM biotin solution (15x molar excess over streptavidin). Beads were washed once with 600 µl each of 0.01% Tween 20/PBS and STRP wash buffer (10 mM Tris pH8.0, 50 mM NaCl, 0.4% Triton X-100). Beads were resuspended in 100 µl of STRP wash buffer, and overhangs were filled in by adding 63 µL water, 1 µl 1 M $MgCl_2$, 10 µl of 10X NEBuffer2, 0.7 µl of 10 mM dATP, 0.7 µl of 10 mM dTTP, 0.7 µl of 10 mM dGTPαS (AXXORA), 15 µl 0.4 mM Biotin-14-dCTP (Invitrogen), 4 µl 10% Triton X-100, and 5 µl of 5 U/µl Klenow enzyme (Enzymatics) and rotating for 40 minutes at room temperature. The reaction was stopped with 5 µl 0.5 M EDTA, beads were washed twice with 600 µl each Klenow wash buffer (50 mM Tris pH 7.5, 0.4% Triton X-100, 0.1 mM EDTA). Beads were suspended in 500 µl Klenow wash buffer, transferred to a 15 ml conical tube, and DNA was ligated under rotation for 4 hours at 16° C in a total volume of 4 ml containing 250 µl 10x T4 DNA ligase buffer (Enzymatics), 180 µl 10 % Triton X-100, 100 µl 1 M Tris pH 7.5, 50 µl 100x BSA (10 mg/ml), 2 µl (1200 U) T4 DNA ligase (Enzymatics). The reaction was terminated by adding 200 µl 0.5 M EDTA. Beads were resuspended in 400 µl extraction buffer (50 mM Tris pH 8.0, 0.2% SDS, 1 mM EDTA, 100 mM NaCl), 20 µl 20 mg/ml proteinase K (Ambion) were added and crosslinks reversed overnight at 65 °C. DNA was extacted once with 400 µl phenol/chloroform/isoamylalcohol (25:24:1) µl tris-buffered to pH 8.0 and once with 400 µl CHCl3, and precipitated overnight at -20 °C with 20 µl 5 M NaCl, 1

µl 20 mg/ml glycogen and 1060 µl 100% EtOH. DNA was pelleted for 25 minutes at 20000x g, 4 °C and washed once with 1 ml 80% EtOH for 5 minutes, 8000x g, 4° C. Pellets were dissolved in 25 µl 10 mM Tris pH 8.0 each, aliquots were pooled and digested for 30 minutes at 37 °C with 1 µl 10 µg/µl RNase A. Reactions were cleaned up on a QiaQuick PCR purification column (Qiagen) and eluted with 50 µl elution buffer. Five micrograms DNA were treated with 300 U exonuclease III in 90 µl 1x NEBuffer 1 for 1 hour at 37 °C, then the enzyme was inactivated with 2 µl 0.5 M EDTA, 2 µl 5 M NaCl and heating to 70 °C for 20 minutes.

Water was added to 100 µl and DNA was sheared to 100-500 bp in a Covaris E220 in a 6x16 AFA fiber microtube at 5 % duty cycle, intensity 5 (= 175 W), 200 cycles/burst for 180" total time. DNA was isolated using a QiaQuick PCR purification column (Qiagen) and eluted with 50 µl EB. The DNA fragment ends were polished with 29 µl water, 10 µl 10x T4 DNA ligase buffer (Enzymatics), 4 µl 10 mM dNTP, and 5 µl (15 U) T4 DNA polymerase, 1 µl (5 U) Klenow, 5 µl (50 U) T4 PNK for 30'at 20 °C. After DNA cleanup on QiaQuick columns and elution into 50 µl EB, fragments were A-tailed with 5.4 µl 10x NEBuffer 2, 0.11 µl 100 mM dATP and 3 µl (15 U) exo$^-$-Klenow enzyme (Enzymatics) for 40' at 37 °C. The reaction was stopped with 1.5 µl 0.5 M EDTA and DNA was captured with 56 µl 2x B&W buffer containing 0.2 % Tween 20 and 15 µl T1 Dynabeads (Invitrogen, washed twice with 1x B&W buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, then suspended in 56 µl 2x B&W buffer), rotating for 30 minutes at room temperature. Beads were washed once

with 500 µl each of 1x B&W/0.1 % Triton-X100, once with TE. Sequencing adapters were ligated to the bead-bound DNA in 100 µl 1x rapid ligation buffer (Enzymatics) containing 0.1% Tween 20, 0.5 µl NextFlex adapters (Bioo, 1:20 diluted), 5 µl (3000 U) T4 DNA ligase (Enzymatics) for 20 minutes at room temperature. The reaction was stopped with 6 µl 0.5 M EDTA, beads washed twice with 1x B&W, twice with 0.1% Tween 20/TE, then resuspended in 40 µl 0.033% Tween20/LoTE, (TE diluted 1:4 with water). Libraries were PCR-amplified using the 20 µl of the bead suspension as template for 10 cycles, size-selected to 225-425 bp insert size on a 2% agarose gel, and paired-end sequenced on an Illumina HiSeq 2000.

Identification of ChIP-Seq Peaks, Heatmap, and Tag Density Analyses.

The identification of ChIP-seq peaks was performed using HOMER similar to previously published methods . Briefly, for transcription factors and histone marks, different parameters were utilized due to different patterns of tag distribution. For MeCP2, we have combined three ChIP-seq samples of biotin-MeCP2, or BirA control, that were sequenced in a single-end or paired-end mode, and aligned to the mouse mm8 genome assembly by using Bowtie. We have employed HOMER to call the MeCP2 peaks by using either a statistical background based on Poisson distribution. For transcription factor/cofactor binding, we searched for high read enrichment regions within a 200bp sliding window. Regions of maximal density exceeding a given threshold were called as peaks. Only one tag from each unique position was considered to avoid clonal

artifacts from sequencing. The tag threshold to call a valid peak was set by a cutoff with a false discovery rate (FDR) of 0.001, determined by peak finding using randomized tag positions in a genome with an effective size of 2×10^9 bp. We also required peaks to have at least 4-fold more tags (normalized to total count) than input control samples. In addition, the defined peaks must have at least four-fold more tags relative to the local background region (10 kb window) to avoid identifying regions with genomic duplications or non-localized binding. All called peaks meeting the criteria were then associated with genes by cross-referencing the RefSeq TSS database. Peaks from individual experiments were considered equivalent/overlapping if their peak centers were located within 200 bp. For all ChIP-seq analysis, the peaks within ± 400 bp of RefSeq gene TSSs were considered as promoter-bound peaks. Annotated positions for promoters, exons, introns and other features were based on RefSeq transcripts and repeat annotations from the University of California, Santa Cruz. Heatmap matrices were created by counting tags using a 6 kb window (±3kb of the peak center) and 100 bp bin size. Heatmaps were later displayed using MultiExperiment Viewer (MeV). For regular enhancer analysis, tag density plots were created by counting tags using a 6 kb window (±3 kb of the peak center) and 10 bp bin size and were plotted using Excel.

Motif Analysis.

Motif discovery was performed using a comparative algorithm as described in HOMER . Motif finding for transcription factors was performed on

sequence from ± 200 bp relative to the peak center. Briefly, sequences were divided into target and background sets for each application of the algorithm. Background sequences were then selectively weighted to equalize the distributions of G+C content in target and background sequences to avoid comparing sequences of different general sequence content. Motifs of length 8 to 12 were identified separately by first exhaustively screening all oligonucleotides for enrichment in the target set compared with the background set using the cumulative hypergeometric distribution to score enrichment. One mismatch was allowed in each oligonucleotide sequence to increase the sensitivity of the method. The top 200 oligonucleotides of each length with the lowest p values were then converted into probability matrices and heuristically optimized to maximize hypergeometric enrichment of each motif in the given data set. As optimized motifs were found, they were removed from the data set to facilitate the identification of additional motifs in subsequent rounds. HOMER also screens for the enrichment of known motifs previously identified through the analysis of published ChIP-ChIP and ChIP-seq data sets by calculating the known motifs' hypergeometric enrichment in the same set of G+C normalized sequences used for de novo analysis. Sequence logos were generated using WebLOGO (http://weblogo.berkeley.edu).

Bioinformatic Characterization of CTCF and MeCP2 Enhancers.

We followed our previously published method to define enhancers and non-active enhancers . The CTCF and MeCP2 enhancer regions are defined as

those in which the distance from the center of a CTCF or MeCP2 peak to the H3K4me1 peak-occupied region is °‹ 1 kb.

GRO-Seq Analysis.

GRO-Seq data analyses were performed as previously reported . The sequencing reads were aligned to mouse genome (mm8) assembly using Bowtie2 using very sensitive parameters. The common artifacts derived from clonal amplification were circumvented by considering maximal three tags from each unique genomic position as determined from the mapping data. To determine siMeCP2 or KCL- dependent changes in gene body, the sequencing reads for RefSeq genes were counted over the first 60 kb of the entire gene body, excluding the 400 bp promoter-proximal region on the sense strand with respect to the gene orientation by using BEDTools or HOMER. Transcript termination sites were defined by either a reduction in reads below 10% as compare to that of TSS or when another transcript's start was identified on the same strand. Transcripts were defined as putative eRNAs if their de novo called start sites were located distal to RefSeq TSS (°›3 kb). We used EdgeR (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) to detect expression changes for either gene body or eRNA between the control and the MeCP2 knockdown treatment samples. P-values returned by EdgeR were adjusted for multiple hypotheses. Transcripts with FDR °‹ 0.001 and fold change (FC) °› 1.5 were selected for further analysis. The p-values were determined by two-tailed Student's t tests.

TCC Analysis.

HOMER (http://homer.salk.edu/homer/interactions/) was used to analyze Tethered Conformation Capture (TCC) data. Alignment of TCC reads was done using Bowtie2 (standard parameters and unique genomic positions) to the mouse genome (mm8) assembly. Read-pairs were processed independently. Once Hi-C reads have been mapped, the makeTagDirectory program was used to process the reads. To find significant interactions, pairs of regions which have more Hi-C reads between them than would be expected by chance were found. The expected number of reads is calculated using the background model. The background model is used to model how many reads we expect to connect to each other region in the genome. If regions are far away (or on separate chromosomes), then one expects only a small number of reads to connect them. Likewise, if two regions are close to one another, one would expect a large number of interactions. These expectations are also dependent on the number of total reads mapping to each locus (which may depend on the number of restriction sites etc. in the region). The background model attempts to take all of these factors into account. The background model creation process can be broken down into the following steps (example uses resolution of 100 kb):

1. Divide the genome into putative regions (i.e. chr1:0-100 kb, chr1:100 kb-200 kb, ... chrY:8900 kb-9000 kb)

2. Calculate the total read coverage in each region

3. Calculate the fraction of interactions spanning any given distance with respect

to read depth

4. Optimize read count model to correctly assign expected interaction counts in regions with uneven sequencing depth

5. Calculate variation in interaction frequencies as a function of distance.

Boundaries and TADs were identified as previously described . Minimum boundary size was set to 40 kb. To perform PCA on Hi-C data, we applied it to the normalized interaction matrix, as previously described .

Immunoprecipitation.

Cells were collected with cold PBS and lysed with extraction buffer (20 mM HEPES pH 7.9, 25% glycerol, 420 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 0.5 mM DTT, protease inhibitors). The lysate was diluted with dilution buffer (20 mM HEPES pH 7.9, 1.5 mM MgCl2, 0.2 mM EDTA, 0.5 mM DTT, protease inhibitors) to the final 150 mM NaCl concentration. 2®C5 μg of antibodies were added into the diluted cell lysate and incubated overnight at 4 °C. The next day, the protein complexes were collected by magnetic Dynabeads G for 2 hr at 4 °C with rotation. The beads®Cantibody®Cprotein complexes were then washed four times with wash buffer (PBS, 0.03% Triton X-100) and boiled for western blot analysis.

Nuclear matrix extraction.

Biochemical extraction of nuclear matrix fraction was performed as previously described with several modifications. Briefly, cells were collected and

nuclei extracted as described. 7.5 mM LIS extraction step was performed and nuclear halos were washed in DNAse digestion buffer. After DNAse I digestion (10 u/50 ul of nuclei, 1hr on ice) nuclear matrices were recovered by centrifugation. The remaining pellet was washed in DNAse digestion buffer and subjected to further steps. Phenol/chloroform extraction of DNA was preceded by proteinase K digestion for 3 hr at 55 deg. DNA was sonicated (Bioruptor, 2x30s) before library preparation for sequencing. For nuclear matrix extraction from mouse brain, 9 week old mice were sacrificed and cortical nuclei were purified from cortex using a 15%: 60% Percoll step gradient centrifugation at 16,000 G for 45 minutes.

Chapter 6 includes unpublished work with Amir Gamliel, Bogdan Tanasa, and Skowronska-Krawczyk.

CHAPTER 7
Condensin I and II complexes license full estrogen receptor α-dependent enhancer activations.

Enhancers instruct spatio-temporally specific gene expression in a manner tightly linked to higher-order chromatin architecture. Critical chromatin architectural regulators condensin I and condensin II play non-redundant roles controlling mitotic chromosomes. But the chromosomal locations of condensins and their functional roles in interphase are poorly understood. Here we report that both condensin complexes exhibit an unexpected, dramatic estrogen-induced recruitment to estrogen receptor α (ER-α)-bound eRNA+ active enhancers in interphase breast cancer cells, exhibiting non-canonical interaction with ER-α via its DNA-binding domain (DBD). Condensins positively regulate ligand-dependent enhancer activation at least in part by recruiting an E3 ubiquitin ligase, HECTD1, to modulate the binding of enhancERassociated coactivators/corepressors, including p300 and RIP140, permitting full eRNA transcription, formation of enhancer:promoter looping, and the resultant coding gene activation. Collectively, our results reveal an important, unanticipated transcriptional role of interphase condensins in modulating estrogen-regulated enhancer activation and coding gene transcriptional program. Enhancers empower the genome with a precise control of temporally and spatially necessary gene expression patterns (Plank and Dean, 2014). The recent discovery of pervasive transcription of enhancer RNAs (eRNAs) revealed enhancers themselves as transcription units (Kim et al., 2010). eRNA levels correlate highly with enhancer activities (Andersson et al., 2014, Hah et al., 2013, Kim et al.,

2010, Melgar et al., 2011, Wang et al., 2011 and Wu et al., 2014), and both enhancer transcription and transcripts were found to contribute to enhancer function (Hsieh et al., 2014, Kaikkonen et al., 2013, Lai et al., 2013, Lam et al., 2013, Li et al., 2013a, Melo et al., 2013, Mousavi et al., 2013, Pnueli et al., 2015 and Schaukowitch et al., 2014), adding an important layer of understanding into the fundamental mechanisms underlying enhancer action (Lam et al., 2014). However, the complete molecular mechanisms that control the appropriate transcriptional output of enhancers and subsequent activation of coding genes remain eluisive. The long-range nature of enhancer functions tightly connects their regulation to chromatin architectures (Plank and Dean, 2014). Cohesin has recently been shown to positively regulate transcription by modulating enhancer function and enhancer-promoter looping (Kagey et al., 2010, Li et al., 2013a and Schmidt et al., 2010), raising the possibility that other architectural complexes important in mitosis/meiosis, particularly condensins, might as well assume critical roles on enhancers and/or in transcription regulation (Wood et al., 2010; Hirano, 2012). Condensins are highly conserved multi-subunit complexes containing structural maintenance of chromosome (SMC) proteins. Together with two other such SMC-containing complexes, cohesin and SMC5/SMC6 complexes, they contribute to the formation, maintenance, and dynamics of eukaryotic chromosome architecture (Wood et al., 2010; Hirano, 2012 and Jeppsson et al., 2014). In vertebrates, two related condensin I and II pentameric complexes, exhibiting similar topological structures (Wood et al., 2010; Hirano, 2012), play non-overlapping but critical roles for chromosome packing in mitosis

(Green et al., 2012, Hirano, 2012 and Ono et al., 2003). Compared to roles in mitosis, less is known about condensin functions in interphase. Condensin I was originally considered mainly cytoplasmic during interphase, whereas condensin II has been recognized to exhibit a nuclear localization, thought to concentrate on chromatin until prophase (Hirano, 2012 and Ono et al., 2003). In particular, it remains largely unclear where condensin I and condensin II are localized on the interphase chromatin, how they are recruited, and by what mechanisms they exert functions, if any, in transcription regulation. In this study, we found that, surprisingly, multiple condensin I and condensin II subunits are rapidly, specifically, and strongly recruited to estrogen receptor α (ER-α)-bound, functionally active enhancers in response to estrogen stimulation in human breast cancer cells. The loading of interphase condensins to these active enhancers was likely achieved by interaction with ER-α via the DNA-binding domain (DBD) of the latter. Mechanistically, condensins were required for full ligand-activated eRNA transcription, at least in part based on its recruitment of an E3 ubiquitin ligase, HECTD1, which modulates proper recruitment of transcriptional coactivators and corepressors via ubiquitinating and dismissing a specific corepressor, RIP140. This regulatory event then licenses full RNA polymerase II (Pol II) loading to enhancers, eRNA transcription, and enhancer:promoter (E:P) chromosomal interactions, leading to upregulation of target coding genes. Our current data have thus identified an unexpected, enhancer-based important function of condensin complexes in regulated transcriptional control, which is likely to be required for at least some other classes of DNA-binding transcription

factors in diverse cell types. About 10%– 30% and ~50% of condensin I and condensin II proteins, respectively, were found as chromatin-associated in MCF-7 breast cancer cells, consistent with findings in other cell types (Heale et al., 2006 and Ono et al., 2003). Following 3–4 days of culture in serum-deficient "stripping" medium, MCF-7 cells were largely (~80%– 95%) blocked in the G0/G1 phase (Villalobos et al., 1995), in contrast to the status without stripping (~42% G0/G1). Even double thymidine block did not further enrich cells in G1/S, providing an ideal cell line model to study potential interphase functions of condensins. ChIP-seq with an antibody against the NCAPG subunit of condensin I identified 2,916 peaks genome-wide in cells cultured in the absence of estrogen treatment (i.e., 100 nM of 17-β-estradiol or E2), strikingly increasing to 7,292 peaks 1 hr after E2 treatment. NCAPG in the majority bound to intergenic (55%) and intronic regions (36%), with only ~4.6% located on RefSeq gene promoters. Remarkably, ~77% (5,623/7,292) of all NCAPG binding peaks overlapped with those of ER-α, and ESR1/ERE was the most enriched motif for all the NCAPG binding sites by HOMER motif analysis (Heinz et al., 2010). Analogous experiments for condensin II (i.e., NCAPH2) revealed similar enrichment to intergenic/intronic regions, remarkable gain of peaks after E2 treatment (from 3,636 to 10,192), and high overlap with ER-α binding. Specificity of the antibodies was confirmed as the knockdown of the mRNAs encoding these two proteins by small interfering RNAs (siRNAs) resulted in dramatic reduction of their binding. ChIP-seq by a specific antibody against NCAPH, another subunit of condensin I, despite being less robust, also yielded

predominantly intergenic and intronic locations, consistent with NCAPG results, overlapping ER-α binding sites E2 treatment caused a switch of the motif enriched on NCAPH-bound intergenic sites to ERE. These ChIP-seq results were yet further confirmed by using a second antibody from independent source against NCAPG (NCAPG [Y.K.], a generous gift from K. Yokomori). Previous work from our lab (Li et al., 2013a) and others (Hah et al., 2013) established that a sub-group of ER-α/H3K27Ac co-bound enhancers exhibiting E2-upregulated eRNAs, high intensity of ER-α binding, and close proximity to estrogen target genes constitute the major E2-activated functional enhancers in MCF-7 cells, referred to as E2-induced "active enhancers" or "eRNA+ enhancers". Besides these active enhancers (n = 1,248), ER-α/H3K27Ac co-bound sites contained another group, which we referred to as non-active/"primed" enhancers, displaying no significant eRNA induction, a lower ER-α binding intensity, and lack of Pol II or p300 increase in response to E2; these "primed" enhancers also exhibited higher levels of H3K27me3.

ChIP-seq analyses showed that both condensin I (NCAPG) and condensin II (NCAPH2) were strongly induced by E2 to bind the active enhancers, but not the primed cohort, and their binding on active enhancers was highly correlated with that of ER-α. By analyzing published ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) datasets, we found that the ER-α binding sites involved in chromosomal looping (Fullwood et al., 2009) exhibited stronger recruitment of NCAPG and NCAPH2 than those without looping, further suggesting condensins' enrichment on functional enhancers.

When we compared the binding of condensins at active enhancers to that of other transcription factors/cofactors and histone marks, we found that NCAPG and NCAPH2 exhibited the most dramatically induced recruitment in response to E2, similar to that of SRC2 and SRC3, the classical ER-α coactivators (CoA), and ER-α itself. A hierarchical cluster analysis (HCA) of the E2-induced binding of these factors indicated that condensins exhibited the highest correlation with ER-α. Together, these data reveal that both condensin I and II complexes are preferentially recruited to ER-α-regulated, functionally active enhancers upon E2 treatment and exhibited the most dramatic induction in response to ligand.

As a comparison, we also examined NCAPG recruitment in mitotic MCF-7 cells. The binding of NCAPG to ER-α-bound sites was partially diminished in asynchronized MCF-7 cells that contain □30% mitotic cells, as exemplified by TFF1 locus, which further decreased to very minimal levels in mitosis-enriched cells. This was observed on most of the active enhancers, indicating that the observed binding of condensins to active enhancers represents interphase events.

To test for potential interactions between condensin complexes and ER-α, we first performed gel filtration experiments using MCF-7 nuclear extract and found highly overlapped co-fraction profiles of them, which are particularly prominent in a range between □1 and 1.5 MDa. It is noted that some subunits (e.g., SMC2 and NCAPH2) and ER-α itself are also present in additional fractions. Co-immunoprecipitation (coIP) experiments revealed that specific condensin subunits co-precipitate with ER-α. Reciprocally, both the endogenous

and overexpressed ER-α can pull down multiple condensin subunits and these interactions were unlikely bridged by DNA. ChIP-western result suggests that NCAPG/ER-α interaction takes place on chromatin. Importantly, condensin/ER-α interaction was not disrupted when lysine 539 of ER-α was mutated to alanine, a mutation that precludes ligand-dependent binding of LxxLL-containing nuclear receptor coactivators (CoAs) and corepressors (CoRs) (Ruff et al., 2000), as exemplified by its failure to bind SRC3 or RIP140. By contrast, the DBD of ER-α exhibited the strongest association with condensins. These data suggested that condensins interact with ER-α in a non-canonical manner, very distinctive from those "LxxLL"-containing coregulators.

ICI 182780 (ICI), a downregulator of ER-α, completely abolished E2-induced condensin recruitment to active enhancers, suggesting the recruitment depends on ER-α, likely via their direct interaction. Condensin protein levels were not obviously altered by E2 or ICI. No direct interaction between condensin I and II complexes was detected in our coIP experiments. Consistent with this, a mass spectrometry experiment following NCAPG IP showed that NCAPG pulls down condensin I subunits and SMCs, but not condensin II constituents. Interestingly, several E3 ubiquitin ligases, as well as ubiquitin itself, were found to coIP with NCAPG, suggesting that condensins may associate with unappreciated partners in breast cancer cells.

Condensin I and II co-localized on 5,253 sites in the genome, and most of these co-bound regions are occupied by ER-α (□94%). This high overlap differs from their largely non-overlapping chromatin localization in mitosis

(Green et al., 2012, Hirano, 2012 and Ono et al., 2003). Two-step ChIP experiment followed by qPCR revealed that ER-α-co-occupied chromatin regions were simultaneously bound by condensins. Interestingly, NCAPH2 and NCAPG also reciprocally re-ChIP each other. These data suggested that, although condensin I and II do not directly interact (Ono et al., 2003), they simultaneously co-occupy ER-α-bound active enhancers. But these data could not prove if condensin I, condensin II, and ER-α all bind the same regions simultaneously. Knockdown of NCAPH subunit reduced the binding of NCAPG to ER-α binding sites, suggesting that the binding of condensins happens as a complex; it also decreased the protein levels of the whole condensin I complex. Condensin I knockdown (i.e., siNCAPG) did not apparently affect either the binding or complex stability of condensin II. These data together suggested that it is unlikely to have a mix-and-match condensin complex recruited to enhancers in MCF-7 cells, but we could not exclude the possibility of any sub-stoichiometric condensin complex formed in the MCF-7 cells, and/or only present in certain genomic regions.

We used at least two different siRNAs to effectively knock down multiple condensin subunits, which resulted in significantly dampened E2 activation of multiple coding genes interrogated, as shown by RT-qPCR results of TFF1, FOXC1, SMAD7, SIAH2, and PGR expression. This did not involve any changes of ER-α mRNA or protein levels. Results from global run-on sequencing (GRO-seq) confirmed this inhibition. Upon depletion of either NCAPG or NCAPD3, the E2-induced fold change (E2-FC) of all E2-upregulated coding

genes is significantly reduced. E2 target genes exhibited a dramatic transcriptional attenuation upon NCAPG knockdown, as shown by their whole-gene profiles. Correspondingly, RNA Pol II loading was also decreased over their gene bodies. Either NCAPG or NCAPD3-activated gene groups showed a high overlap with E2-upregulated targets, suggesting that condensins are widely involved in the E2-dependent gene activation program. In support of this, ESR1 appeared as a top Gene Ontology (GO) term for NCAPG-activated gene cohorts. In addition, the two condensin complexes regulated a partially overlapping category of coding genes, consistent with their partially overlapped chromatin localization. Representative browser images of GRO-seq and Pol II ChIP-seq are shown for TFF1 locus. Together, these data indicate that condensins play an important role in activating the expression of estrogen target genes, acting at the level of transcription.

Flow cytometry experiments revealed no significant change of cell cycle after condensin depletion in MCF-7 cells, excluding the possibility that the observed transcription inhibition was caused by indirect effects. This is consistent with previous results that single condensin knockdown did not obviously impact mitotic index (Ono et al., 2003).

Promoter binding was a rare event for condensin I, condensin II, or ER-α, for the coding genes they regulate. This made us focus on their possible roles acting on enhancers, given their strong enrichment there. We first tested if condensins have any role in controlling eRNA transcription, a key marker of active enhancers (Andersson et al., 2014, Hah et al., 2013, Kim et al., 2010, Lam

et al., 2014, Wang et al., 2011 and Wu et al., 2014). RT-qPCR results revealed clear inhibition of eRNAs when specific condensin subunits were knocked down, accompanied by reduced Pol II loading. Genome-wide analyses of GRO-seq and RNA Pol II ChIP-seq confirmed the quantitative but significant inhibition of enhancer transcription upon depletion of condensins. Screenshots of these data for TFF1 and FOXC1 enhancers are shown. Interestingly, eRNA levels are higher from enhancers neighboring genes regulated by condensin I or II, or both, as compared to those next to genes regulated only by E2, consistent with the possibility that condensins function through modulating highly transcribing enhancers.

In accord with the role/contribution of eRNA to the formation of E:P looping (Hsieh et al., 2014, Lai et al., 2013, Li et al., 2013a and Pnueli et al., 2015), we found that the knockdown of a condensin subunit (siNCAPH2) inhibited E:P contact frequency in the TFF1 locus by 3D-DSL assays. The effect is specific for the interrogated E:P looping, as there is no clear change of other interactions surrounding the acceptor sites. The reduced E:P looping was confirmed by 3C-PCR assays using another restriction enzyme, SacI. A PCR product migrating at the predicted size (i.e., □1.1 kb) was detected only in ligated samples and exhibited clear reduction upon depletion of a condensin subunit. Sanger sequencing confirmed the identity of this PCR product. A similar reduction of specific E:P looping was found in the FOXC1 locus by both 3D-DSL and 3C-PCR, with concomitant decrease of the eRNA. As a control, a condensin-independent gene GATA3 did not exhibit obvious change of its E:P looping.

These data suggest that condensins play an important role during estrogen-induced enhancer activation, including allowing Pol II recruitment, eRNA transcription, and E:P looping.

Previous studies have revealed sequential/dynamic cofactor recruitment to classical ER-α-bound sites (Métivier et al., 2003 and Shang et al., 2000) preceding Pol II loading. To examine at which step condensins act, we performed ChIPs for pioneer factor FOXA1, several conventional CoAs and CoRs, and ER-α itself, upon knockdown of condensin subunits. This resulted in no significant change of either FOXA1 or ER-α binding at active enhancers. However, E2-induced increment of p300 binding was markedly inhibited, as indicated by both ChIP-qPCR and ChIP-seq. Reduction was also observed for recruitment of other CoAs including SRC1, SRC3, and TIP60. As many of these CoAs possess histone acetyltransferase activity, it is consistent that the histone H3 lysine 27 acetylation (H3K27Ac) displayed a significant decrease on the active enhancers, corroborating the conclusion that enhancer activation was compromised. The consequence also included a quantitative decrease of MED1 binding, consistent with its role in both transcription activation and E:P looping (Hsieh et al., 2014, Kagey et al., 2010 and Lai et al., 2013). Interestingly, loading of condensins seems to be a downstream event during enhancer activation subsequent to the trans-recruitment/assembly of the "MegaTrans" complex (Liu et al., 2014), as demonstrated by the reduction of condensin binding upon dual knockdown of RARα/RARγ, while RARα binding was unaltered when condensin was knocked down. A representative genome browser image of aforementioned ChIP-seq

results is shown for the TFF1 locus. To appreciate the extent of condensin effects, transient knockdown of p300 and NCAPG followed by RT-qPCR were performed, which revealed comparable inhibition of E2-target coding genes and eRNAs, although a partial effect of p300 knockdown could be attributed to the reduced level of ER-α itself. We also tested condensin knockdown on binding of two CoRs, RIP140 and CtBP1 (Watson et al., 2012 and White et al., 2005). Interestingly, while RIP140 displayed an E2-induced binding on several ER-α/condensin co-bound sites, its recruitment became further augmented when condensin was depleted. Knockdown of RIP140 increased transcription of several interrogated eRNAs by RT-qPCR, confirming its role as a CoR. Intriguingly, a concomitant increase of CtBP1 binding was detected. These observed changes of CoA and CoR binding should not be attributed to alterations of their protein amounts. These data together suggest that condensins license enhancer activation by maintaining a fine balance of E2-dependent recruitment of CoAs and CoRs.

We next explored the finding that some E3 ubiquitin ligases co-immunoprecipitated with NCAPG in mass spectrometry. We were particularly intrigued by a potential importance of HECTD1 (Sarkar and Zohn, 2012 and Zhou et al., 2012), a member of the HECT family, which possesses several cofactors for nuclear receptors (Nawaz et al., 1999 and Sun et al., 2014). CoIP experiments confirmed the interaction between condensin subunits and HECTD1 in MCF-7 cells. Like condensins, HECTD1 interacted with the DBD of ER-α, and the interaction was independent of the ER-α L539 residue. HECTD1 elution profile in gel filtration coincided well with those of condensins. Mapping of

interacting domains showed that the C terminus and a central fragment of HECTD1 interact with condensin subunit NCAPH. ChIP-qPCR using two different commercial antibodies against HECTD1, despite different affinities, both revealed an E2-induced binding to several condensin/ER-α binding sites. ChIP-seq using one of these HECTD1 antibodies identified 3,274 peaks genome-wide in liganded MCF-7 cells, about □45% and □41% of which overlapped the sites of ER-α and NCAPG, respectively. Heatmap analysis revealed the presence of HECTD1 and its E2 induction on the eRNA+ active enhancers, as exemplified by the TFF1 locus. Furthermore, HECTD1 and NCAPG binding exhibited high correlation. This is an interesting observation consistent with the finding that active ubiquitination and protein proteolysis events are enriched on active enhancers (Catic et al., 2013). When we knocked down NCAPG, the binding of HECTD1 on interrogated ER-α/condensin co-bound sites was significantly reduced, but the HECTD1 protein level was not affected, suggesting that HECTD1 was recruited in a condensin-dependent manner. Similar to condensin knockdown, depletion of HECTD1 caused an inhibition of p300 recruitment and increase of RIP140 binding to ER-α-regulated sites. This was accompanied by reduced transcription of eRNAs and coding genes in response to E2. To elucidate if the E3 ligase activity is important for HECTD1 function, a “rescue” experiment were performed that showed that an HA-tagged mouse wild-type HECTD1 (mHECTD1-WT) expression plasmid could rescue, at least in part, the eRNA inhibition resulting from HECTD1 knockdown in MCF-7 cells. In contrast, a HECTD1 mutant (C2579G) that is defective of E3 ligase activity (Sarkar and

Zohn, 2012) failed to produce rescue. These results together indicate that the condensin-dependent recruitment of HECTD1 is needed for full activation of eRNA transcription.

Previously, RIP140 was found polyubiquitinated in macrophages for proper inflammatory gene transcription (Ho et al., 2012). Considering that the absence of HECTD1 augmented RIP140 binding on active enhancers, we sought to test if RIP140 might be dismissed by or a direct substrate of HECTD1. Polyubiquitination (Ubn) of RIP140 could be detected in MCF-7 cells after E2 treatment and was enhanced by MG-132, a proteasome inhibitor. E2 treatment did not alter the total protein levels of RIP140 in MCF-7 cells until the addition of translation inhibitor cycloheximide (CHX) during E2 stimulation. A likely explanation for this is that E2 induced polyubiquitination of RIP140 and hence its degradation, but this was counteracted by new RIP140 synthesis. When HECTD1 was knocked down in the presence of CHX, the E2-triggered RIP140 reduction was disrupted, suggesting HECTD1 is the enzyme responsible for RIP140 turnover. In vivo ubiquitination assays were performed in 293T cells by ectopically co-expressing Flag-tagged RIP140 and HA-tagged wild-type HECTD1 or C2579G mutants, together with either wild-type or K48R mutant ubiquitin. Wild-type but not C2579G mutant of HECTD1 promoted RIP140 polyubiquitination in the presence of wild-type ubiquitin but not K48R mutant. Functionally, the reduced activation of E2 target genes and eRNAs due to siHECTD1 could be at least partially rescued by RIP140 knockdown. These data together suggest that RIP140 could be one of the functional substrates of

HECTD1 during E2-induced enhancer activation.

Where are interphase condensin I and condensin II on the chromatin? Our current study reveals a previously unsuspected interphase chromatin loading of condensin I and condensin II to the ER-α-bound active enhancers in a rapid, simultaneous manner in response to estrogen stimulus in human cancer cells. Their binding represents probably the most robust signature of the eRNA+ active enhancers and is distinctive from other chromatin structural molecules (e.g., cohesin). This dramatic enhancer enrichment is quite surprising, especially for mammalian condensin I, as it was considered to display low nuclear/chromatin abundance in interphase (Hirano, 2012). Intriguingly, chromatin-associated protein levels of condensins do not change by E2 treatment, implying that the enhancer-bound portion was re-distributed from other regions, reminiscent of the relocation of cohesin on yeast chromatin after initial loading (Lengronne et al., 2004).

The quite high co-localization between the two condensin complexes in interphase is distinct from their "non-overlapping" localization in mitosis (Hirano, 2012 and Ono et al., 2003), thus extending observations in other organisms (D'Ambrosio et al., 2008, Kim et al., 2013 and Kranz et al., 2013) or a study in murine stem cells reporting the presence of condensin II on (super-enhancers (Dowen et al., 2013). Interestingly, condensin II is not enriched at enhancers in Drosophila (Van Bortle et al., 2014), raising a possibility that their enhancer-based roles represent evolved functions. Importantly, our data provide insight into the poorly understood process of condensin loading to chromatin, that they are

recruited to regulatory elements by interacting with a transcription factor (i.e., ER-α). Of course, it is also possible that this initial recruitment by ER-α allows subsequent direct association of condensin with enhancer DNA by other strategies, such as topological entrapment (Piazza et al., 2014).

Functionally, GRO-seq data reveal that condensins activate gene expression and that they act at the level of transcription. These results are rather unexpected, as condensins were long considered to "condense" chromatin, which supposedly might attenuate transcription, as exemplified by the roles of condensin-like dosage compensation complex (DCC) in X chromosome gene repression in C. elegans (Wood et al., 2010). Interestingly, even in DCC-defective mutant worms, while expression of X genes increases, many autosomal genes seem to be reduced (Jans et al., 2009).

Mechanistically, condensins modulate the activation of ER-α-bound enhancers by regulating the balanced recruitment of coactivators versus corepressors. In turn, these events license RNA Pol II binding, eRNA transcription, and enhancer full activation. The effects of condensin knockdown on eRNA transcription were comparable to those observed with sip300, but appeared quantitative, likely suggesting certain redundancy or yet unknown mechanisms underlying eRNA transcription. On the basis of a role of eRNA in E:P looping formation and gene activation (Hsieh et al., 2014, Lai et al., 2013 and Li et al., 2013a), we suggest that the modulation of eRNAs by condensins is an important component of the full activation of coding target genes in response to regulatory signals. Consistent with this, condensin depletion reduces the

intensity/stability of interrogated E:P loopings. But our data could not clearly define if the looping defect is completely or partially a consequence of reduced eRNA levels. Indeed, there could well be a possibility that condensins directly control higher-order chromatin architecture, potentially by regulating topological domain borders (Hirano, 2012, Jeppsson et al., 2014 and Van Bortle et al., 2014). This role is actually reported for Drosophila NCAPH2 by a recent study (Li et al., 2015).

Mitotic condensins are thought to play structural roles in regulating chromatin, possibly through ATPase or DNA super-coiling activities (Hirano, 2012, Hudson et al., 2008 and St-Pierre et al., 2009), or topological entrapment of chromosomes (Cuylen et al., 2013 and Hirano, 2012). But in interphase, their regulatory mechanism on gene expression/transcription is rather elusive. Our data suggest that, at least in part, condensins exert interphase actions via recruiting specific ubiquitination machinery to control enhancer activation. These data provide insight into the long-observed dynamic/cyclic recruitment of CoAs (e.g., p300 and SRC3) and CoRs (e.g., RIP140) to nuclear receptors (Foulds et al., 2013, Métivier et al., 2003 and Shang et al., 2000). In accord with this finding, a recent genome-wide study of transcription factor ubiquitination revealed that active protein turnover by ubiquitination is required for gene regulation and often takes place on active enhancer regions (Catic et al., 2013).

Finally, the functions of condensins reported here in breast cancer cells have important disease implications. Indeed, mutations and altered expression of condensin subunits are associated with several cancer types (Emmanuel et al.,

2011, Leiserson et al., 2015, Murakami-Tonami et al., 2014 and Ryu et al., 2007); HECTD1 expression is elevated in ER-α-positive breast cancer patients in Oncomine databases and is important for ER-α-negative breast cancer cell invasion and metastasis (Li et al., 2013b). Given the important roles of condensins acting on tightly regulated specific genomic regions (e.g., enhancers), we are tempted to propose that dysregulated temporal or spatial loading of interphase condensins may lead to aberrant gene expression, possibly underlying human cancers.

Materials and Methods.

3C and 3D-DSL.

The procedures of 3C and 3D-DSL followed previous methods (Li et al., 2013a). 3C-PCRs were performed using SacI restriction enzyme, using primers pre-tested for their efficiency and linearity. For results presented, 30 cycles of PCR were performed, and lower cycles gave similar results. 3D-DSL was performed using BamHI

(for 3C step) and followed by DSL using oligonucleotides.

Chapter 7, in full, is a reprint of the material as it appears in Molecular Cell 2015. Enhancer activation requires trans-recruitment of a mega transcription factor complex. Li W, Hu Y, Oh S, Ma Q, Merkurjev D, Song X, Zhou X, Liu Z, Tanasa B, He X, Chen AY, Ohgi K, Zhang J, Liu W, Rosenfeld MG., 2015.

CHAPTER 8

What are some computational ways in which HiC analysis can be improved?

Introduction.

Chromosome conformation capture, or 3C, is a high-throughput molecular biology technique used to analyze the organization of chromosomes in a cell's natural state. Studying the structural properties and spatial organization of chromosomes is important for the understanding and evaluation of the regulation of gene expression, DNA replication and repair, and recombination. One example of chromosomal interactions influencing gene expression is a chromosomal region which can fold in order to bring an enhancer and associated transcription factors within close proximity of a gene, as was first shown in the beta-globin locus. Chromosome conformation capture has enabled researchers to study the influences of chromosomal activity on the aforementioned cellular mechanisms. This technology has aided the genetic and epigenetic study of chromosomes both in model organisms and in humans.

Primary Sources of Technical Bias in Hi-C Interaction Counts.

One of the goals when analyzing Hi-C data is to understand which loci in the genome tend to "interact" or "don't interact" in a biologically meaningful way. Unfortunately, the total number Hi-C reads between any two loci is dependent on many factors, many of which need to be normalized before meaningful conclusions can be reached. First, the read depth per region: Since Hi-C is an unbiased assay of genomic structure, one expects to observe equal read coverage across the genome. However, factors such as the ability to map reads uniquely

(e.g. density of genomics repeats), the number of restriction sites, and genomic duplications/structural variation in the experimental sample will all influence the number of reads. In addition, linear distance between loci along the chromosome can be a source of bias. If two loci are along the same polymer/chromosome of DNA, the loci are constrained with respect to one another independent of any specific structures adopted by the chromosome. More to the point, loci closely spaced along a chromosome are almost guaranteed to be 'near' one another if for no other reason their maximal separation is the length of DNA between them. As a result, closely spaced loci with have very high Hi-C read counts, regardless of their specific conformation. This is generally true of all 3C-based assays. Third, GC% bias, ligation preferences during library construction, normal sequencing problems can result in bias Hi-C counts. Finally, when looking for specific interaction in a particular chromatin environment, it can be useful to understand the properties of your local region of DNA to help interpret what is meaningful or just normal for that type of chromatin environment.

Normalizing Hi-C Data.

Since Hi-C analysis is an unbiased assay of nuclear topology, the expectation is that roughly equal numbers of Hi-C reads should originate from each region of equal size in the genome. If different numbers of Hi-C reads are observed, this is likely due to bias in mapping (e.g. repeats or duplicated genomic sequences in the region where no reads can be mapped), a variable number of restriction enzyme recognition sites (e.g. HindIII), or a technical artifact (i.e.

inaccessibility of HindIII, or any other restriction enzyme used, to DNA). For this reason, Hi-C reads between any two regions of a given size must be normalized for sequencing depth by dividing the total number of interaction reads that each region participates in. To calculate the expected Hi-C reads between two given regions, we use the following equation: $e_{ij} =(n_i)(n_j)/N$, where N is the total number of reads in the Hi-C experiment and n the total number of reads at each region i and j. This formulation assumes that each region has a uniform probability of interacting with any other region in the genome. In addition to sequencing depth, it is also useful to normalize data based on the distance between interacting regions. The constrained proximity of regions along linear DNA is by far the strongest signal in Hi-C data, with regions found at close linear distances much more likely to generate Hi-C reads than regions located at large linear distances along the chromosome or located on separate chromosomes. By computing the average number of Hi-C reads as a function of distance and sequencing depth, read frequencies between specific regions can be reported relative to the average Hi-C read density for their linear distances to help reliably identify proximity relationships that are not simply a result of the general linear compaction along the DNA.

The general idea was to modify the calculation of the expected number of HiC reads between any two loci to account for both their linear distance and sequencing depth. However, this calculation requires that the true number of interaction reads originating from a given locus, and not the measured number of interaction reads, are known. For example, if region A (mapped) is adjacent to a

region B (repeat region, unmappable), the total number of Hi-C reads mapping to A will likely be much less than a scenario in which B can be mapped as well. This is particularly important of regions adjacent to regions that cannot be mapped where their linear proximity would predict a significant numbers of Hi-C reads to connect the regions. This phenomenon causes a problem when trying to estimate the expected number of reads connecting region A to another region C. Since the total number of reads mapped to A is less because of region B adjacent to A that cannot be mapped, we likely underestimate the number of Hi-C reads mapping A to C.

To account for this, we find the expected number of Hi-C reads can be found with the following equation: e_{ij} = f(i-j) (n*i)(n*j)/N*, where f is the expected frequency of Hi-C reads as a function of distance, N* is estimated total number of reads, and n* is the estimated total number of interaction reads at each region. The goal was to identify n*i such that the total number of expected reads at each region i as a function of sequencing depth and distance (S_i=sum(e_{ij})) is equal to the observed number of reads at each region n_i.

Because the expected number of reads in any given region depends on the number of reads in all other regions, the resulting nonlinear system is difficult to solve directly. Instead, a simple hill climbing optimization was used to estimate inferred total reads. To calculate the inferred number of reads at each region, the model for expected interactions above is used to compute the expected number of read totals for each region, using the actual numbers of interaction reads as the initial values. The difference between the observed number of reads and the

expected number of reads is then used to scale the values for the estimated numbers of reads. This is computed for each region iteratively and repeated until the error between expected and observed Hi-C read totals per region is near zero.

Creating Hi-C Background Models.

The primary/default normalization method attempts to normalize the data for sequencing depth and distance between loci. To speed-up analysis, HOMER creates a "background model" that saves important parameters from normalization so that the background model only has to be computed once for a given resolution. The background model creation process can be broken down into the following steps (example uses resolution of 100 kb): 1. Divide the genome into putative regions (i.e. chr1:0-100 kb, chr1:100 kb-200 kb, ... chrY:8900 kb-9000 kb). 2. Calculate the total read coverage in each region. 3. Calculate the fraction of interactions spanning any given distance with respect to the read depth. 4. Optimize read count model to correctly assign expected interaction counts in regions with uneven sequencing depth. 5. Calculate variation in interaction frequencies as a function of distance.

Counting significant interactions.

One can search can search for pairs of loci that have a greater number of Hi-C reads that expected by chance, which below will be referred to as a significant interaction. The enhanced proximity of these regions may have relevant biological interpretation. It is important to note up front that it is

basically unheard of for two loci to always co-localize next to one another (i.e. in same cross-linked complex). Usually regions simply show "enrichment" for their co-localization, meaning that evidence for their co-fixation comes from only a fraction of the total cells used in the experiment. This also means that some regions may have significant interactions with multiple other loci in the same experiment.

The premise behind finding significant interactions is simple enough: Look for pairs of regions that have more Hi-C reads between them than would be expected by chance. The expected number of reads is calculated using the background model explained previously. The background model is used to model how many reads one expects to connect to each other region in the genome. If regions are far away (or on separate chromosomes), then only a small number of reads to connect them. Likewise, if two regions are close to one another, one would expect a large number of interactions. These expectations are also dependent on the number of total reads mapping to each locus (which may depend on the number of restriction sites etc. in the region). The background model attempts to take all of these factors into account. Since the total number of reads per region is fixed (and more or less constant for each region due to the unbiased nature of Hi-C), one tests how these reads are distributed relative to the expectation. For two given loci that could potentially interact, one models their randomly expected read counts using the cumulative binomial distribution, where the total number of trials is the number of reads that could possibly map between the loci (i.e. the region total), the rate of success is the expected interaction

frequency, and the number of observed successes is the number of observed reads mapping between the loci. In this setting, regions with only 1 or 2 reads between them will have high p-values, regardless of their expected interaction frequency, while regions with many interactions above expected will have low p-values.

The 3D chromatin structure modeling by chromatin interactions derived from Hi-C experiments is significantly challenged by the intrinsic sequencing biases in these experiments. Conventional modeling methods only focus on the bias among different chromatin regions within the same experiment but neglect the bias arising from different experimental sequencing depth and biological/technical replicates. However, the wide use of these methods is limited by the sequencing biases of Hi-C derived data. First, it is pointed out that the raw Hi-C chromatin interactions have systematic biases resulted from experiment, such as restriction enzymes, GC content and sequence uniqueness. The current bias reduction and 3D modeling schemes only focus on the sequencing bias within the same experiment caused by differences in enzyme efficiency and sequence coverage for different chromatin regions but neglect the bias arising from another important factor, experimental sequencing depth. Second, a lot of modeling approaches are performed at megabase resolution because it is difficult to reduce systematic bias at higher resolution. It is known that functional structural rearrangement often occur in genomic sizes ranging from hundreds of kilobases to megabases. These concerns stimulated our interest to propose a novel strategy to reduce sequencing-dependent biases by normalizing Hi-C data with the inherent characteristics of chromatin interactions. Thus, the low-resolution

modeling can only provide information on global chromatin structure but prevent its application for investigating 3D structure of functional chromatin regions. In order to investigate a potential algorithmic solution to this problem, 5C in MCF7 cells using minus and plus estradiol conditions is used. To improve the mappability of this part of reads, all reads are scanned to identify the existence of potential ligation junctions for the expected sequence 'AAGCTAGCTT' from HindIII libraries, and then the junction and all bases after the 3′ of the junction are removed. The kept reads are then aligned to the reference human genome hg18 by using bwa-0.6.1-r104 with default settings. Only the uniquely mapped paired-reads (mapping quality > 30 for both reads) are selected for the next processing. The pairs that originated from PCR duplication are removed, and the pairs without enzyme restriction site after the downstream 500 bp is also removed to exclude incomplete exonuclease action. Finally, all pairs aligned < 10000 bp are considered as self-ligation and further eliminated from data set. In addition, an additional experiment was performed by a different technology: tethered Hi-C (called tethered conformation capture). Tethered conformation capture (TCC) is a method for genome-wide mapping of chromatin interactions. By performing ligations on solid substrates rather than in solution, TCC substantially enhances the signal-to-noise ratio, thereby facilitating a detailed analysis of interactions within and between chromosomes. The challenge of normalizing Hi-C derived data sets from disparate regions and experiments lies in developing an algorithm which improves on the bias from two or more replicates is proposed.

Algorithm for combining replicate interactions.

For replicates, interactions can be estimated by two various algorithm from replicates of HiC experiments 1,…,n. Let x_1, …, x_n be HiC interaction counts for each replicate 1,…,n. For this simple model, the design matrix is: X= [1…1 $x_1..x_n$] ^{T}, and the hat matrix H is the matrix of the orthogonal projection onto the column space of the design matrix: $H= X(X^TX)^{-1}X^T$. The "leverage" h_{ii} is the ith diagonal entry in the hat matrix. The variance of the ith residual is: $Var(e_i)= d^2(1-h_{ii})$. In case the design matrix X has only two columns (as in the example above), this is equal to: $Var(e_i)= d^2(1-1/n- (x_i-mean(x)^2/S(x_j-mean(x)^2)$. The corresponding studentized residual is then: $e_i/(f(1-h_{ii})^{0.5})$, where f is an appropriate estimate of d. This algorithm provides a way to combine replicates of HiC/TCC by approximating the “compromised” HiC/TCC interaction measurement.

A second algorithm can be used for 2 replicates, x and y. Consider a set of m data points (x_1,y_1), (x_2,y_2), …, (x_m,y_m), or HiC interactions counts, and a curve (model function) y=f(x,b), that in addition to the variable x also depends on n parameters, with m>=n. It is desired to find the vector b of parameters such that the curve fits best the given data in the least squares sense, that is, the sum of squares: $S= Sr_i^2$ is minimized, where the residuals (errors) r_i are given by: $r_i= y_i-f(x_i,b)$ for i=1,2,…,m. The minimum value of S occurs when the gradient is zero. Since the model contains n parameters there are n gradient equations. This algorithm provides a way to combine replicates of HiC/TCC by approximating the “compromised” HiC/TCC interaction measurement. However, only two replicates can be used. Sometimes, linear approximations are not representative of the data. In that case, a non-linear solution to combining replicates can be used.

In a non-linear system, the derivatives $\P r_i/\P b_j$ are functions of both the independent variable and the parameters, so these gradient equations do not have a closed solution. Instead, initial values must be chosen for the parameters. These initial values can represent the observed HiC interaction counts. Then, the parameters are refined iteratively, that is, the values are obtained by successive approximation, $b_j \sim b_j^{k+1} = b_j^k + Db_j$.

Here, k is an iteration number and the vector of increments, Db is known as the shift vector. At each iteration the model is linearized by approximation to a first-order Taylor series expansion about b^k: $f(x_i,b) = f(x_i,b^k) + S\P f(x_i,b^k)/\P b_j(b_i - b_j^k) \sim f(x_i,b^k) + SJ_{ij}Db_j$. The Jacobian, J, is a function of constants, the independent variable and the parameters, so it changes from one iteration to the next. Thus, in terms of the linearized model, $S\P r_i/\P b_j = -J_{ij}$ and the residuals are given by $r_i = Dy_i - SJ_{is}Db_s$; $Dy_i = y_i - f(x_i,b^k)$. Substituting these expressions into the gradient equations, they become $-2SJ_{ij}(Dy_i - SJ_{is}Db_s) = 0$, which, on rearrangement, become n simultaneous linear equations, the normal equations: $SSJ_{ij}J_{is}Db_s = SJ_{ij}Dy_i$ $(j=1,\ldots,n)$.

The normal equations are written in matrix notation as: $(J^TJ)Db = J^TWDy$. When the HiC interaction counts are not equally reliable, a weighted sum of squares may be minimized, $S = SW_{ii}r_i^2$. Each element of the diagonal weight matrix W should, ideally, be equal to the reciprocal of the error variance of the measurement. The weights might represent the sequencing depth of each HiC experiment. The normal equations are then: $(J^TWJ)Db = J^TWDy$. These equations form the basis for the Gauss–Newton algorithm for a non-linear least squares

problem.

Before combining the replicates, all replicates need to be comparable. Several algorithms are possible, the Mann-Whitney U test and a multivariate version of the algorithm. In statistics, the Mann–Whitney U test (also called the Mann–Whitney–Wilcoxon (MWW), Wilcoxon rank-sum test (WRS), or Wilcoxon–Mann–Whitney test) is a nonparametric test of the null hypothesis that two populations are the same against an alternative hypothesis, especially that a particular population tends to have larger values than the other.

It has greater efficiency than the t-test on non-normal distributions, such as a mixture of normal distributions, and it is nearly as efficient as the t-test on normal distributions. The Wilcoxon rank-sum test is not the same as the Wilcoxon signed-rank test, although both are nonparametric and involve summation of ranks. The test involves the calculation of a statistic, usually called U, whose distribution under the null hypothesis is known. In the case of small samples, the distribution is tabulated, but for sample sizes above ~20 approximation using the normal distribution is fairly good. Some books tabulate statistics equivalent to U, such as the sum of ranks in one of the samples, rather than U itself. The U test is included in most modern statistical packages. It is also easily calculated by hand, especially for small samples. There are two ways of doing this test.

Method one involves comparing two small sets of observations, a direct method is quick, and gives insight into the meaning of the U statistic, which corresponds to the number of wins out of all pairwise contests. For each observation in one set, count the number of times it wins over any observations in

the other set (the other value loses if it is larger). Count 0.5 for any ties. The sum of wins and ties is U for the first set. U for the other set is the converse.

Method 2 involves the following steps: 1. Assign numeric ranks to all the observations, beginning with 1 for the smallest value. Where there are groups of tied values, assign a rank equal to the midpoint of unadjusted rankings [e.g., the ranks of (3, 5, 5, 9) are (1, 2.5, 2.5, 4)]. 2. Now, add up the ranks for the observations, which came from sample 1. The sum of ranks in sample 2 is now determinate, since the sum of all the ranks equals N(N + 1)/2 where N is the total number of observations. 3. U is then given by: $U_1= n_1n_2+n_1(n_1+1)/2-R_1$, where n_1 is the sample size for sample 1, and R_1 is the sum of the ranks in sample 1.

Note that it doesn't matter which of the two samples is considered sample 1. An equally valid formula for U is: $U_2= n_1n_2+n_2(n_2+1)/2-R_2$. The smaller value of U_1 and U_2 is the one used when consulting significance tables. The sum of the two values is given by: $U_{1+}U_2= n_1n_2+n_1(n_1+1)/2-R_1+ n_1n_2+n_2(n_2+1)/2-R_2$. Knowing that $R_1 + R_2 = N(N + 1)/2$ and $N = n_1 + n_2$, and doing some algebra, we find that the sum is: $U_1+U_2= n_1n_2$.

Now one can look at a multivariate version of the problem, namely to test whether the population means of the $k \times 1$ random vectors and Y are equal, i.e. the null hypothesis H0: $\mu X = \mu Y$.

Definition 1: The Two sample Hotelling's T-square test statistic is: $T^2=(\text{mean}(x)-\text{mean}(y))^T(S(1/n_x+1/n_y)]^{-1}(\text{mean}(x)-\text{mean}(y))$, where S is the pooled sample covariance matrix of X and Y, namely: $S=((n_x-1)S_x+(n_y-1)S_y)/((n_x-1)+(n_y-1))$, where S_X is the covariance matrix of the sample for X and the sample for each

random variable x_i in X has n_x elements, and similarly S_Y is the covariance matrix of the sample for Y and the sample for each random variable y_i in Y has n_y elements. Note the similarity between the expression for T^2 and the expression for T^2 given above.

Theorem 1: For n_x and n_y sufficiently large, $T^2 \sim \chi 2(k)$, there exists an observation: For small n_x and n_y, T^2 is not sufficiently accurate and a better estimate is achieved using the following theorem.

Theorem 2: Under the null hypothesis: $F= (n-k)/k(n-1)T^2 \sim F(k,n-k)$ where $n = n_x + n_y - 1$. If $F > F_{crit}$ then we reject the null hypothesis.

Sometimes constrains are needed for HiC interactions. Using HiC interactions as data, one can create a linear programming problem. Constrains can be based on maximum tags for all replicates, and weights based on tags in each HiC interactions.

Linear programming (LP; also called linear optimization) is a method to achieve the best outcome (such as maximum profit or lowest cost) in a mathematical model whose requirements are represented by linear relationships. Linear programming is a special case of mathematical programming (mathematical optimization). More formally, linear programming is a technique for the optimization off a linear objective function, subject to linear equality and linear inequality constrains. Its feasible region is a convex polytope, which is a set defined as the intersection of finitely many half spaces, each of which is defined by a linear inequality. Its objective function is a real-valued affine function defined on this polyhedron. A linear programming algorithm finds a point in the

polyhedron where this function has the smallest (or largest) value if such a point exists.

Linear programs are problems that can be expressed in canonical form: maximize C^Tx subject to AX<=b. and x>=0, where x represents the vector of variables (to be determined), c and b are vectors of (known) coefficients, A is a (known) matrix of coefficients, and $(.)^T$ is the matrix transpose. The expression to be maximized or minimized is called the objective function (c^Tx in this case). The inequalities $Ax \leq b$ and $x \geq 0$ are the constraints which specify a convex polytope over which the objective function is to be optimized. In this context, two vectors are comparable hen they have the same dimensions. If every entry in the first is less-than or equal-to the corresponding entry in the second then we can say the first vector is less-than or equal-to the second vector. Standard form is the usual and most intuitive form of describing a linear programming problem. It consists of the following three parts:

1) A linear function to be maximized. e.g. $f(x_1,x_2)= c_1x_1+c_2x_2$. 2) Problem constraints of the following form e.g. $a_{11}x_1+a_{12}x_2< -b_1$. $a_{21}x_1+a_{22}x_2< -b_2$. $a_{31}x_1+a_{32}x_2< -b_3$. 3) Non-negative variables e.g. $x_1>=0$, $x_2>=0$.

The problem is usually expressed in matrix form, and then becomes: max{c^Tx| Ax<-b && x>=0} Other forms, such as minimization problems, problems with constraints on alternative forms, as well as problems involving negative variables can always be rewritten into an equivalent problem in standard form.

The linear programming problem can be solved using the simplex

algorithm. Having observed non-linear regression and confirmed that the replicates are comparable, one can proceed with the algorithm as following: use k-means clustering for find weights and then weighted non-linear regression to combine HiC replicates. K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. The problem is computationally difficult (NP-hard); however, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

Given a set of observations (x_1, x_2, …, x_n), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k (≤ n) sets S = {S_1, S_2, …, S_n} so as to minimize the within-cluster sum of squares (WCSS). In other words, its objective is to find: argminSSêêx-$m_i$êê^2. One can first divide the genomic region into k*5 regions where k is the number of bins. K-means clustering is done for all of the k bins, resulting in k clusters. Each cluster would represent a bin in the HiC cluster map. These k clusters are then

evaluated using the Davies–Bouldin index, resulting in weights which are later used in non-linear regression to calculate the "combined" interaction values for all the HiC replicates.

Algorithm for assigning significance to HiC interactions with replicates.

Another problem using HiC replicates is assigning interaction p-values to interactions which appear multiple times in HiC replicates. One strategy would be to combine the p-values for each replicates. In order to assign significance to replicate interactions, Fisher's method can be used. First, Fisher's method combines extreme value probabilities from each test, commonly known as "p-values", into one test statistic (X^2) using the formula: $X^2_{2k} \sim -2\mathrm{Sln}(p_i)$, where p_i is the p-value for the i^{th} hypothesis test, or HiC intensity values for the ith replicate. When the p-values tend to be small, the test statistic X^2 will be large, which suggests that the null hypotheses are not true for every test.

When all the null hypotheses are true, and the p_i (or their corresponding test statistics) are independent, X^2 has a chi-squared distribution with 2k degrees of freedom, where k is the number of tests being combined. This fact can be used to determine the p-value for X^2. The distribution of X^2 is a chi-squared distribution for the following reason. Under the null hypothesis for test i, the p-value p_i follows a uniform distribution on the interval [0,1]. The negative natural logarithm of a uniformly distributed value follows an exponential distribution. Scaling a value that follows an exponential distribution by a factor of two yields a quantity that follows a chi-squared distribution with two degrees of freedom.

Finally, the sum of k independent chi-squared values, each with two degrees of freedom, follows a chi-squared distribution with 2k degrees of freedom.

A closely related approach to Fisher's method is based on Z-scores rather than p-values. If we let $Z_i = \Phi^{-1}(1-p_i)$, where Φ is the standard normal cumulative distribution function, then: $Z \sim Z_i/(k)^{0.5}$ is a Z-score for the overall meta-analysis. This Z-score is appropriate for one-sided right-tailed p-values; minor modifications can be made if two-sided or left-tailed p-values are being analyzed. One advantage of the Z-score approach is that it is straightforward to introduce weights, or, for example, the sequencing depth of each HiC experiment. If the i^{th} Z-score is weighted by w_i, then the meta-analysis Z-score is which: $Z \sim w_i Z_i/(w_i^2)^{0.5}$ follows a standard normal distribution under the null hypothesis. While weighted versions of Fisher's statistic can be derived, the null distribution becomes a weighted sum of independent chi-squared statistics, which is less convenient to work with.

Discussion.

This algorithm allows the ability to combine interactions and assign a statistical value using HiC/TCC replicates. This is a first attempt, and many improvements can be made. For example, other distributions for modeling chromatin interactions can be used.

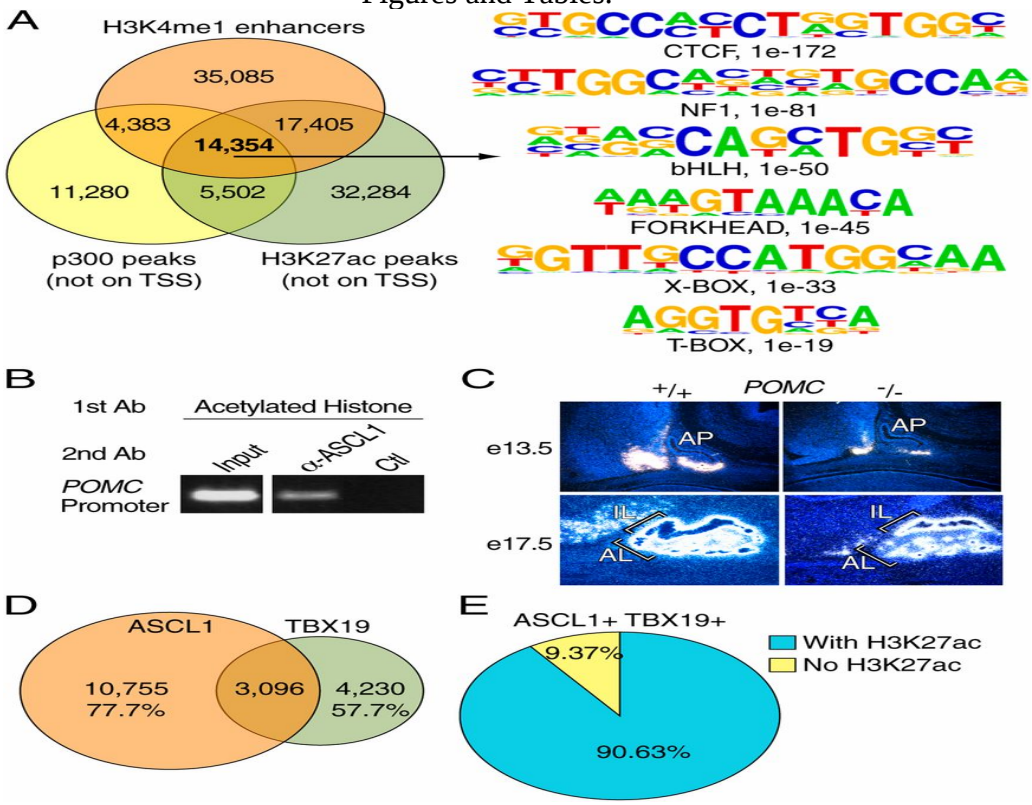


Figure 1 (Chapter 1): ASCL1 is required for the development of the POMC lineage and, together with TBX19, is highly correlated with the H3K27ac marker. (A, Left) Venn diagram showing the intersection of H3K4me1, p300, and H3K27ac ChIP-seq peaks excluding TSSs. The number of the peaks in all possible combinations of the three markers based on ChIP-seq detected in AtT20 cells are indicated. (Right) The enriched motifs of enhancers with p300 and H3K27ac markers. The top-ranked enriched motifs are shown with P values. (B) Sequential ChIP of adult pituitary glands by anti-H3K27ac followed by anti-ASCL1 compared with no-antibody control (beads) to demonstrate that ASCL1 is recruited to the H3K27Ac^{+}POMC promoter in vivo. The panel was sliced to remove unrelated content, as indicated by the blank space. (C) Null mutants of Ascl1 exhibit reduced POMC transcripts in both the pituitary gland and ventral diencephalon at E13.5 and E17.5. AL, anterior lobe; AP, adenohypophysis; IL, intermediate lobe. (D) Venn diagram showing the intersection of ASCL- and TBX19-containing enhancers. Numbers of enhancers are indicated. (E) The majority of ASCL1^{+} and TBX19^{+} enhancers are decorated with the H3K27ac histone marker.

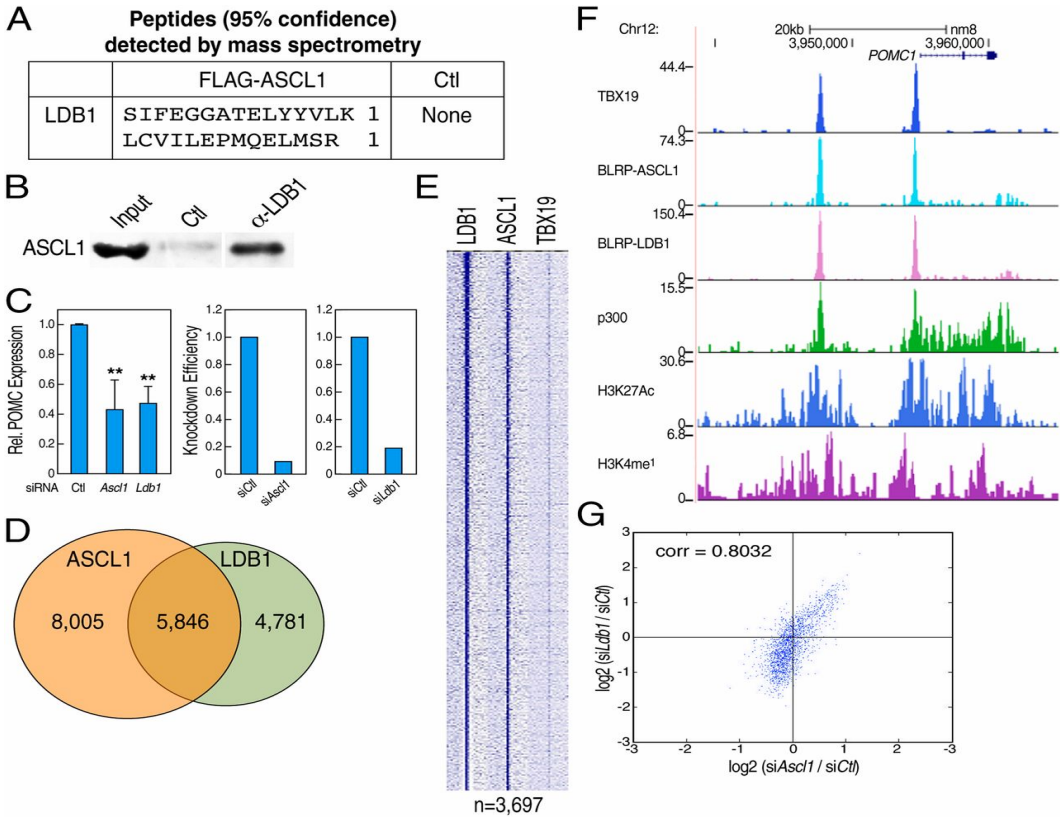


Figure 2 (Chapter 1): The ASCL1/LDB1 complex coregulates a large gene program in AtT20 cells.

The ASCL1/LDB1 complex coregulates a large gene program in AtT20 cells. (A) Mass spectrometry analysis of the FLAG–ASCL1 complex detected unique LDB1 peptides. The parental cell line was used as a control. (B) Coimmunoprecipitation of LDB1 followed by Western blotting confirmed endogenous LDB1–ASCL1 interaction in AtT20 cells. Beads alone served as a control. The panel was sliced to remove unrelated content, as indicated by a blank space. (C, Left) siRNAs against Ascl1 and Ldb1 down-regulate the expression of the POMC gene compared with AllStars Negative Control siRNA. (Center and Right) The knockdown efficiency of Ascl1 (Center) and Ldb1 (Right) in AtT20 cells. Data are presented as mean ± SEM (n = 3; **P < 0.01). (D) Venn diagram showing 5,846 ASCL1 peaks on enhancers colocalized with LDB1. (E) Heatmap showing 3,697 $ASCL1^{+}$ $TBX19^{-}$ peaks on enhancers colocalized with LDB1. (F) Snapshot of the genome browser shows the binding of TBX19, ASCL1, LDB1, and p300, together with histone markers H3K27ac and H3K4me1, on the POMC locus. Strong bindings of all factors are shown on both the POMC promoter and enhancer. (G) The Ldb1 transcriptional program was highly correlated with that of Ascl1. Genes in the proximity of cobound ASCL1 and LDB1 enhancers are shown. Data were obtained from the GRO-seq assay after knockdown of either the Ascl1 or Ldb1 gene compared with AllStars Negative Control siRNA. Corr, correlation coefficient.

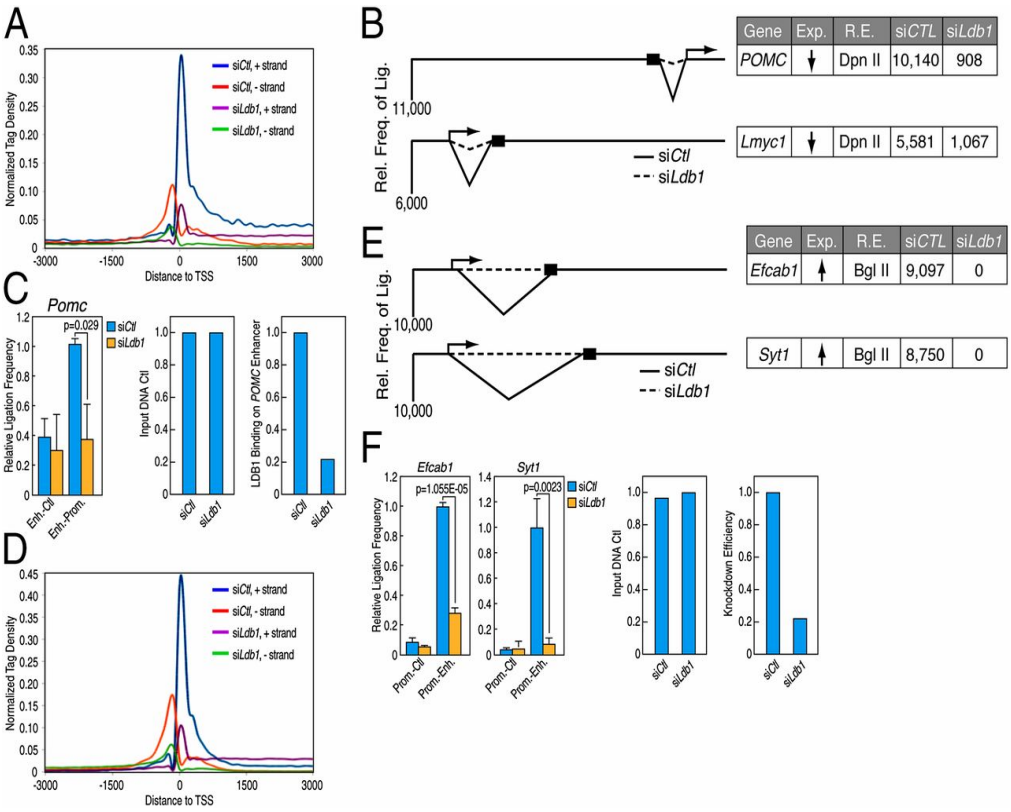


Figure 3 (Chapter 1): LDB1 is required for enhancer:promoter looping in both gene activation and gene repression.

(A) Tag density plot showing that LDB1 activates genes mainly at the transcriptional initiation level. (B) Promoter:enhancer looping is reduced for LDB1-activated genes upon knockdown of Ldb1 compared with AllStars Negative Control siRNA, as detected by PE3C-DSL (See Material and Methods and SI Materials and Methods for details). (C) 3C assays to confirm the promoter:enhancer interactions for LDB1-activated genes. (Left) The relative ligation efficiency is detected by quantitative PCR (qPCR). (D) Tag density plot shows that LDB1 represses genes by regulating transcriptional pausing. (E) Promoter:enhancer looping detected by PE3C-DSL is lost for LDB1-repressed genes upon knockdown of Ldb1, as compared with cells treated with AllStars Negative Control siRNA. DpnII was used as the restriction digestion enzyme (R.E.). Exp., expression level. (F) 3C assay to confirm the promoter:enhancer interaction in LDB1-repressed genes. (Left) The relative ligation efficiency is detected by qPCR. Data are presented as mean ± SEM; n = 3. (Center) Comparable amounts of DNA as determined by qPCR were used for ligation. (Right) Knockdown efficiency was evaluated by relative mRNA level, using qPCR. AllStars Negative Control siRNA was used as the control siRNA in all experiments.

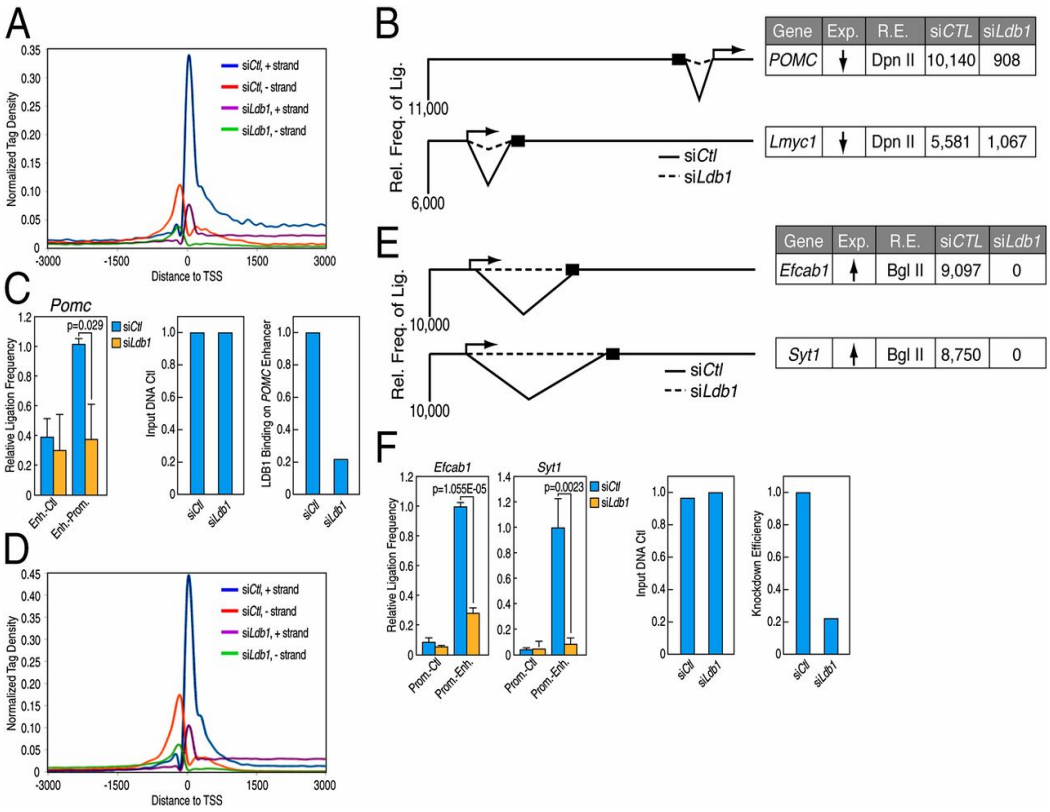


Figure 4 (Chapter 1): MTA is involved in LDB1 repression.

(A) Tag density plot showing that LDB1 activates genes mainly at the transcriptional initiation level. AllStars Negative Control siRNA was used as control. (B) Promoter:enhancer looping is reduced for LDB1-activated genes upon knockdown of Ldb1 compared with AllStars Negative Control siRNA, as detected by PE3C-DSL. The numbers show read counts from deep sequencing. DpnII was used as the restriction digestion enzyme (R.E.). Exp., expression level. (C) 3C assays to confirm the promoter:enhancer interactions for LDB1-activated genes. (Left) The relative ligation efficiency is detected by quantitative PCR (qPCR). Data are presented as mean ± SEM; n = 3. (Center) Comparable amounts DNA were used for ligation as determined by qPCR. (Right) The reduction of LDB1 binding on the POMC enhancer upon knockdown of Ldb1 was evaluated by ChIP assay with LDB1 antibody. (D) Tag density plot shows that LDB1 represses genes by regulating transcriptional pausing. (E) Promoter:enhancer looping detected by PE3C-DSL is lost for LDB1-repressed genes upon knockdown of Ldb1, as compared with cells treated with AllStars Negative Control siRNA. DpnII was used as the restriction digestion enzyme (R.E.). Exp., expression level. (F) 3C assay to confirm the promoter:enhancer interaction in LDB1-repressed genes. (Left) The relative ligation efficiency is detected by qPCR. Data are presented as mean ± SEM; n = 3. (Center) Comparable amounts of DNA as determined by qPCR were used for ligation. (Right) Knockdown efficiency was evaluated by relative mRNA level, using qPCR.

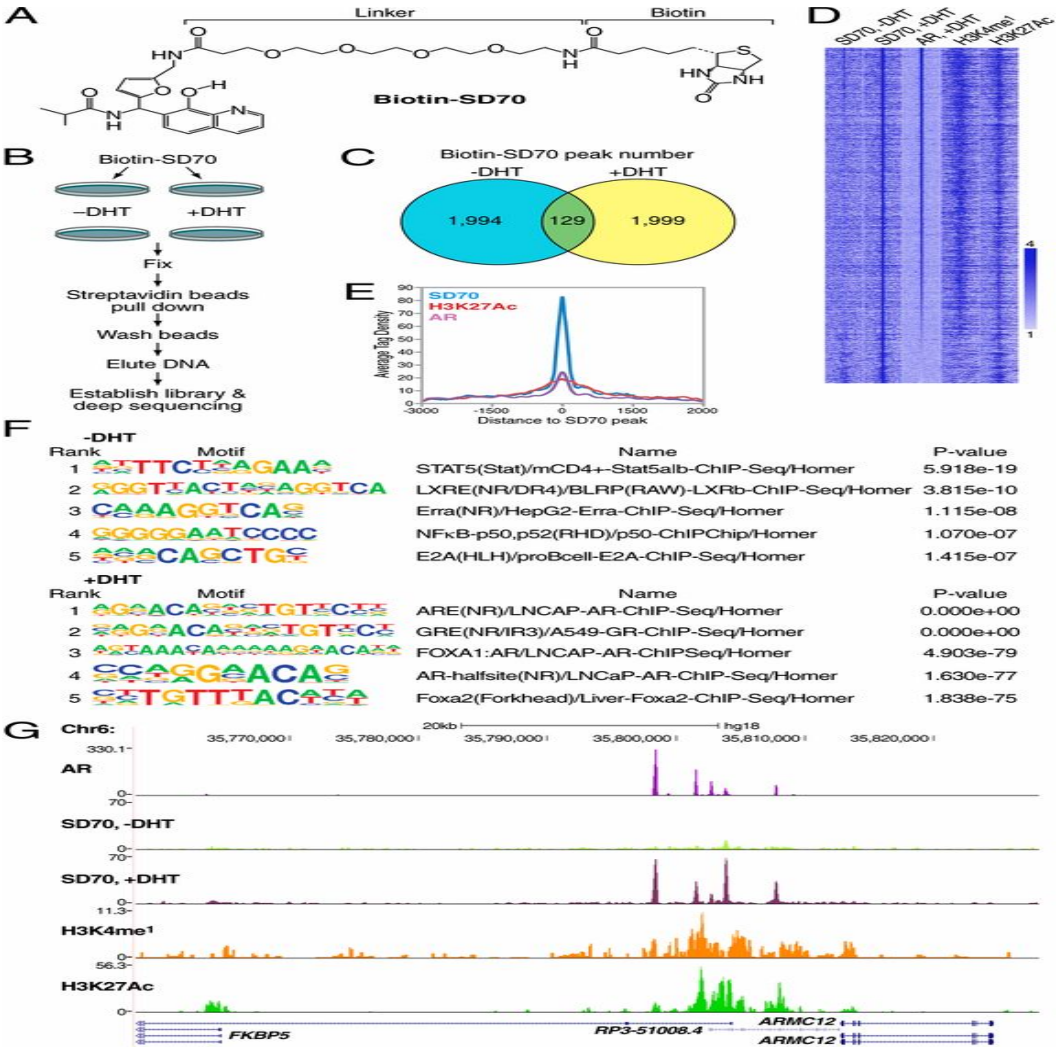


Figure 5 (Chapter 2): Chem-seq revealed colocalization of SD70 and AR enhancer genomic binding sites.

(A) Structure of Biotin-SD70. (B) Flowchart of SD70 Chem-seq procedure. (C) HOMER peak number of SD70 Chem-seq in the basal or DHT treatment condition. Detailed parameters are described in SI Methods. (D) Heatmap showing the distribution of Biotin-SD70, AR (GSM699631, +DHT), and the two enhancer marker [H3K4me1 (GSM686928) and H3K27Ac (GSM686937)] binding sites (−3kb/+3kb relative to the center of 1999 DHT induced Biotin-SD70 peak LNCaP cells). Each horizontal bar represents a single Biotin-SD70 binding site, and the color scale indicates the normalized tag density in 100-nt bins. Scale: lower limit = 0; midpoint value = 1; higher limit = 4. (E) Average tag profile analysis of the aligned 2,128 SD70 peaks showing focal SD70 binding in association with AR occupancy and H3K27Ac. (F) Motif enrichment of SD70 binding sites in the absence or presence of DHT condition. (G) UCSC genome browser shot for SD70 occupancy at the enhancer of AR target gene FKBP5, as an example, overlaying regions.

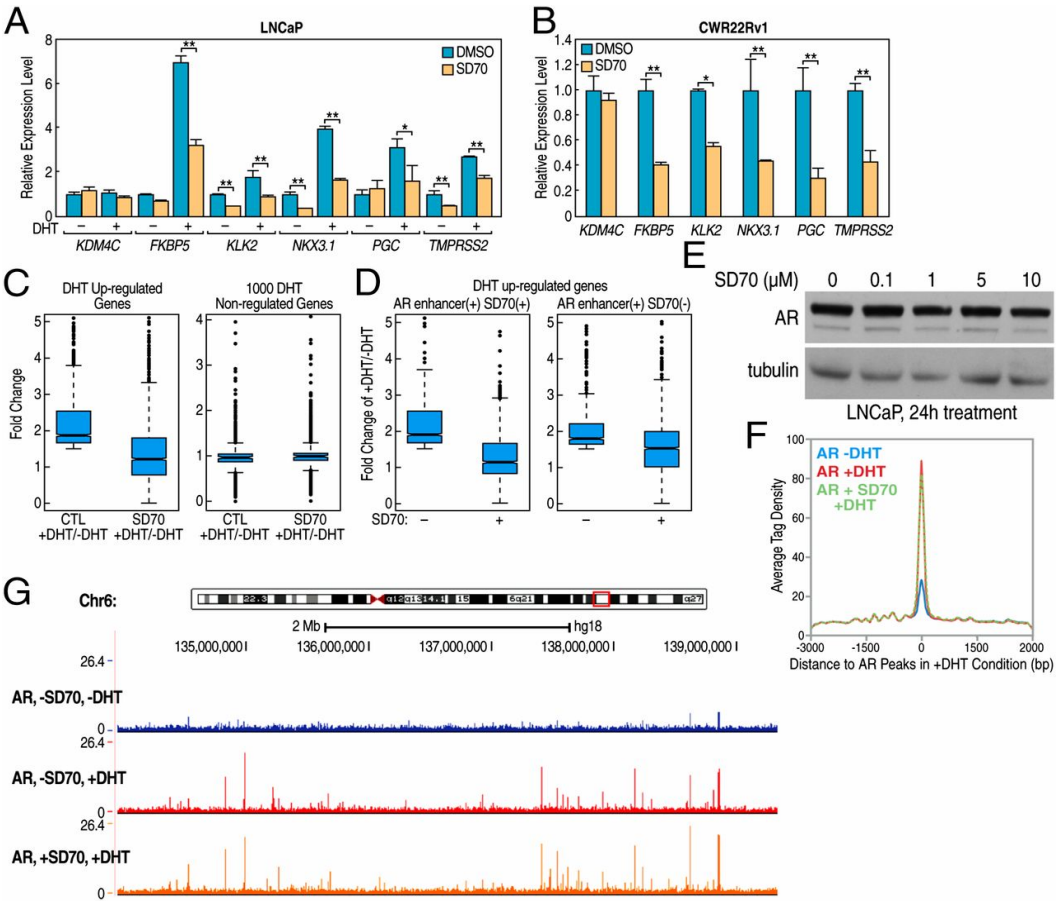


Figure 6 (Chapter 2): SD70 inhibits AR target gene expression.

(A) RT-qPCR analysis of KDM4C and some canonical AR target gene expression level in LNCaP cells treated with SD70 (10 μM) or vehicle (0.1% DMSO) for 2 h followed by DHT (100 nM) treatment for 4 h. (B) RT-qPCR analysis of KDM4C and some canonical AR target gene expression level in CWR22Rv1 cells treated with SD70 (10 μM) or vehicle (0.1% DMSO) for 2 h. Error bars represent SD for three repeats (*P < 0.05 and **P < 0.01). (C) Global DHT induced expression FCs for DHT up-regulated genes (n = 2,445, FC >1.5) in SD70 (10 μM) pretreated for 2 h versus vehicle before DHT (100 nM)-treated or nontreated LNCaP cells, determined by GRO-Seq. (C, Left) DHT induction effect was significantly reduced with treatment of SD70 ($P < 10^{-5}$). (C, Right) One thousand randomly chosen DHT nonregulated genes exhibiting no changes upon SD70 treatment as a negative control. (D) Global DHT induction changes in DHT-induced AR target genes by SD70 treatment. (E) AR protein level measurement upon indicated concentration of SD70 treatment for 24 h before Western blot. AR(441) antibody was used to detect endogenous AR protein, and tubulin served as a loading control. (F) LNCaP cells were hormone-stripped for 3 d and pretreated by SD70 (10 μM, 24 h), which were further treated with vehicle or DHT (100 nM, 1 h) and subjected to AR ChIP-seq. (G) Snapshot of AR peaks does not change.

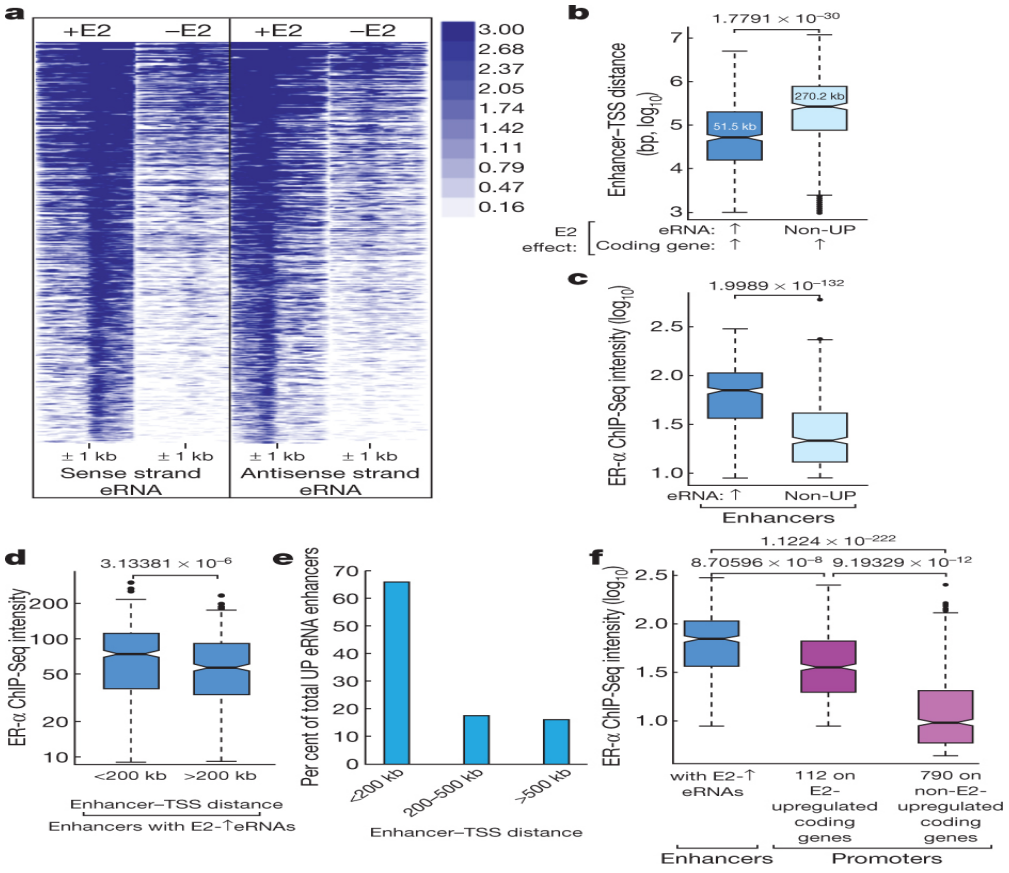


Figure 7 (Chapter 3): E2 induction of eRNA in breast cancer cells.
A. Heatmap of GRO-seq showing bidirectional eRNA transcription at enhancers induced by E2. B. UP enhancers are closer to the UP genes (median ~52kb) in comparison to enhancers with non-upregulated eRNAs (median ~270kb). Up arrows indicate upregulation. C. ER-α binds more robustly to UP enhancers than to the enhancers with non-upregulated eRNA. D. Among the UP enhancers, the proximal ones within 200 kb from any E2-upregulated gene TSSs exhibit higher ER-α binding intensity than the distal cohort of UP enhancers located farther away. E. Most of the UP enhancers are in close proximity to E2-upregulated coding genes. F. ER-α binding intensity on UP enhancers is higher than on 112 promoters of E2-activated genes, which itself is higher than the 790 ER-α-bound promoters of genes did not show upregulation by E2. A log10 scale is used for panels b, c and f. P values are given at the top of graphs, and were calculated using the Student's t-test.

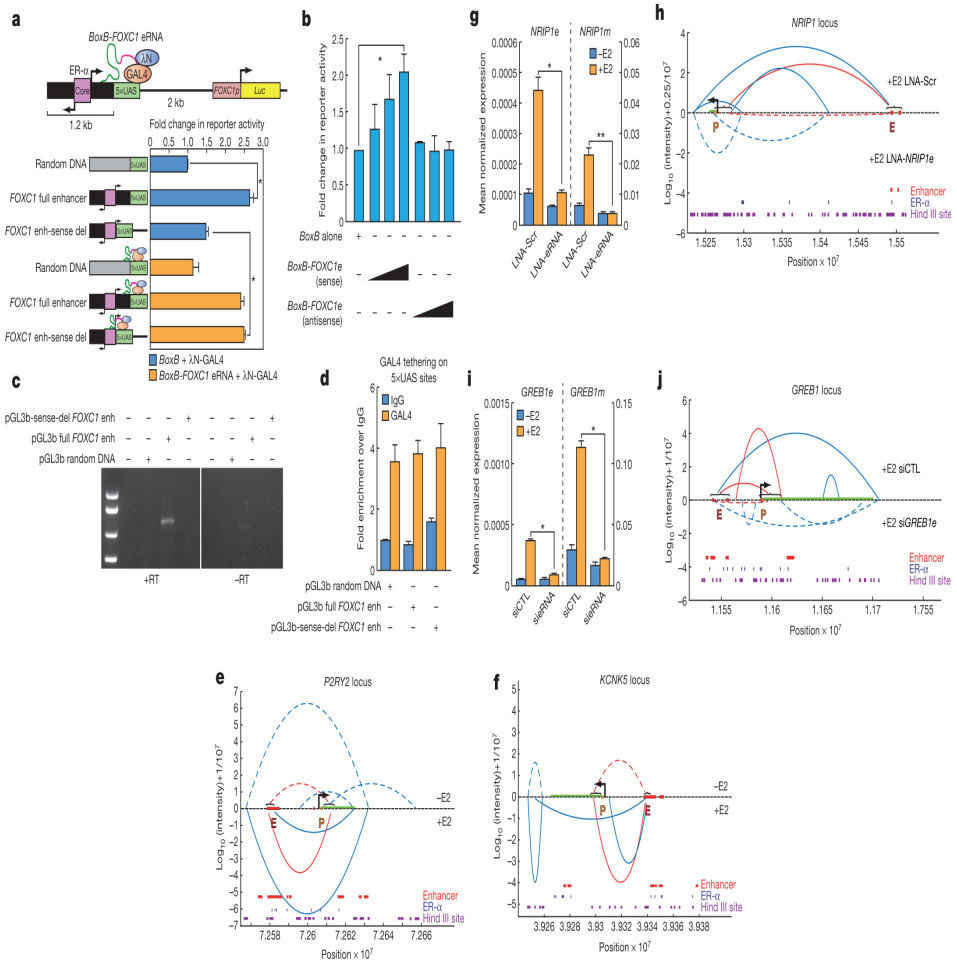


Figure 8 (Chapter 3): Ligand-induced eRNA is functionally important.
A. Schematic diagram of the BoxB-λN tethering system on the FOXC1 enhancer, which is upstream of a FOXC1 native-promoter-linked luciferase (Luc). B. eRNA function is sequence specific. C. Gel picture showing plasmid-based eRNA expression from full-length enhancer (enh) but not from the sense-eRNA-deleted enhancer construct (pGL3b-sense-del FOXC1 enh). RT, reverse transcriptase. D. Bar graph showing efficiency of GAL4 tethering on various pGL3b constructs. E. 3D-DSL data for the P2RY2locus, revealing strengthened promoter–enhancer interactions over basal conditions after 1 h E2 treatment. F. 3D-DSL data for the KCNK5 locus after 1 h E2 treatment. G, LNA knockdown of NRIP1eRNA effectively reduced the levels of both eRNA and associated coding gene transcripts. H. 3D-DSL data demonstrating significant reduction in promoter–enhancer interaction upon treatment of LNA against NRIP1 eRNA. I. GREB1e siRNA knockdown diminished the levels of eRNA and associated coding gene transcript. CTL, control. J. 3D-DSL data for the GREB1 locus showing significantly reduced enhancer–promoter looping as well as other genomic interactions afterGREB1e-specific siRNA treatment. Dotted lines in panels e and f represent −E2 condition, but knockdown situation in panels h and j.

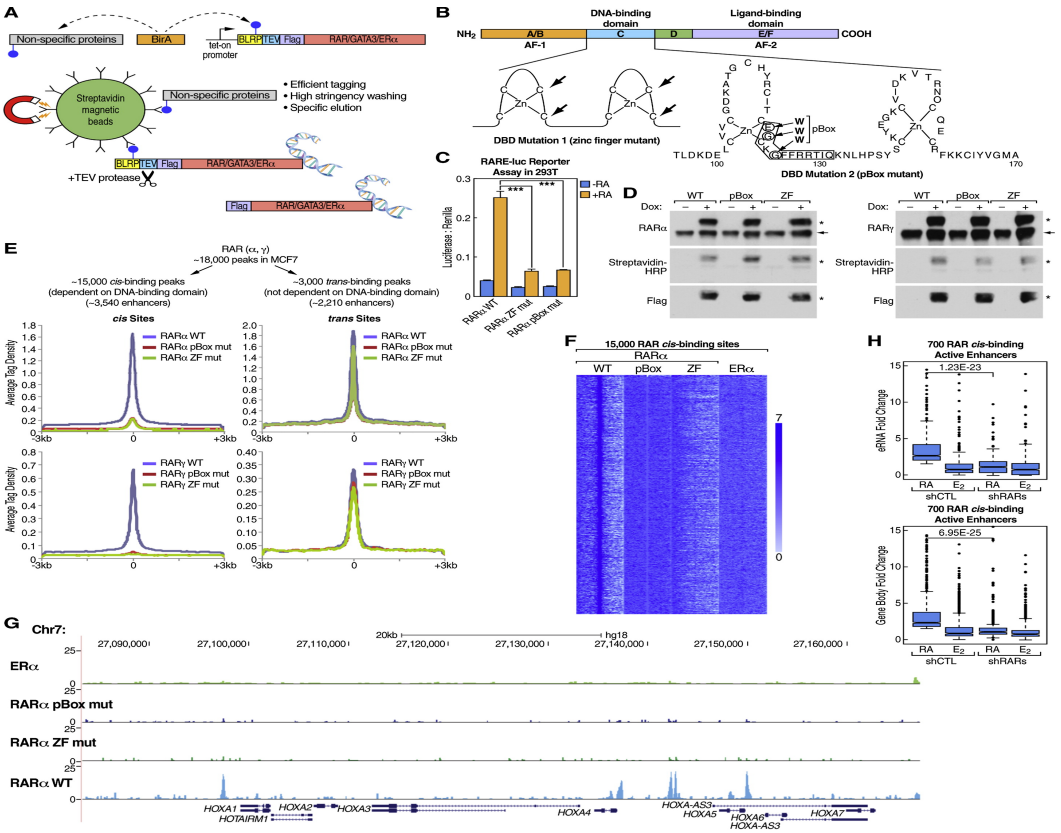


Figure 9 (Chapter 4): RARs Have cis- and trans-Binding Functional Models in MCF7.

(A) Schematic diagram depicting the in vivo biotinylation BirA-BLRP system, in which BirA biotin ligase can biotinylate a lysine residue in the BLRP tag, and BLRP-tagged proteins but not nonspecific biotin labeled proteins can be eluted by TEV protease digestion. (B) Diagram of two different DNA-binding domain mutations for RARα and RARγ. Arrows point out the mutation sites. (C) Both zinc finger and pBox mutations abolish RAR ability to activate RARE luciferase reporter. Mean ± SEM based on three independent experiments (***$p < 0.001$). (D) Western blots showing that the doxycycline-induced expression levels of BLRP-tagged RARα/γ (marked by asterisk) are similar to their endogenous protein (marked by arrow) levels. The blots with streptavidin-HRP show the in vivo biotinylation levels for tagged-RARα/γ. (E) cis and trans binding sites of RAR in MCF7 cells. The average tag density plots show the differential binding of wild-type and mutant RAR on these sites. (F) ChIP-seq heatmap of RARα and ERα for 15,000 RAR cis-binding sites, showing no binding of ERα at these sites. (G) UCSC browser snapshot of ChIP-seq showing the cis-binding of RARα to the HoxA cluster depends on its DNA-binding ability. (H) GRO-seq boxplots showing that RARs are required for RA-liganded activation of both eRNA and gene body transcription.

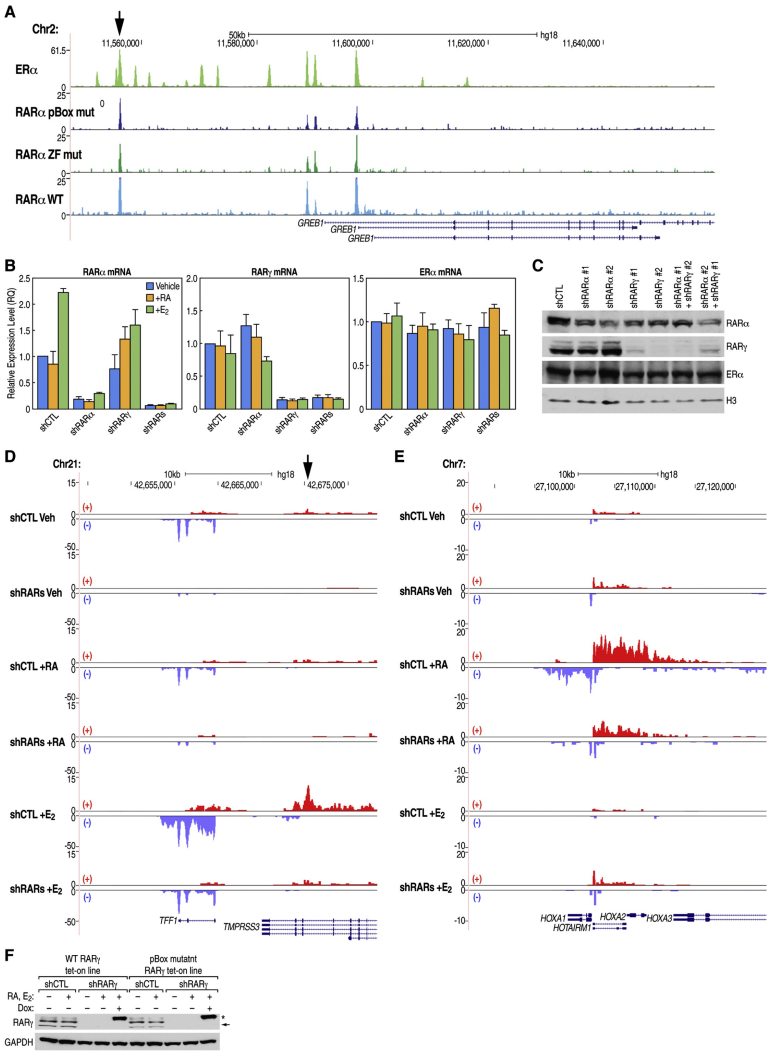


Figure 10 (Chapter 4): Both cis- and trans-Binding of RARs Regulates Ligand-Dependent Transcription Activation.

(A) UCSC browser snapshot of ChIP-seq showing the trans-binding of RARα at the GREB1 locus does not depend on its DNA-binding ability. (B) qPCR results show robust shRAR knockdown efficiency and that ERα RNA levels are not affected by knockdown of RAR. Data are represented as mean ± SEM based on three independent experiments. (C) Western blots show the protein knockdown efficiency of two different shRNA constructs for both RARα/γ. Knockdown of RARα/γ does not affect ERα protein level. Histone H3 was used as loading control. (D) UCSC browser image of GRO-seq showing knockdown of RARs affects E2-induced activation of TFF1 gene. The arrow points to an ERα active enhancer, which also has lower E2 induction of eRNA upon knockdown of RARs. (E) UCSC browser snapshot of GRO-seq showing knockdown of RARs affects RA-induced activation of HoxA1 gene. (F) A Tet-On inducible system was used to express either wild-type or pBox mutant RARγ in rescue experiments.

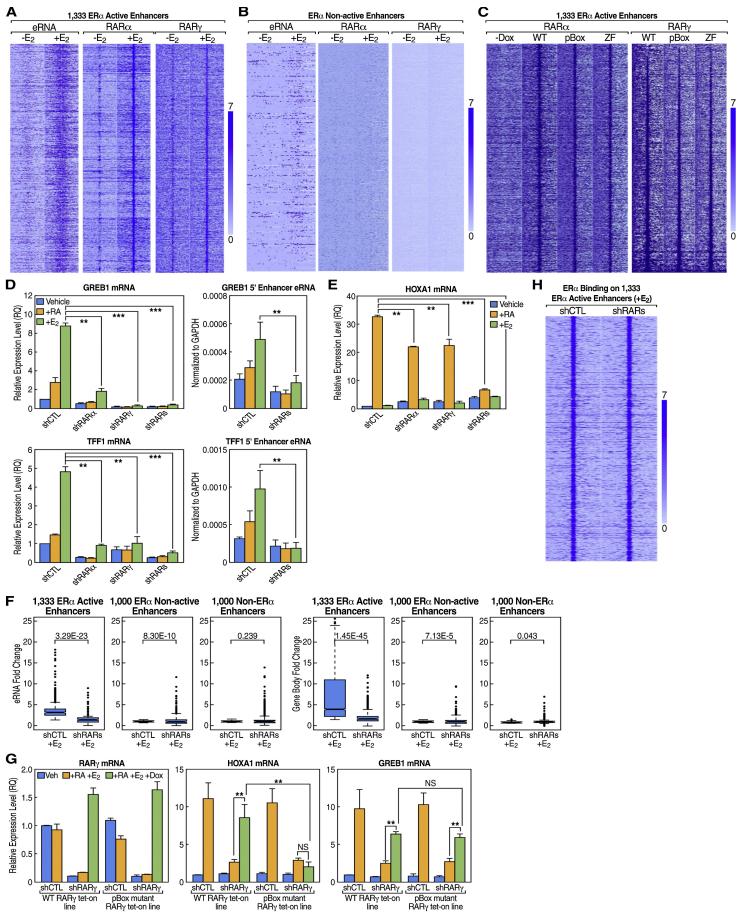
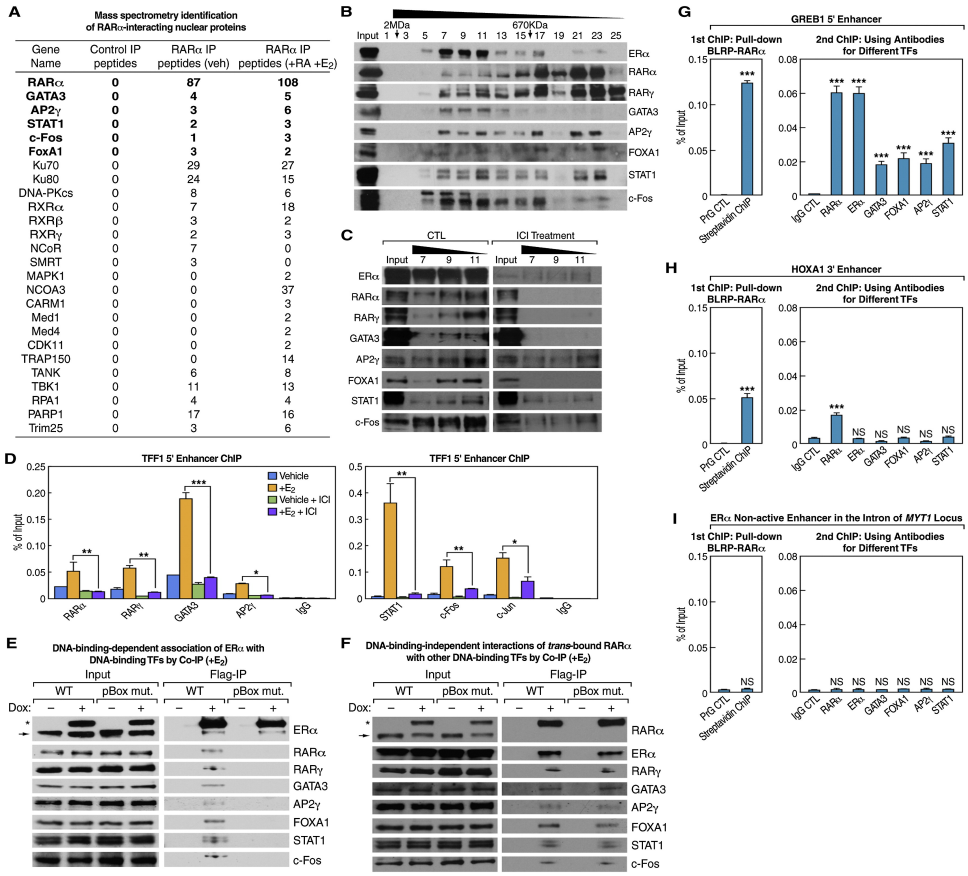


Figure 11 (Chapter 4): trans-Bound RARs on ERα Active Enhancers Regulate E2-Liganded Transcription Activation.

(A) Heatmaps of GRO-seq and ChIP-seq data (±E2) for 1,333 ERα active enhancers showing strong E2-induced eRNA transcription and E2-enhanced binding of both RARα and RARγ, respectively. (B) Heatmaps of GRO-seq and ChIP-seq for a control group of ERα non-active enhancers exhibiting no RARα/γ binding and no significant E2-induced eRNA transcription. (C) For the 1,333 ERα active enhancers, heatmaps of ChIP-seq data for the wild-type and two DNA-binding mutants of RARα/γ (+RA and E2) show that their association with these enhancers is DNA binding independent. (D) Knockdown of either RARα or RARγ by shRNA inhibits ERα target gene induction by E2, as demonstrated by qPCR analysis. (E) Knockdown of either RARα or RARγ using shRNA inhibits expression of the RAR cis-binding target HoxA1 gene in response to RA, as shown by qPCR analysis. (F) RARs are required for the E2-liganded activation of ERα active enhancers and their targets, as shown by GRO-seq boxplots. (G) The pBox mutant RARγ fails to rescue expression of its cis-binding target HoxA1 after knockdown of endogenous RARγ. (H) Heatmap showing that knockdown of RARs does not affect ERα binding at the 1,333 active enhancers.



Gene Name	Control IP peptides	RARα IP peptides (veh)	RARα IP peptides (+RA +E2)
RARα	**0**	**87**	**108**
GATA3	**0**	**4**	**5**
AP2γ	**0**	**3**	**6**
STAT1	**0**	**2**	**3**
c-Fos	**0**	**1**	**3**
FoxA1	**0**	**3**	**2**
Ku70	0	29	27
Ku80	0	24	15
DNA-PKcs	0	8	6
RXRα	0	7	18
RXRβ	0	3	2
RXRγ	0	2	3
NCoR	0	7	0
SMRT	0	3	0
MAPK1	0	0	2
NCOA3	0	0	37
CARM1	0	0	3
Med1	0	0	2
Med4	0	0	2
CDK11	0	0	2
TRAP150	0	0	14
TANK	0	6	8
TBK1	0	11	13
RPA1	0	4	4
PARP1	0	17	16
Trim25	0	3	6

Figure 12 (Chapter 4): ERα Interacts with a Mega Complex of DNA-Binding Transcription Factors at ERE-Containing Active Enhancers.

(A) RARα associates with several DNA-binding TFs, as shown by mass spectrometry analysis after pull down of biotin-tagged RARα and elution with TEV protease digestion. The same inducible BLRP-tagged RARα stable cell line without doxycycline induction was used as a control. (B) Western blots of gel filtration samples from MCF7 nuclear lysates (+E2) show various DNA-binding TFs associate with ERα in 1–2 MDa fractions. (C) Knockdown of ERα by ICI 182780 causes loss of the DNA-binding TFs in 1–2 MDa ERα-containing complex, as revealed by immunoblotting of gel filtration fractions from the 1–2 MDa range (fractions 7, 9, and 11). (D) DNA-binding TFs in the ERα complex bind to an ERα active enhancer at TFF1 locus upon E2 signal, and knockdown of ERα reduces their binding. ChIP signals are presented as percentage of input. (E) The interaction of ERα with other DNA-binding TFs is dependent on its DNA-binding ability, as shown by coimmunoprecipitation using BLRP-tagged WT or pBox mutant ERα. (F) The interaction of RARα with other DNA-binding TFs is independent of its DNA-binding ability, as demonstrated by coimmunoprecipitation of BLRP-tagged WT or pBox mutant RARα and other TFs. (G–I) ChIP-reChIP analysis confirms the co-binding of RARα, ERα, and other DNA-binding TFs on ERα active enhancers but not on the ERα non-active enhancers or RAR-bound HoxA1 enhancer.

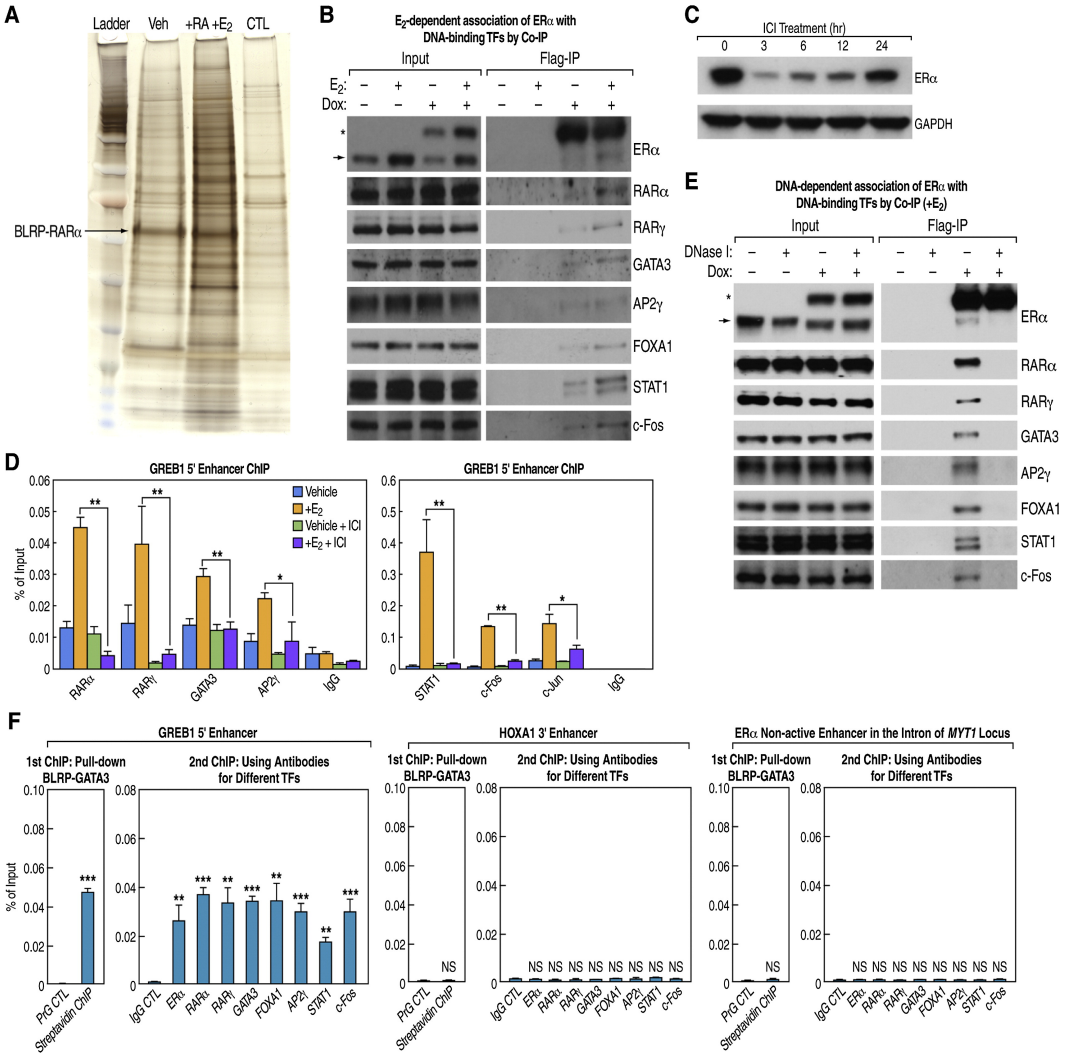


Figure 13 (Chapter 4): ERα Recruits In Situ a DNA-Binding TF Complex to ERE-Containing Active Enhancers.

(A) Silver staining of RARα complex to detect the pull-down of BLRP-RARα and its associated nuclear proteins. The same cell line without doxycycline treatment was used as a control to filter background proteins after mass spectrometry. (B) The E2-dependent interaction between ERα and other DNA-binding TFs was confirmed by coimmunoprecipitation of BLRP-tagged ERα and the other Tfs. (C) Test of ERα protein levels after ICI 182780 knockdown showing that a treatment of 3 hr yielded the best ERα protein knockdown. (D) DNA-binding TFs in the ERα complex bind to an ERα active enhancer located 5′ of GREB1 upon E2 signal and ERα knockdown reduces their binding. ChIP signals are presented as percentage of input. (E) The DNA-dependent interactions between ERα and other DNA-binding TFs were confirmed by co-immunoprecipitation of BLRP-tagged ERα and the other TFs. (F) ChIP-reChIP analyses were performed to confirm the co-binding of GATA3, ERα, and other DNA-binding TFs to ERα active enhancers but not HoxA1 enhancer or ERα non-active enhancers. ChIP signals are presented as percentage of input.

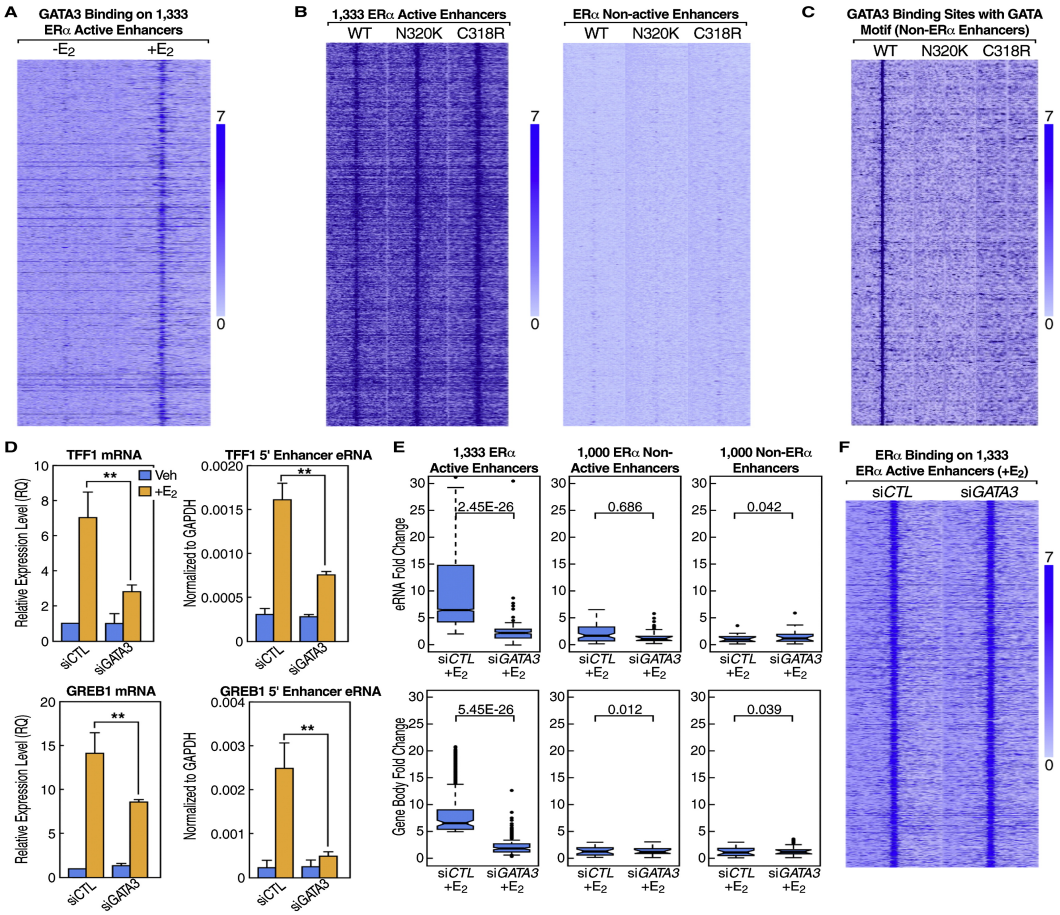


Figure 14 (Chapter 4): trans-Bound GATA3 on ERα Active Enhancers Regulates ERα E2-Liganded Transcription Activation.

(A) Heatmap displaying GATA3 binding at the 1,333 ERα active enhancers is enhanced by E2. (B) Heatmaps of ChIP-seq data for wild-type and two DNA-binding mutants of GATA3 (+E2) show the binding of GATA3 to these ERα active enhancers is not dependent on its DNA-binding ability. There is no binding of either wild-type or mutant GATA3 to ERα non-active enhancers. (C) Heatmap of ChIP-seq data for wild-type and two DNA-binding mutants of GATA3 (+E2) shows the binding of GATA3 to these non-ERα enhancers that contain the GATA motif requires its DNA-binding ability. (D) Knockdown of GATA3 affects ERα-dependent activation of eRNA transcription and coding gene expression for GREB1 and TFF1 genes. Mean ± SEM based on three independent qPCR experiments ($^{**}p < 0.01$). (E) GRO-seq boxplots showing that GATA3 is required for the E2-liganded activation of ERα active enhancers and their coding gene targets. (F) Heatmap showing that knockdown of GATA3 does not affect ERα binding at the 1,333 active enhancers.

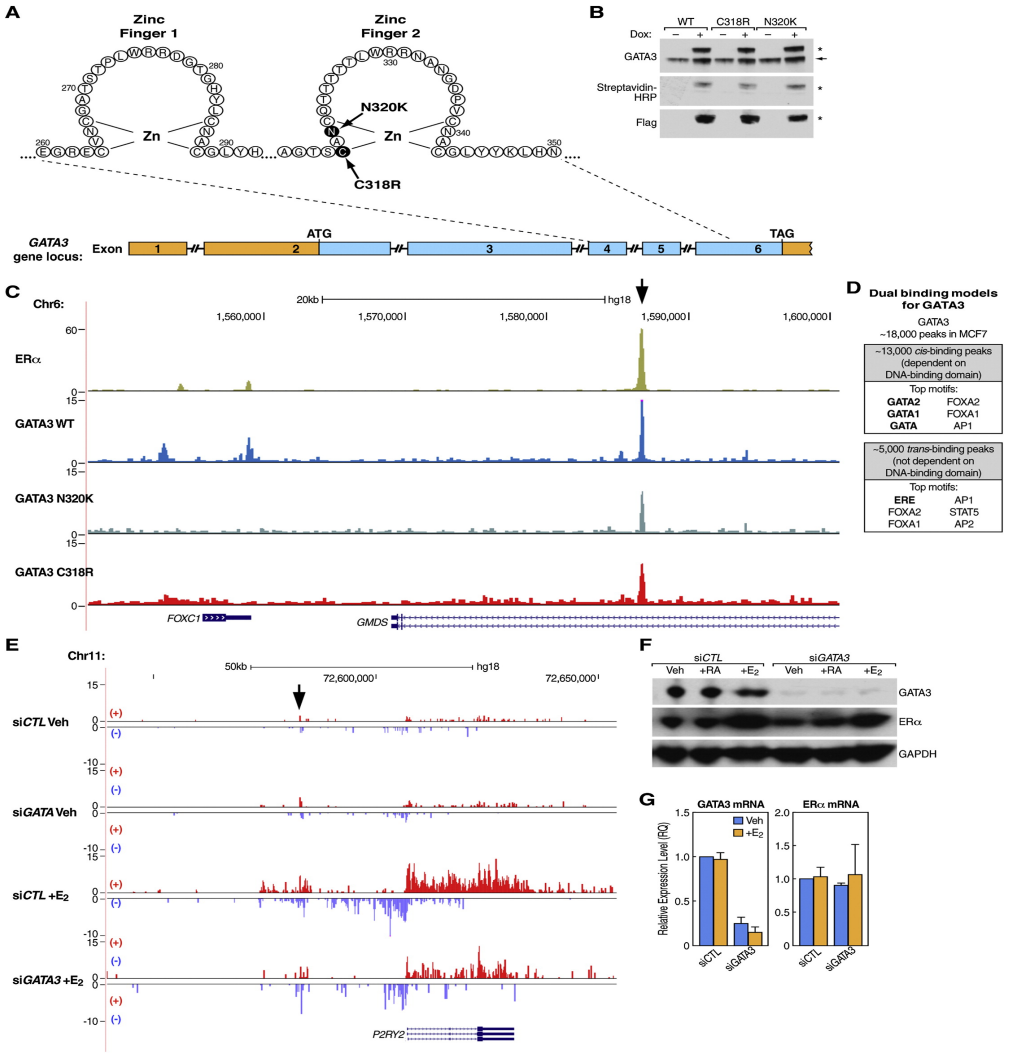


Figure 15 (Chapter 4): Dual-Binding Models for GATA3 and trans-bound GATA3 Also Regulates the Function of ERα Active Enhancers.

(A) Schematic diagram showing two different mutagenesis approaches for disrupting the GATA3 DNA-binding domain. (B) Western blots show the expression levels of BLRP-tagged GATA3 (marked by asterisk) are similar to their endogenous protein (marked by arrow) levels. (C) UCSC browser snapshot of ChIP-seq illustrating the cis-binding and trans-binding of GATA3 at the FoxC1 gene locus. In the snapshot, the trans-binding of GATA3 at one ERα active enhancer (marked by arrow) is evident. (D) The cis and trans dual binding models of GATA3 in MCF7. (E) UCSC browser snapshot of GRO-seq showing knockdown of GATA3 affects the E2-induced activation of P2RY2 gene at both the enhancer and the gene body. The arrow points to an ERα active enhancer. (F) Western blots confirm the effective siRNA knockdown of GATA3 and that its knockdown does not affect ERα protein level. GAPDH was used as a loading control. (G) qPCR results showing robust siGATA3 knockdown efficiency of GATA3 RNA levels and no concomitant change in ERα RNA level. Data are represented as mean ± SEM based on three independent experiments.

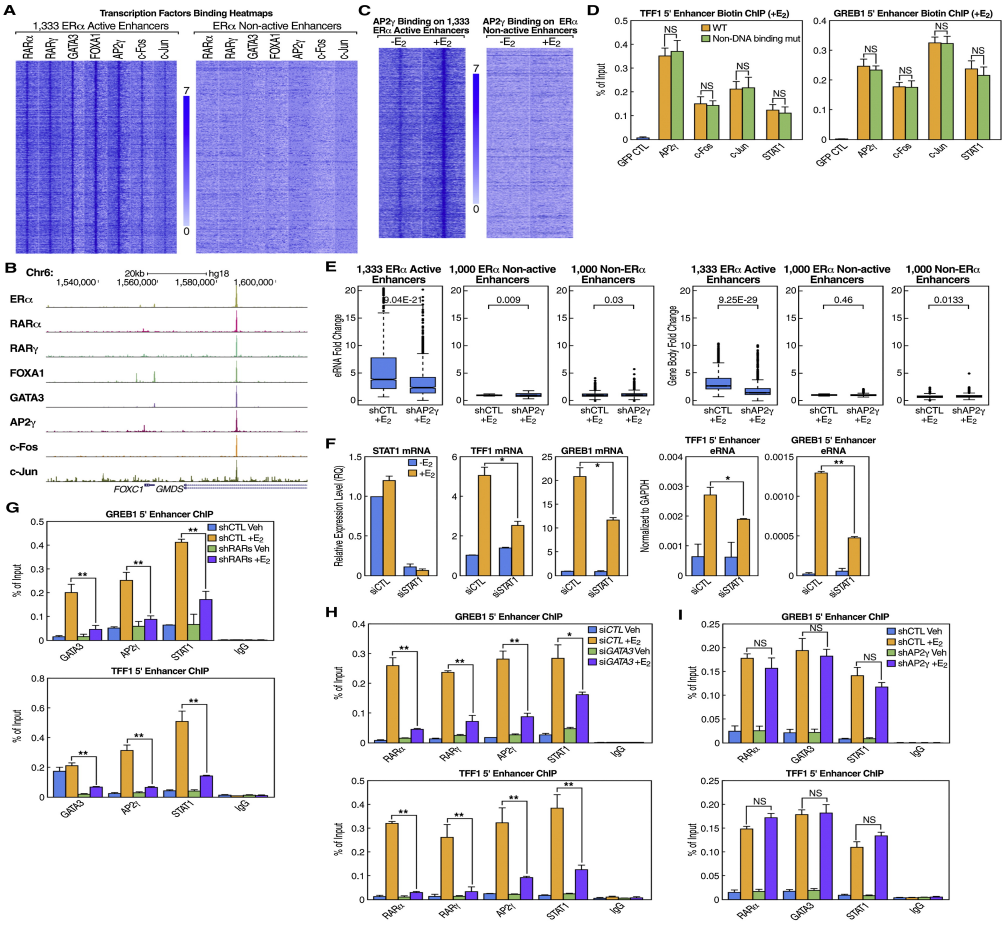


Figure 16 (Chapter 4): ERα Active Enhancers Are MegaTrans Enhancers Regulated by DNA-Binding TFs (part 1).

(A) Heatmaps of ChIP-seq data for different TFs (+E2) displaying strong binding of these DNA-binding TFs at the 1,333 ERα active enhancers but not at ERα non-active enhancers. (B) UCSC browser snapshot of an ERα active enhancer for FoxC1, which exemplifies a MegaTrans-bound enhancer (+E2). (C) Heatmap showing AP2γ binding at ERα active enhancers, but not at ERα non-active enhancers, in response to E2. (D) WT and non-DNA-binding mutants of MegaTrans TF components bind equivalently to two ERα active enhancers of TFF1 and GREB1, as demonstrated by biotin ChIP using BLRP-tagged TFs (GFP served as control). For details regarding DNA-binding domain mutagenesis see Extended Experimental Procedures. ChIP signals are presented as percentage of input. (E) GRO-seq boxplots showing that AP2γ is required for ligand-dependent activation of both eRNA and target gene body transcription for ERα active enhancers. (F) STAT1 is required for the activation of ERα active enhancers and coding gene expression by E2 for GREB1 and TFF1 genes, as demonstrated by knockdown and qPCR analysis. (G–I) Knockdown of RARs or GATA3, but not AP2γ, greatly reduces the E2-enhanced occupancy of DNA-binding TFs on ERα active enhancers of TFF1 and GREB1. ChIP signals are presented as percentage of input.

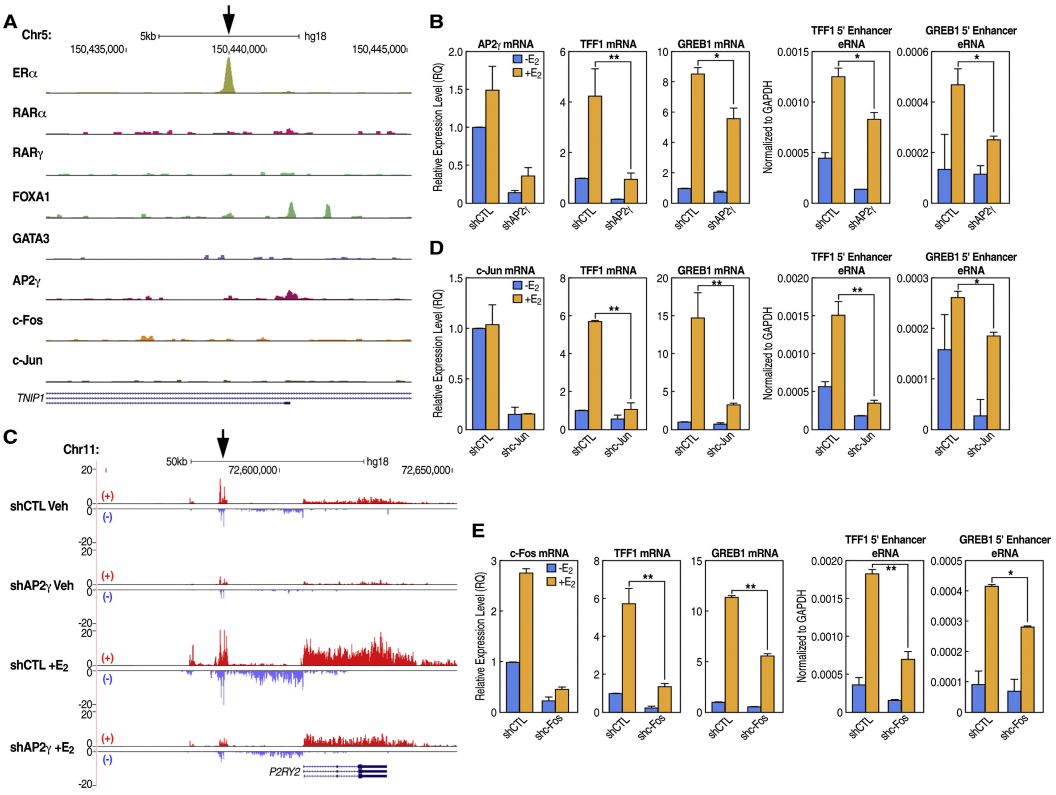


Figure 17 (Chapter 4): ERα Active Enhancers Are MegaTrans Enhancers Regulated by DNA-Binding TFs (part 2).

(A) UCSC browser snapshot of ChIP-seq showing an ERα non-active enhancer (pointed by arrow) at the TNIP1 locus without the binding of MegaTrans complex. (B) Knockdown of AP2γ by shRNA affects ligand-dependent activation of both eRNA transcription and coding gene expression for GREB1 and TFF1 genes. (C) UCSC browser image of GRO-seq showing knockdown of AP2γ affects the E2-induced activation of P2RY2 gene at both the enhancer and the gene body. The arrow points to an ERα active enhancer. (D-E) Knockdown of c-Jun or c-Fos by lentivirus shRNA affects E2-dependent ERα activation of the coding gene body and eRNA transcription for GREB1 and TFF1 genes, as demonstrated by qPCR.

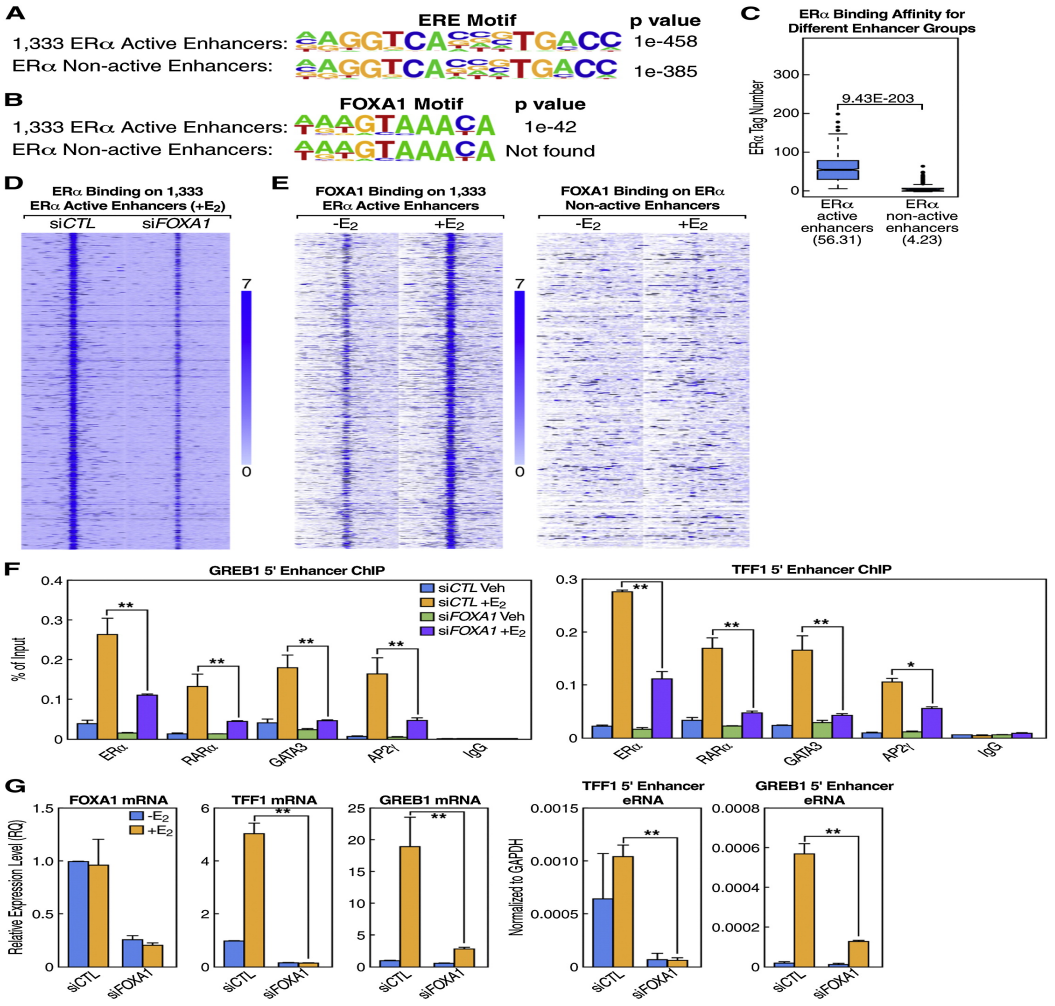


Figure 18 (Chapter 4): FoxA1 Performs Dual Roles on ERα Active Enhancers. (A and B) ERα and FoxA1 motif analyses using Homer program for 1,333 ERα active enhancers and ERα non-active enhancers (see Extended Experimental Procedures for analysis details). (C) Boxplot based on ERα ChIP-seq data (+E2) showing higher binding affinity of ERα at 1,333 ERα active enhancers than at ERα non-active enhancers. (D) Heatmap showing that knockdown of FoxA1 greatly reduces ERα binding at the 1,333 active enhancers.(E) Heatmap showing FoxA1 binding at 1,333 ERα active, but not at ERα non-active enhancers, is enhanced in response to E2. (F) Conventional ChIP assays for TFF1 and GREB1 enhancers showing knockdown of FoxA1 substantially reduced binding of ERα and the MegaTrans components following E2 treatment. ChIP signals are presented as percentage of input. (G) FoxA1 is required for the activation of ERα active enhancers in response to E2, as exemplified by the effects of FoxA1 knockdown on coding gene expression and eRNA transcription for GREB1 and TFF1 genes.

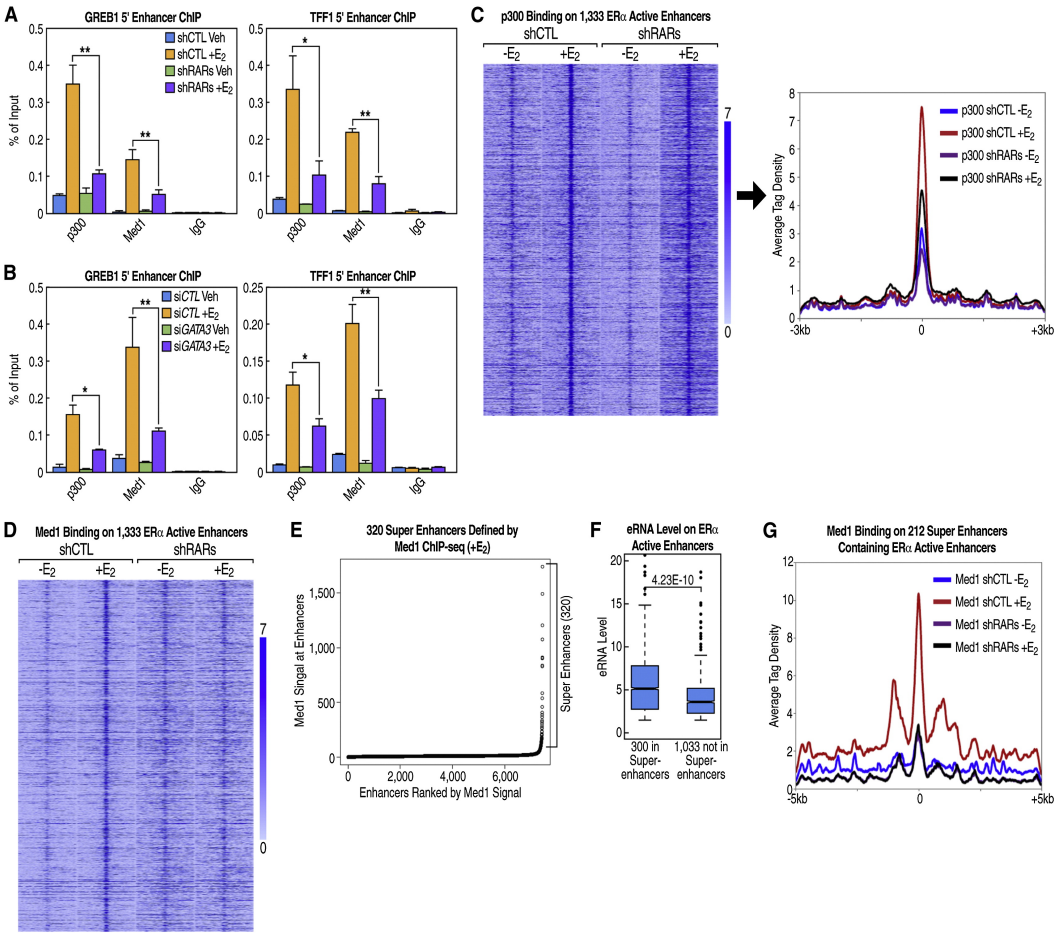


Figure 19 (Chapter 4): trans-Bound TFs on MegaTrans Enhancers Are Required for Recruitment of ERα Coactivators and Super-Enhancer Function (part 1).

(A and B) Knockdown of RARs or GATA3 greatly reduces the E2-enhanced binding of p300 and Med1 to ERα active enhancers. ChIP signals are presented as percentage of input. (C) Heatmap and tag density plot of p300 ChIP-seq data for four different conditions demonstrating that knockdown of RARs by shRNA reduces E2-enhanced p300 recruitment on 1,333 ERα active enhancers. (D) trans-bound RARs are required for E2-enhanced recruitment of the coactivator Med1 to ERα active enhancers, as shown by a heatmap of Med1 ChIP-seq data on 1,333 ERα active enhancers. (E) A Med1 tag density plot based on Med1 ChIP-seq (+E2) data and clustering of enhancers identifies ~320 super-enhancers in MCF7 cells (see Extended Experimental Procedures for analysis details). (F) A boxplot analysis based on GRO-seq data (+E2) of eRNA expression levels for two groups of ERα active enhancers: the 300 ERα active enhancers located in super-enhancers (median: 5.14) and 1,033 ERα active enhancers that are not located in super-enhancers (median: 3.59). (G) Tag density plot showing knockdown of trans-bound RARs, which affects the function of ERα active enhancers, reduces the E2-enhanced Med1 signal at 212 super-enhancers that contain ERα active enhancers.

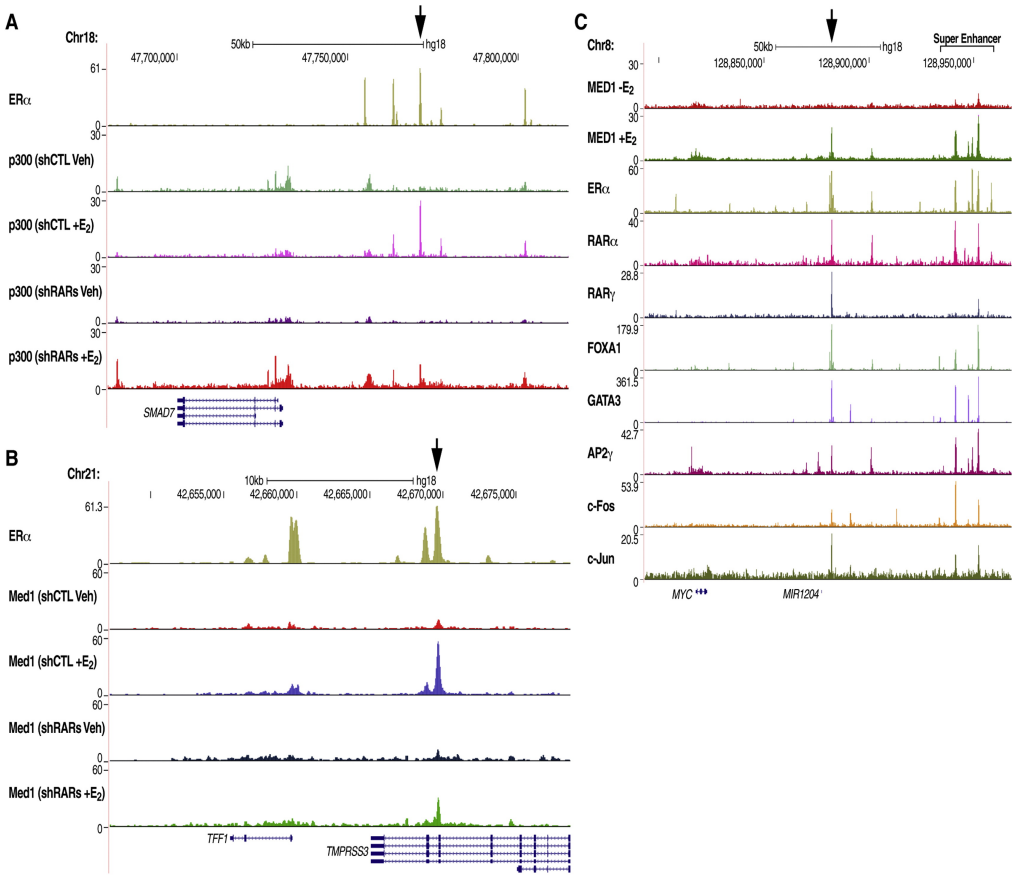


Figure 20 (Chapter 4): trans-Bound TFs on MegaTrans Enhancers Are Required for Recruitment of ERα Coactivators and Super-Enhancer Function (part 2).

(A) UCSC browser image of ChIP-seq showing the ERα active enhancer (marked by arrow) at the Smad7 locus exhibits a lower level of p300 recruitment upon E2 signal after RARα and RARγ shRNA knockdown. (B) UCSC browser snapshot of ChIP-seq showing the ERα active enhancer (marked by arrow) at the TFF1 locus has a lower level of E2-induced Med1 recruitment after RARα and RARγ shRNA knockdown. (C) UCSC browser image of ChIP-seq showing one regular ERα active enhancer (marked by arrow) and one super-enhancer containing several ERα active enhancers at the Myc gene locus. Both regions have stronger Med1 ChIP-seq signals upon E2 treatment, suggesting that a small percentage of ERα MegaTrans enhancers locate in super-enhancers and regulate their function.

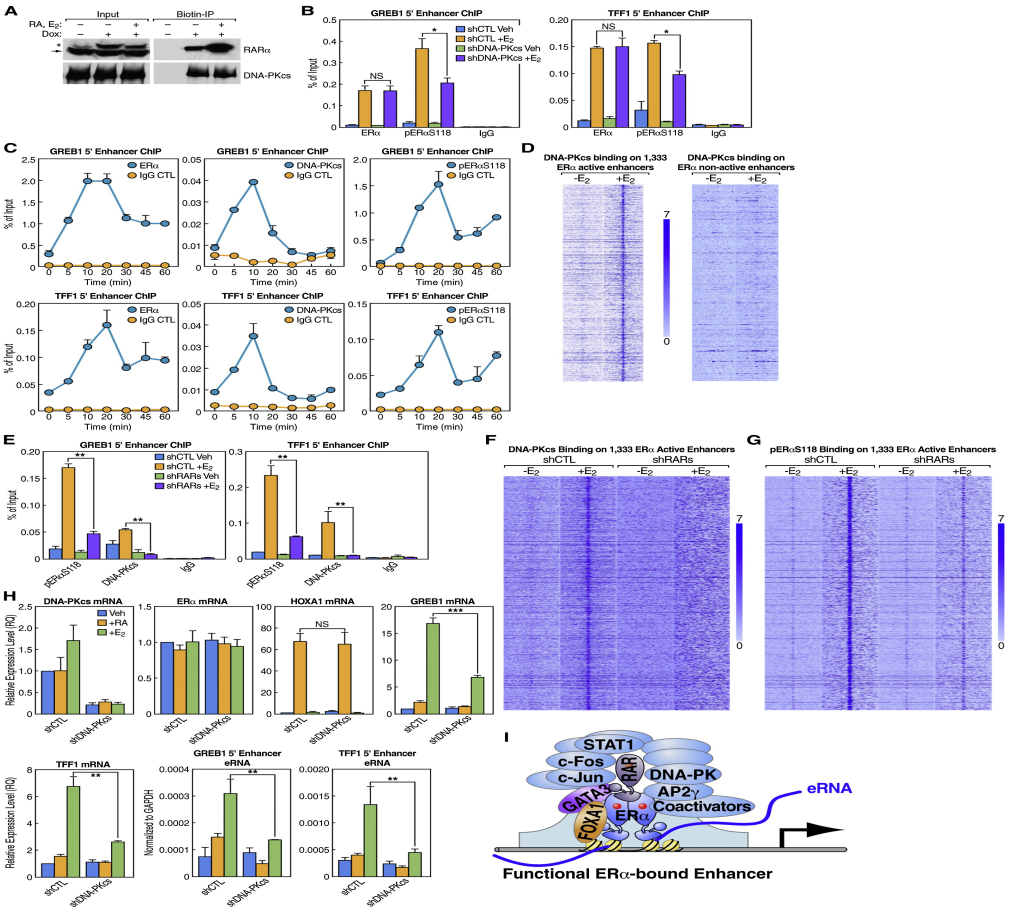


Figure 21 (Chapter 4): trans-Bound RAR May Contribute to Recruit DNA-PK Kinase as a Coactivator for ERα Active Enhancers (part 1).

(A) Western blots demonstrating interaction of doxycycline-induced BLRP-RARα protein with DNA-PKcs after pull down by streptavidin magnetic beads. (B) Conventional ChIP assays for TFF1 and GREB1 enhancers showing DNA-PKcs is not required for the occupancy of ERα but is partially required for the presence of S118-phosphorylated ERα (pERαS118). (C) The kinetics of ERα, DNA-PKcs, and pERαS118 occupancy at ERα active enhancers. ChIP signals are presented as percentage of input. (D) Heatmaps, based on ChIP-seq data, showing DNA-PKcs binding at 1,333 ERα active enhancers is enhanced by E2, while its binding is not apparent at ERα non-active enhancers. (E) Knockdown of RARs by shRNA greatly reduces DNA-PKcs binding to ERα active enhancers and affects enrichment of pERαS118. (F) Heatmap of DNA-PKcs ChIP-seq data showing loss of E2-enhanced DNA-PKcs binding to ERα active enhancers upon knockdown of both RARs. (G) Heatmap of pERαS118 ChIP-seq data demonstrating partial reduction of E2-enhanced pERαS118 binding to ERα active enhancers upon knockdown of both RARs. (H) Knockdown of DNA-PKcs by shRNA affects E2-liganded activation of gene body and eRNA transcription for GREB1 and TFF1 genes but does not affect ERα levels or RA induction of the HoxA1 gene, as demonstrated by qPCR. (I) Working model of a MegaTrans enhancer.

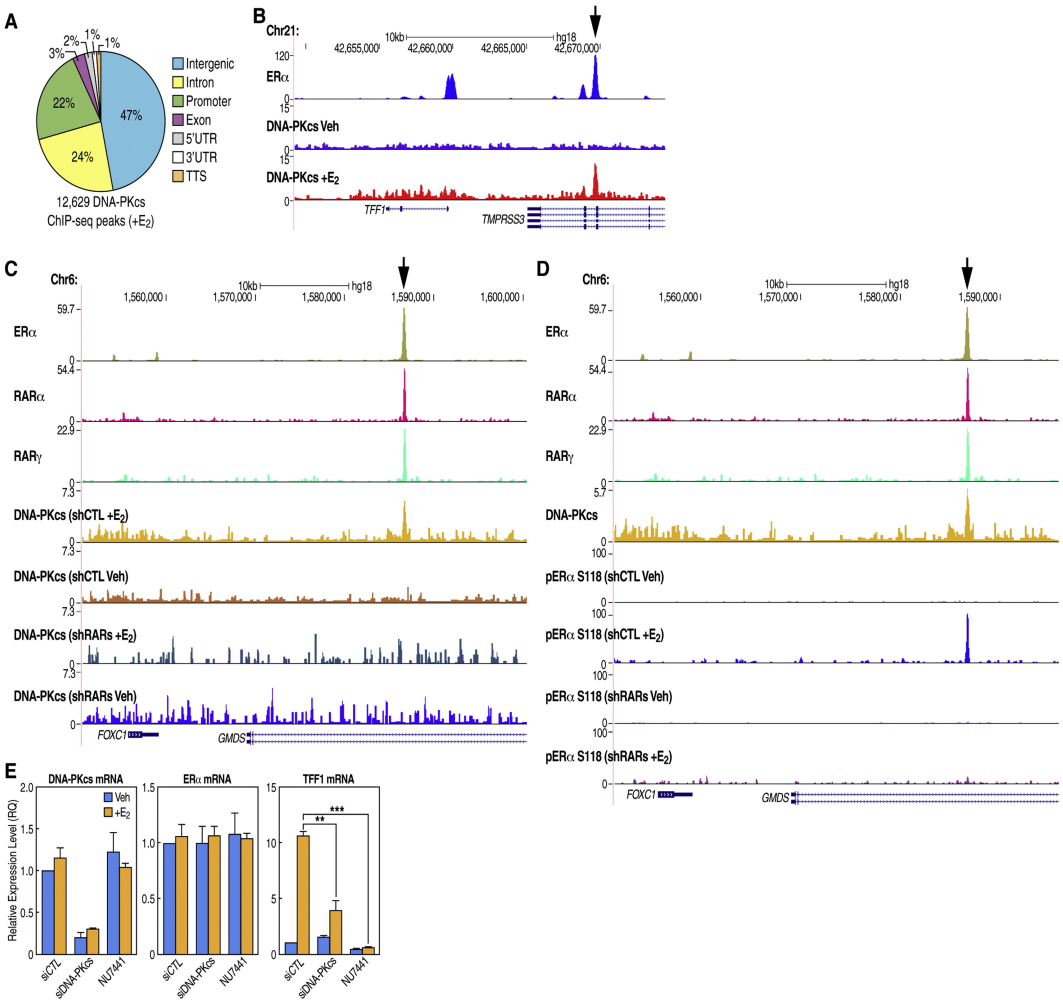


Figure 22 (Chapter 4): trans-Bound RAR May Contribute to Recruit DNA-PK Kinase as a Coactivator for ERα Active Enhancers (part 2).

(A) The distribution of 12,629 DNA-PKcs ChIP-seq (+E2) peaks. (B) UCSC browser snapshot of DNA-PKcs ChIP-seq data (-E2 and +E2) showing E2-enhanced binding of DNA-PKcs to an ERα active enhancer (pointed by arrow) at the TFF1 locus. (C) UCSC browser snapshot of ChIP-seq showing the ERα active enhancer (marked by arrow) at the TFF1 locus loses the recruitment of DNA-PKcs upon E2 stimulation after RARα and RARγ knockdown by shRNAs. (D) UCSC browser snapshot of ChIP-seq showing the ERα active enhancer (marked by arrow) at the TFF1 locus has a lower level of S118-phosphorylated ERα upon E2 stimulation after RARα and RARγ shRNA knockdown. (E) Knockdown of DNA-PKcs by siRNA or inhibition of DNA-PKcs kinase activity using NU7441 affects the activation of TFF1 by E2. Knockdown of DNA-Pkcs did not affect ERα levels. Data are represented mean ± SEM based on three independent qPCR experiments.

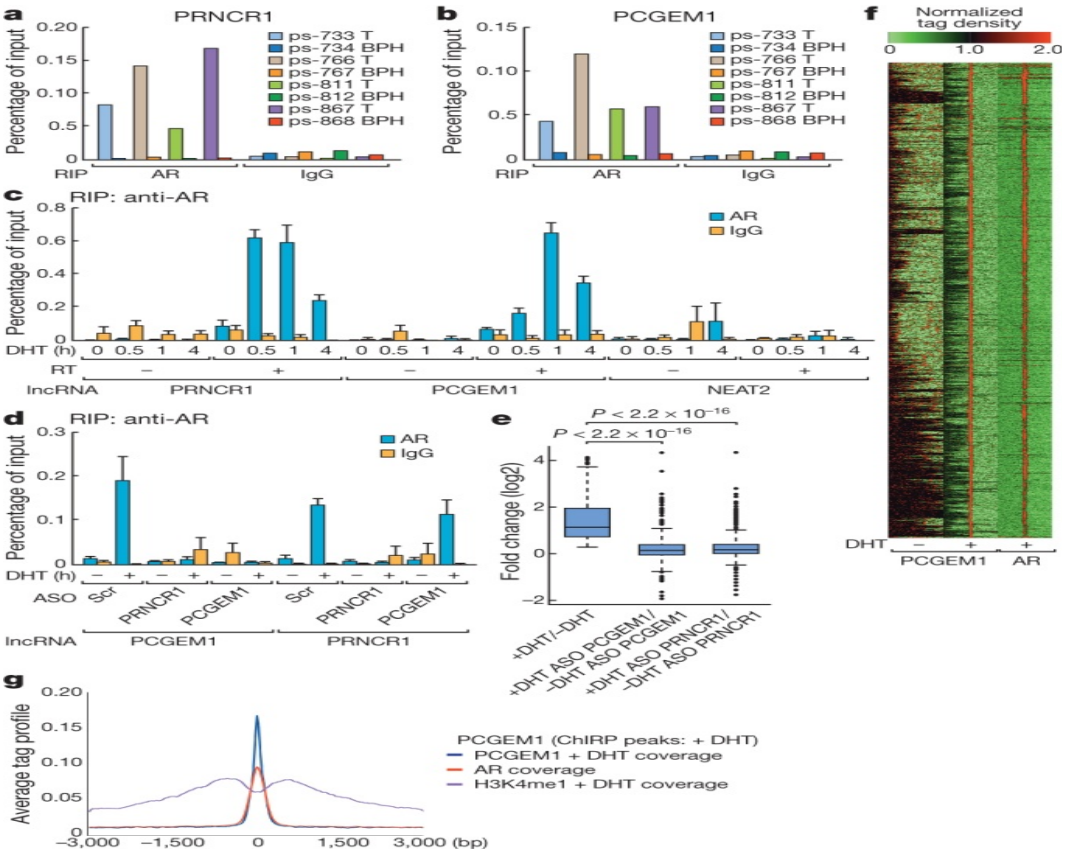


Figure 23 (Chapter 5): Signal-Dependent Interaction between AR and Prostate Specific LncRNAs.

(A and B) RIP assay performed in paired benign prostatic hyperplasia (BPH) and prostate tumour (T) tissues. (C) RIP assay in DHT-treated LNCaP cells (100 nM) at the indicated time points. (D) RIP assay in LNCaP cells transfected with indicated ASO followed by DHT treatment (100nM). Scr, scrambled control. (E) Global changes in DHT-induced AR target genes in PCGEM1 or PRNCR1 depleted LNCaP cells. (F) Heatmap showing the distribution of PCGEM1 and AR binding sites in DHT-stimulated LNCaP cells. (G) Average tag profile analysis of the aligned 2,142 PCGEM1 ChIRP peaks. For (C) and (D)(n= 3).

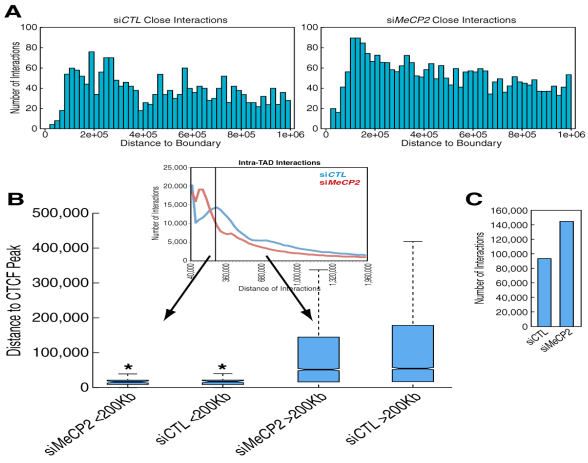


Figure 24 (Chapter 6): Distribution of interactions within TADs.
(A) Histogram showing the distribution of short-range interactions (<200kb) relative to the boundary in siCtrl (left) and siMeCP2 (right). (B) Distance of short-range intra-TAD interactions (<200kb), which are increase upon MeCP2 knockdown, relative to intra-TAD long-range interactions (>200kb) to CTCF binding sites. For each pair of interaction sites, the closest CTCF site to either interaction site was chosen for analysis. (C) Number of intra-TAD interactions that span across MeCP2 and CTCF co-bound sites in control and MeCP2 knockdown conditions. (D) Analysis of the distribution of all interaction distances within TADs (intra-TAD) comparing “shifted” (left panel) vs. “non-shifted” (right panel) boundary containing TADs.

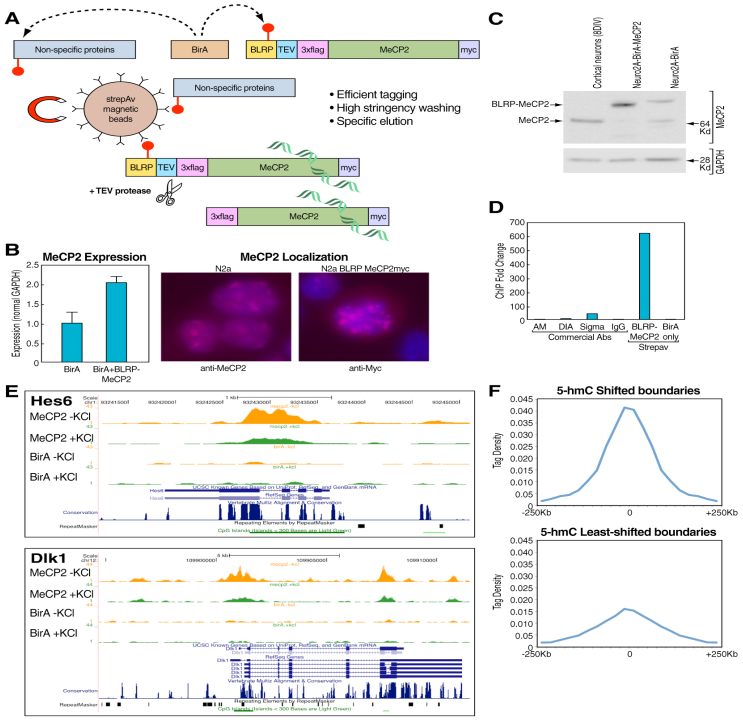


Figure 25 (Chapter 6): Biotin tagging system for genome wide ChIP-seq in mouse neuroblastoma cells.

(A) Biotin ligand recognition peptide (BLRP) tagging of MeCP2. The biotin tagging system enables endogenous biotinilation of MeCP2 by co-expressing the bacterial BirA protein. Specific elution of MeCP2 bound DNA is ensured by elution with TEV protease. (B) MeCP2 mRNA expression levels (left) and immunostaining showing similar intracellular localization of BLRP tagged MeCP2 relative to endogenous MeCP2 (right). (C) Comparison of MeCP2 expression levels of Neuro2a cell lines expressing biotin tagged MeCP2 to those of cortical neurons. Western blots showing the relative expression of MeCP2 in cells extracts from parental Neuro2a expressing the BirA ligase only (Neuro2a-BirA) stable expressing Biotin- MeCP2 and e17.5 cortical neurons cultured for 8 days in vitro. (D) Comparison between commercial MeCP2 antibodies to biotin-MeCP2 ChIP on mouse BDNF promoter. (AM – Active Motif, anti-MeCP2 DIA – Diagenode anti-meCP2, Sigma anti MeCP2). (E) Examples of genome wide binding of MeCP2. UCSC browser examples of MeCP2 binding compared to the background BirA expressing cells, with and without 3h KCl depolarization.

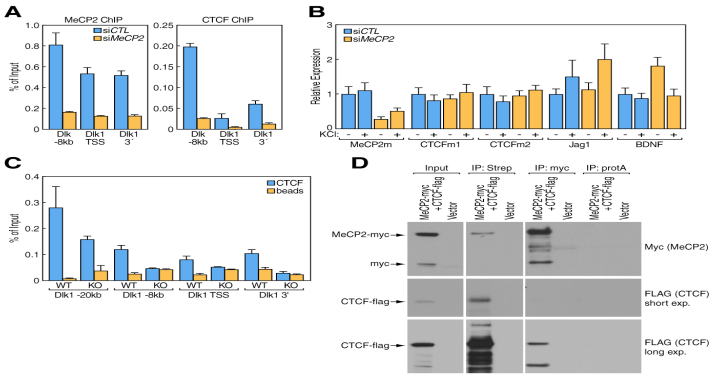


Figure 26 (Chapter 6): ChIP on select MeCP2 binding regions show loss of CTCF binding upon MeCP2 knockdown and co-immunoprecipitation.

(A) MeCP2 and CTCF ChIP on select co-bound sites near Dlk1 following MeCP2 knockdown reveal reduced of CTCF binding. (B) qRT-PCR shows that while MeCP2 is efficiently knocked down after siRNA treatment, there is no change in CTCF mRNA levels. Two different primers for CTCF are shown, as well as a few of MeCP2 target genes that are induced upon knock down. (C) CTCF ChIP on hippocampal tissue from wild type and MeCP2 knock out mice show reduced CTCF binding. (D) Co-IP of MecP2 and CTCF co-expressed in Neuro2a cells. Western blots showing co-immunoprecipitation of MeCP2 and CTCF in Neuro2a transfected with myc tagged MeCP2 and flag tagged CTCF.

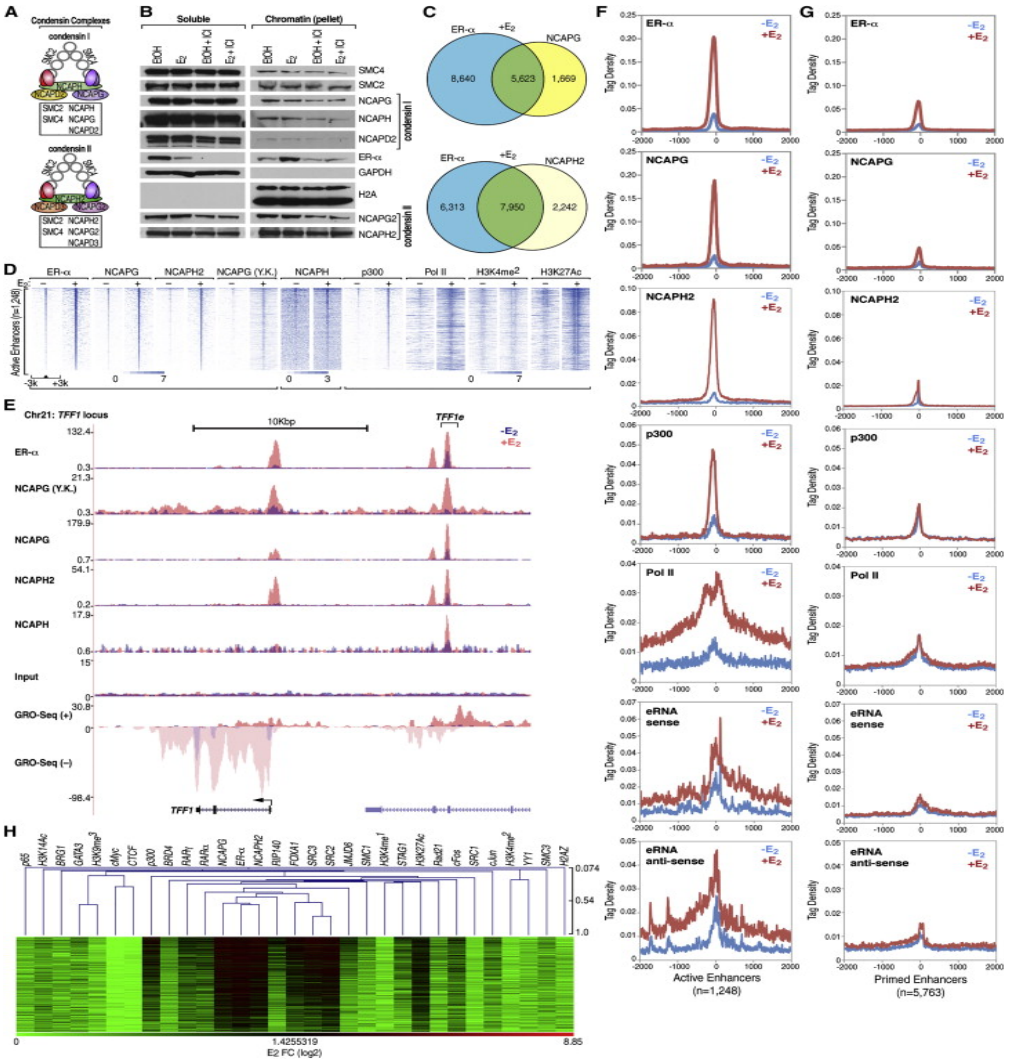


Figure 27 (Chapter 7): Estrogen-Induced Loading of Condensins to ER-α-Bound Active Enhancers.

(A) A cartoon diagram showing the subunit constituents of the condensin I and condensin II complexes. (B) Chromatin fractionation followed by western blots showing the localization of condensin subunits in MCF-7 cells upon E_2 or ICI treatment. (C) Venn diagram showing the genome-wide ChIP-seq peak numbers of NCAPG and NCAPH2 and their overlap with that of ER-α in E_2-treated MCF-7 cells. (D) Heatmaps showing ChIP-seq data of condensin I and condensin II together with p300, RNA Pol II, and active histone marks $H3K4me^2$ and H3K27Ac on active enhancers. (E) A snapshot of the UCSC genome browser (hg18) showing the ChIP-seq tracks of condensin subunits, ER-α, input control, and GRO-seq in TFF1 locus. (F and G) Profile plots showing normalized ChIP-seq or GRO-seq tag intensities. (H) Hierarchical cluster analysis showing the correlation between the E_2-induced recruitment of the interrogated transcription factors and histone modifications on the 1,248 active enhancers.

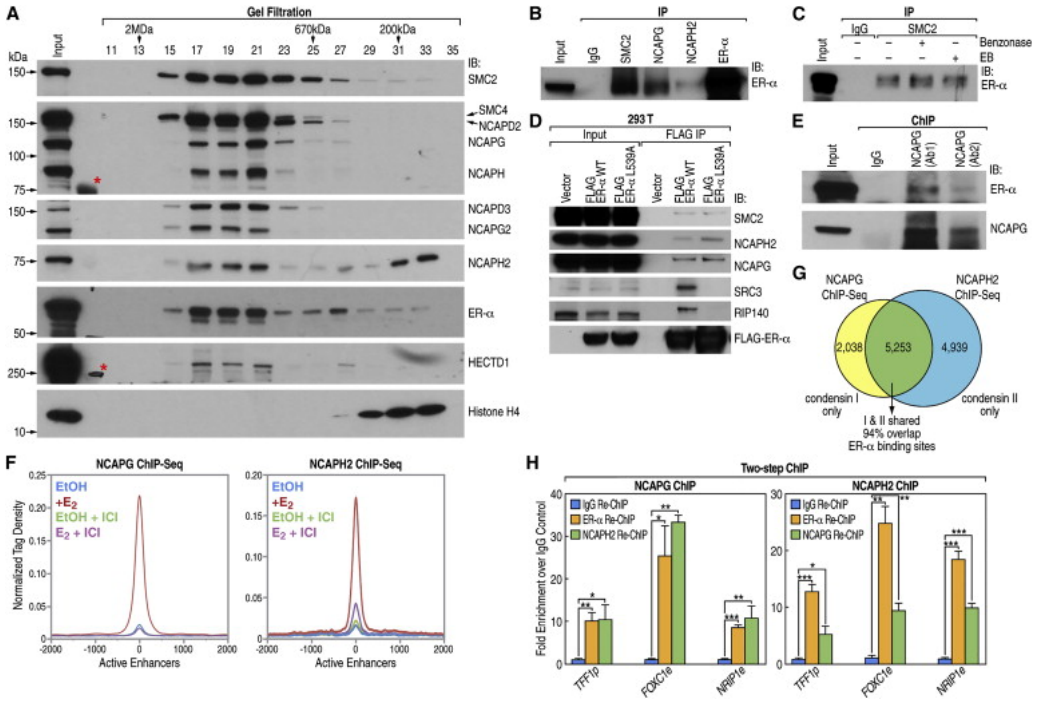


Figure 28 (Chapter 7): ER-α Interacts with Condensins.

(A) Gel filtration of E_2-treated MCF-7 nuclear extracts followed by western blots showing the elution profiles of target proteins as indicated, with histone H4 as a control. Red asterisks denote irrelevant bands. (B) CoIP in the whole-cell lysate followed by western blots showing that endogenous condensin subunits co-precipitated with ER-α (without benzonase). (C) The presence of benzonase or ethidium bromide (EB) did not cause detectable change of condensin/ER-α interaction.(D) CoIP followed by western blots showing that overexpressed FLAG-tagged ER-α pulls down condensin subunits (benzonase added). The L539A mutant of ER-α interacts with condensin subunits, but not with canonical coregulators SRC3 and RIP140. WT: wild type. (E) ChIP-western data showing that two antibodies against NCAPG pull down ER-α. (F) ChIP-seq profile plots (centered on ER-α binding peaks in $+E_2$ situation) showing the binding to active enhancers of both condensin I (NCAPG) and condensin II (NCAPH2) in presence of E_2 or ICI treatments. (G) Venn diagram showing the genome-wide overlap of ChIP-seq peaks of condensin I (i.e., NCAPG) and condensin II (i.e., NCAPH2) in E_2-liganded MCF-7 cells. (H) Two-step ChIP-qPCR results are shown using antibodies against condensin I and II (NCAPG and NCAPH2) and ER-α in liganded cells; experiment was repeated three times; “p” and “e” after gene names denote promoter and enhancer, respectively. Data are presented as mean ± SD. (two-tailed Student’s t test). IP/coIP experiments were performed in MCF-7 cells unless otherwise indicated.

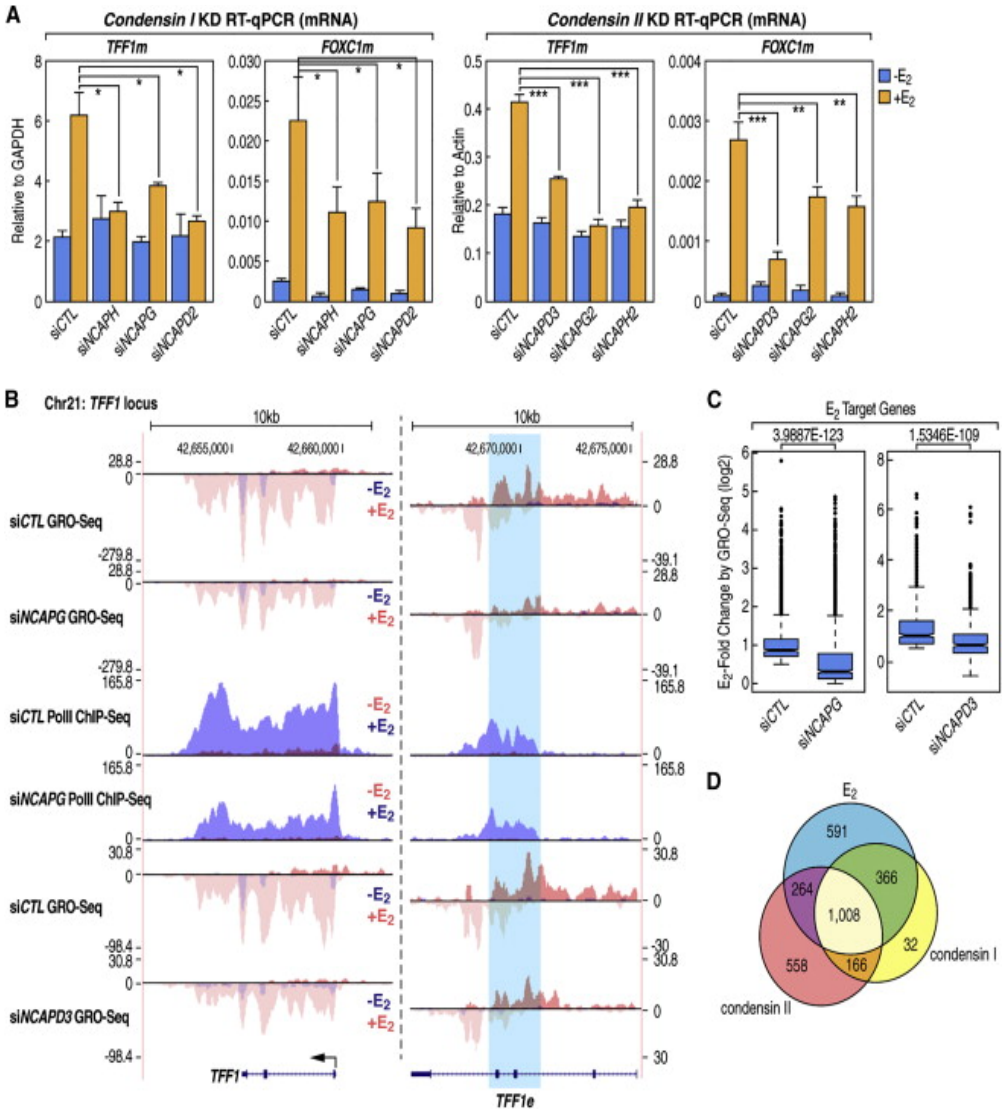


Figure 29 (Chapter 7): Condensin I and Condensin II Control ER-α-Regulated Gene Activation in a Partially Overlapping Manner.

(A) RT-qPCR data showing the expression of TFF1 and FOXC1 mRNA levels in the wild-type or in cells with condensin I or condensin II knockdown. (B) Genome browser image showing the results of GRO-seq and Pol II ChIP-seq at TFF1 locus in the presence of condensin knockdown (siNCAPG or siNCAPD3) versus siCTL transfected cells. (TFF1e, highlighted area). (C) Boxplots showing the E_2-induced fold changes (E_2-FC) of all the E_2-upregulated coding genes in thesiCTL group and siNCAPG or siNCAPD3 groups. p values were calculated by two-tailed Student's t test. (D) Venn diagram showing the overlap of gene groups regulated by condensin I, condensin II, and E_2, as calculated from GRO-seq data.

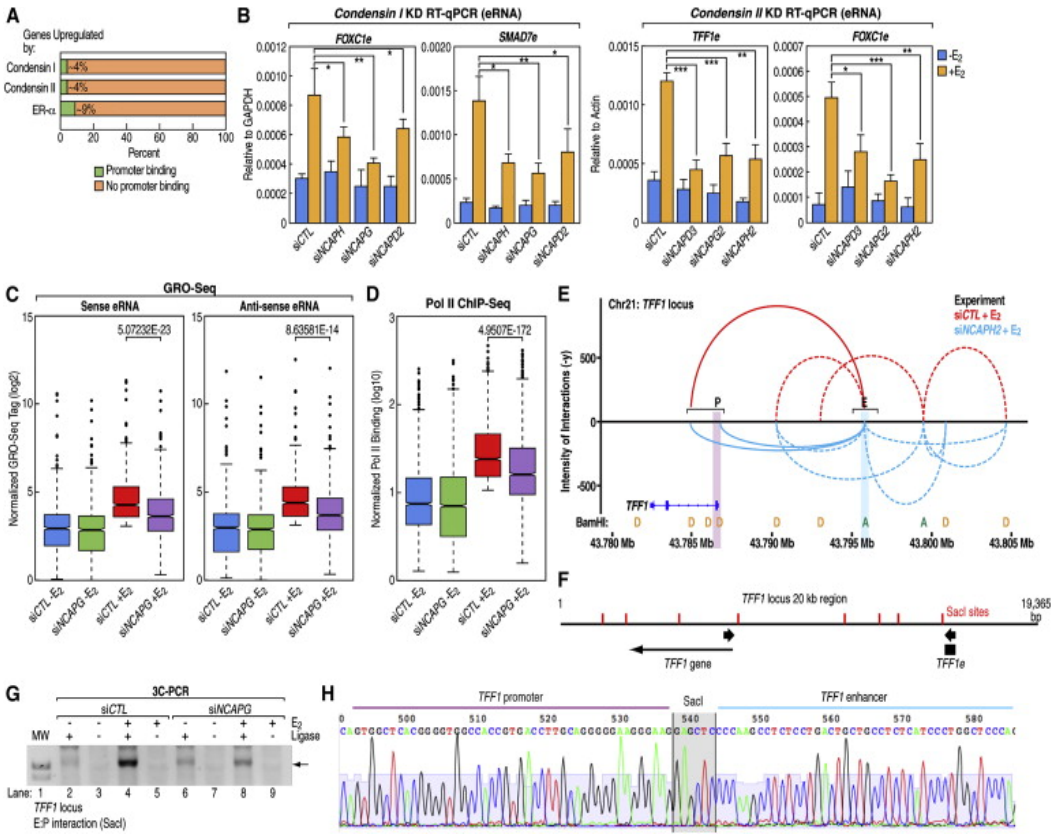


Figure 30 (Chapter 7): Condensins Are Needed for Full eRNA Activation and Enhancer:Promoter Looping.

(A) Bar graphs showing the percentage (green colored) of the RefSeq genes upregulated by condensin I, condensin II, or ER-α that possesses direct promoter-binding of the respective factor. (B) RT-qPCR data indicating the levels of representative E_2-induced eRNAs when either condensin I or condensin II subunit was knocked down ("e" after gene names denotes eRNA) (n = 5). KD, knockdown. (C) Boxplots of normalized GRO-seq tags showing levels of E_2-induced eRNAs on the active enhancers insiCTL or siNCAPG transfected cells, from both sense and antisense directions. (D) Boxplot of normalized ChIP-seq tags showing RNA Pol II recruitment to active enhancers in same group of cells as in (C) (two-tailed Student's t test). (E) Results from 3D-DSL assay showing detected chromatin interactions in the displayed region of TFF1locus in the presence of either siCTL or siNCAPH2 (blue). (F and G) A 3C-PCR assay showing the intensities of a specific E:P looping in the TFF1 locus. The arrow in (G) points to specific PCR product. Control 3C samples without T4 ligase are shown in (G). More details can be found in the Supplemental Experimental Procedures. MW, molecular weight. (H) Sanger sequencing of the 3C-PCR product from (G) (arrow) showing that the ligated fragment comprises regions from TFF1 promoter and enhancer.

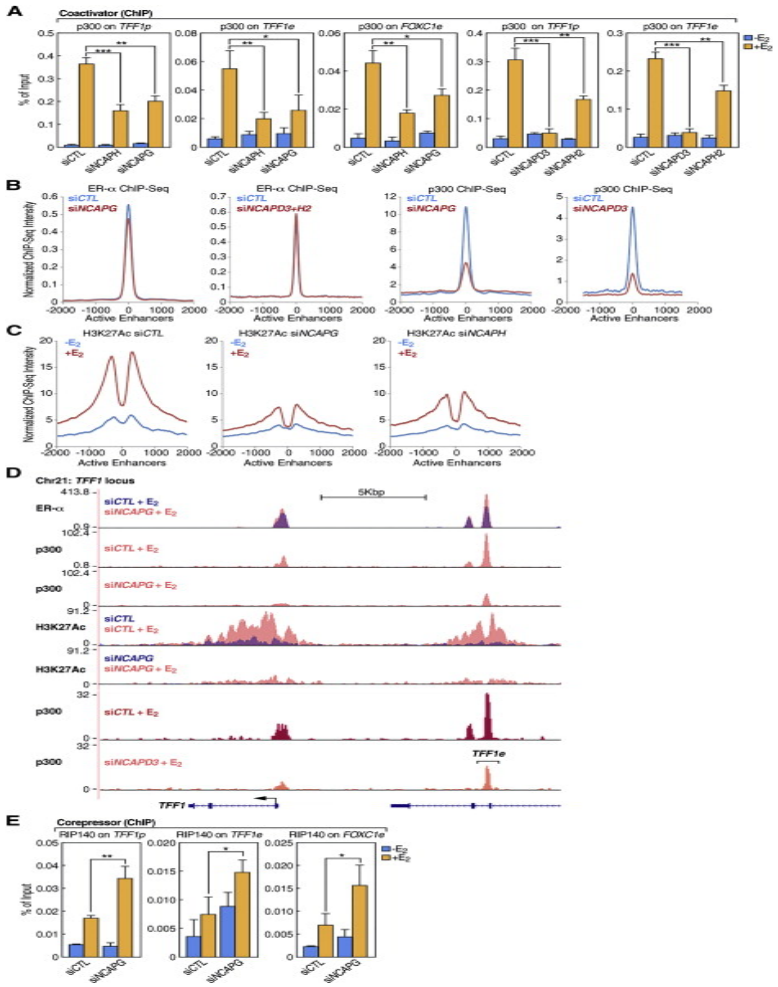


Figure 31 (Chapter 7): Condensins License Appropriate Coactivator and Core-pressor Recruitment during Enhancer Activation.

(A and B) ChIP-qPCR (n = 3) and ChIP-seq profile data showing the effects of condensin knockdown on binding of p300 or ER-α. (C) ChIP-seq profile plots showing the levels of H3K27Ac histone modification on active enhancers in con-trol or knockdown conditions as indicated. (D) Genome browser image showing the binding of p300 and deposition of H3K27Ac, as well as ER-α inTFF1 locus in the presence or absence of NCAPG or NCAPD3. (E) ChIP-qPCR results of RIP140 showing its levels of recruitment to indicated ER-α binding sites upon-siCTL or siNCAPG treatment (n = 3). Profiles in (B) and (C) are centered on ER-α binding sites in $+E_2$ situation. Experiments were repeated as indicated; data are presented as mean ± SD; (two-tailed Student's t test).

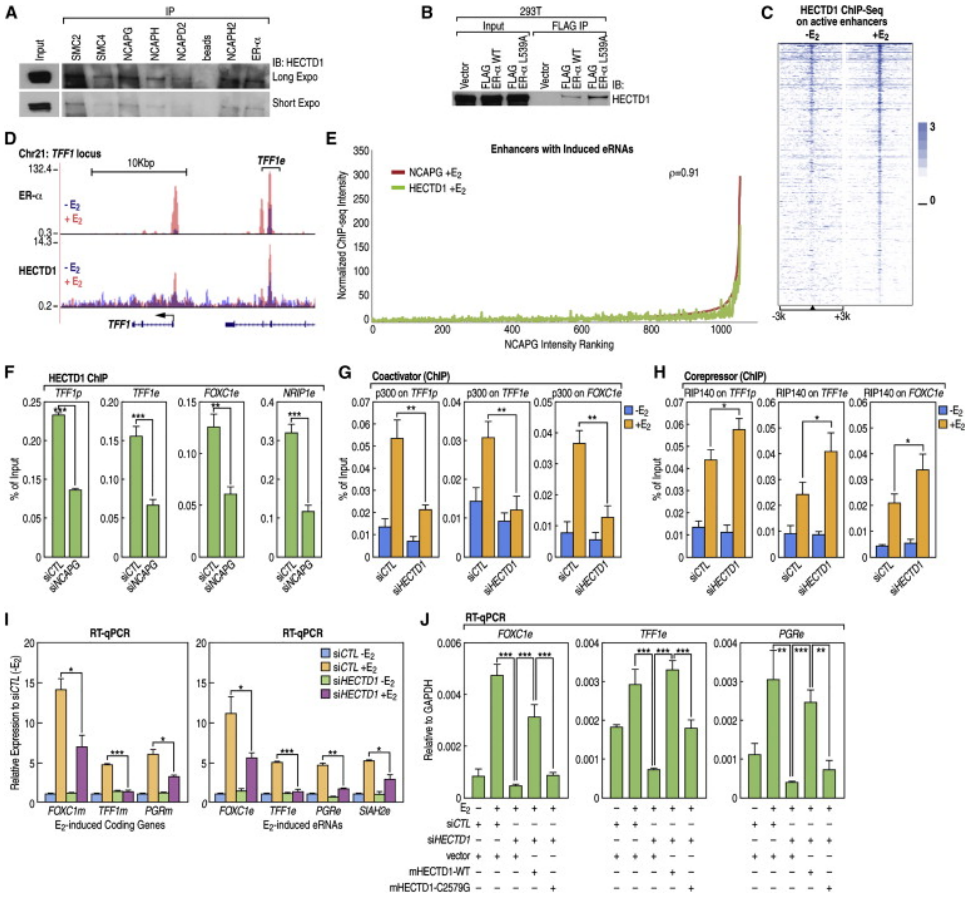


Figure 32 (Chapter 7): Condensins License Appropriate Coactivator and Corepressor Recruitment during Enhancer Activation.

(A) Endogenous coIP followed by western blots showing condensin and HECTD1 interaction using indicated antibodies. Long and Short Expo indicate the lengths of exposure time. (B) Similar to the experiment in Figure 2D, this panel shows the western blots of HECTD1 following coIPs in 293T cells transfected with FLAG-tagged ER-α or its L539A mutant. (C) A heatmap showing HECTD1 ChIP-seq result centered at ER-α binding active enhancers with a scale as indicated. (D) A representative genome browser image from TFF1 locus showing HECTD1 binding to ER-α binding sites. (E) ChIP-seq intensity plots ranked by NCAPG peaks showing the binding of NCAPG (red) and HECTD1 (green) at active enhancers, with the Pearson correlation coefficient shown. (F) ChIP-qPCR results showing the binding of HECTD1 with and without siNCAPG treatment (+E_2) (n = 3). (G and H) ChIP-qPCR results showing the recruitment of p300 and RIP140 to interrogated regions in control cells or siHECTD1 transfectants (n = 3). (I) RT-qPCR data showing the expression levels of interrogated mRNAs and eRNAs in cells with and without siHECTD1 knockdown (n = 4). (J) RT-qPCR data showing the levels of interrogated eRNAs upon HECTD1 knockdown.

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates.

chr1	429713	431931	+
chr1	1003714	1005932	+
chr1	1004502	1006720	+
chr1	6593391	6595609	+
chr1	6941286	6943504	+
chr1	6942240	6944458	+
chr1	7429535	7431753	+
chr1	10734105	10736323	+
chr1	12467336	12469554	+
chr1	12468111	12470329	+
chr1	15124151	15126369	+
chr1	16806436	16808654	+
chr1	16810030	16812248	+
chr1	16812501	16814719	+
chr1	17512943	17515161	+
chr1	17700923	17703141	+
chr1	17718654	17720872	+
chr1	17718654	17720872	+
chr1	17727588	17729806	+
chr1	21439814	21442032	+
chr1	21639621	21641839	+
chr1	21642838	21645056	+
chr1	22667590	22669808	+
chr1	22670763	22672981	+
chr1	30547653	30549871	+
chr1	31431077	31433295	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr1	32067983	32070201	+
chr1	32080110	32082328	+
chr1	32080110	32082328	+
chr1	32117395	32119613	+
chr1	32118109	32120327	+
chr1	32570083	32572301	+
chr1	35062911	35065129	+
chr1	35062911	35065129	+
chr1	42015698	42017916	+
chr1	42027015	42029233	+
chr1	42097574	42099792	+
chr1	42103591	42105809	+
chr1	42105505	42107723	+
chr1	43229730	43231948	+
chr1	44068182	44070400	+
chr1	45756592	45758810	+
chr1	65220475	65222693	+
chr1	65220475	65222693	+
chr1	65234191	65236409	+
chr1	65281708	65283926	+
chr1	108542369	108544587	+
chr1	108580949	108583167	+
chr1	108725031	108727249	+
chr1	108725031	108727249	+
chr1	108799501	108801719	+
chr1	108801335	108803553	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr1	109494467	109496685	+
chr1	109495419	109497637	+
chr1	109583329	109585547	+
chr1	109589453	109591671	+
chr1	109589453	109591671	+
chr1	111856670	111858888	+
chr1	111858018	111860236	+
chr1	114319805	114322023	+
chr1	114377192	114379410	+
chr1	143308074	143310292	+
chr1	149242194	149244412	+
chr1	154851960	154854178	+
chr1	159906157	159908375	+
chr1	159906766	159908984	+
chr1	160394945	160397163	+
chr1	175675707	175677925	+
chr1	175683585	175685803	+
chr1	175683585	175685803	+
chr1	178463347	178465565	+
chr1	178463347	178465565	+
chr1	179361781	179363999	+
chr1	179368424	179370642	+
chr1	179368424	179370642	+
chr1	198251658	198253876	+
chr1	198251658	198253876	+
chr1	200257904	200260122	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr1	200341180	200343398	+
chr1	200341180	200343398	+
chr1	200343073	200345291	+
chr1	200343073	200345291	+
chr1	200347648	200349866	+
chr1	200347648	200349866	+
chr1	200366810	200369028	+
chr1	200367373	200369591	+
chr1	200504484	200506702	+
chr1	200570156	200572374	+
chr1	200766977	200769195	+
chr1	202383563	202385781	+
chr1	202615334	202617552	+
chr1	202615941	202618159	+
chr1	202998478	203000696	+
chr1	205080970	205083188	+
chr1	205082692	205084910	+
chr1	205086549	205088767	+
chr1	205134372	205136590	+
chr1	205134372	205136590	+
chr1	205137641	205139859	+
chr1	205140284	205142502	+
chr1	205142643	205144861	+
chr1	205142643	205144861	+
chr1	205142643	205144861	+
chr1	205142643	205144861	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr1	205148634	205150852	+
chr1	210797010	210799228	+
chr1	210797010	210799228	+
chr10	123376	125594	+
chr10	123376	125594	+
chr10	125145	127363	+
chr10	125145	127363	+
chr10	141966	144184	+
chr10	147193	149411	+
chr10	161977	164195	+
chr10	161977	164195	+
chr10	5737626	5739844	+
chr10	5741285	5743503	+
chr10	6377685	6379903	+
chr10	6377685	6379903	+
chr10	6379031	6381249	+
chr10	6379031	6381249	+
chr10	12953701	12955919	+
chr10	12953701	12955919	+
chr10	12954416	12956634	+
chr10	12954416	12956634	+
chr10	17526773	17528991	+
chr10	28824034	28826252	+
chr10	28824034	28826252	+
chr10	29253443	29255661	+
chr10	29436974	29439192	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr10	29438486	29440704	+
chr10	30140164	30142382	+
chr10	42840909	42843127	+
chr10	42840909	42843127	+
chr10	43988072	43990290	+
chr10	43988072	43990290	+
chr10	49165053	49167271	+
chr10	49165053	49167271	+
chr10	49165909	49168127	+
chr10	51216431	51218649	+
chr10	51216431	51218649	+
chr10	62277494	62279712	+
chr10	62371693	62373911	+
chr10	65059423	65061641	+
chr10	65140955	65143173	+
chr10	65140955	65143173	+
chr10	70240377	70242595	+
chr10	79730859	79733077	+
chr10	80587317	80589535	+
chr10	81762310	81764528	+
chr10	81763451	81765669	+
chr10	81763451	81765669	+
chr10	82039787	82042005	+
chr10	82039787	82042005	+
chr10	95183901	95186119	+
chr10	95189412	95191630	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr10	104459504	104461722	+
chr10	105680246	105682464	+
chr10	105680246	105682464	+
chr10	112856654	112858872	+
chr10	114527508	114529726	+
chr10	114528120	114530338	+
chr10	115670812	115673030	+
chr10	115676163	115678381	+
chr10	115676163	115678381	+
chr10	115731505	115733723	+
chr10	115731505	115733723	+
chr10	115732273	115734491	+
chr10	115805890	115808108	+
chr10	115805890	115808108	+
chr10	116098516	116100734	+
chr10	121436669	121438887	+
chr10	121437673	121439891	+
chr10	127302402	127304620	+
chr10	127302402	127304620	+
chr10	127303495	127305713	+
chr10	127303495	127305713	+
chr10	127311821	127314039	+
chr10	127311821	127314039	+
chr10	128221354	128223572	+
chr10	130228563	130230781	+
chr10	133827204	133829422	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr11	1773431	1775649	+
chr11	1773431	1775649	+
chr11	1805427	1807645	+
chr11	1805427	1807645	+
chr11	1806029	1808247	+
chr11	1807738	1809956	+
chr11	2145974	2148192	+
chr11	2170184	2172402	+
chr11	2170184	2172402	+
chr11	2460443	2462661	+
chr11	10720027	10722245	+
chr11	10720027	10722245	+
chr11	16933863	16936081	+
chr11	16963717	16965935	+
chr11	16966598	16968816	+
chr11	20009055	20011273	+
chr11	44278013	44280231	+
chr11	46340821	46343039	+
chr11	64371885	64374103	+
chr11	66550939	66553157	+
chr11	66585333	66587551	+
chr11	68824853	68827071	+
chr11	69038290	69040508	+
chr11	69038290	69040508	+
chr11	69047708	69049926	+
chr11	69057082	69059300	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr11	69165837	69168055	+
chr11	69417006	69419224	+
chr11	69417006	69419224	+
chr11	69629103	69631321	+
chr11	69639416	69641634	+
chr11	70328665	70330883	+
chr11	70869149	70871367	+
chr11	72170492	72172710	+
chr11	72174203	72176421	+
chr11	72200761	72202979	+
chr11	72580118	72582336	+
chr11	72580118	72582336	+
chr11	72581699	72583917	+
chr11	72581699	72583917	+
chr11	72582379	72584597	+
chr11	72582379	72584597	+
chr11	72588042	72590260	+
chr11	72588042	72590260	+
chr11	72615506	72617724	+
chr11	76160341	76162559	+
chr11	76160341	76162559	+
chr11	76161578	76163796	+
chr11	77274101	77276319	+
chr11	77533487	77535705	+
chr11	77538486	77540704	+
chr11	77544111	77546329	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr11	77544111	77546329	+
chr11	77647861	77650079	+
chr11	78741524	78743742	+
chr11	82391875	82394093	+
chr11	100409154	100411372	+
chr11	100553559	100555777	+
chr11	100674574	100676792	+
chr11	100675461	100677679	+
chr11	100711908	100714126	+
chr11	100811490	100813708	+
chr11	100816718	100818936	+
chr11	100816718	100818936	+
chr11	116183313	116185531	+
chr11	116183313	116185531	+
chr11	116207344	116209562	+
chr11	116211183	116213401	+
chr11	116678083	116680301	+
chr11	117432871	117435089	+
chr11	117432871	117435089	+
chr11	133632773	133634991	+
chr12	502767	504985	+
chr12	703639	705857	+
chr12	1820021	1822239	+
chr12	1910774	1912992	+
chr12	2774385	2776603	+
chr12	6241703	6243921	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr12	6243231	6245449	+
chr12	6243231	6245449	+
chr12	6257042	6259260	+
chr12	47867015	47869233	+
chr12	48483125	48485343	+
chr12	49444657	49446875	+
chr12	52657085	52659303	+
chr12	52662417	52664635	+
chr12	52662417	52664635	+
chr12	62697105	62699323	+
chr12	62760339	62762557	+
chr12	67472232	67474450	+
chr12	67472232	67474450	+
chr12	67473449	67475667	+
chr12	67483373	67485591	+
chr12	67483373	67485591	+
chr12	68119252	68121470	+
chr12	88381935	88384153	+
chr12	100615491	100617709	+
chr12	105890550	105892768	+
chr12	109168718	109170936	+
chr12	109168718	109170936	+
chr12	114790253	114792471	+
chr12	115205050	115207268	+
chr12	115205050	115207268	+
chr12	115220355	115222573	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr12	115307081	115309299	+
chr12	115307081	115309299	+
chr12	115311649	115313867	+
chr12	115311649	115313867	+
chr12	115415424	115417642	+
chr12	115415424	115417642	+
chr12	115416538	115418756	+
chr12	115416538	115418756	+
chr12	118092468	118094686	+
chr12	118092468	118094686	+
chr12	118100886	118103104	+
chr12	120097207	120099425	+
chr12	120670224	120672442	+
chr12	122112477	122114695	+
chr12	122119308	122121526	+
chr12	123795355	123797573	+
chr12	126215402	126217620	+
chr12	130910361	130912579	+
chr12	130910361	130912579	+
chr13	27506320	27508538	+
chr13	38127917	38130135	+
chr13	38155995	38158213	+
chr13	39274492	39276710	+
chr13	40488357	40490575	+
chr13	40496636	40498854	+
chr13	40523672	40525890	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr13	45591593	45593811	+
chr13	97722420	97724638	+
chr13	97742514	97744732	+
chr13	97869272	97871490	+
chr13	98099634	98101852	+
chr13	106202695	106204913	+
chr13	106310566	106312784	+
chr13	106310566	106312784	+
chr13	106311274	106313492	+
chr13	108873482	108875700	+
chr13	109449075	109451293	+
chr13	109449075	109451293	+
chr13	109449634	109451852	+
chr13	109449634	109451852	+
chr13	112318092	112320310	+
chr13	112318092	112320310	+
chr13	112614875	112617093	+
chr13	112614875	112617093	+
chr13	113481015	113483233	+
chr14	45430588	45432806	+
chr14	49553865	49556083	+
chr14	49553865	49556083	+
chr14	50459762	50461980	+
chr14	52687622	52689840	+
chr14	53479081	53481299	+
chr14	53482407	53484625	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr14	58720518	58722736	+
chr14	64653946	64656164	+
chr14	64653946	64656164	+
chr14	64856471	64858689	+
chr14	64856471	64858689	+
chr14	66997325	66999543	+
chr14	67015587	67017805	+
chr14	67993750	67995968	+
chr14	67994590	67996808	+
chr14	68048205	68050423	+
chr14	68089604	68091822	+
chr14	73078672	73080890	+
chr14	73078672	73080890	+
chr14	73110533	73112751	+
chr14	73137617	73139835	+
chr14	73137617	73139835	+
chr14	73169321	73171539	+
chr14	73169321	73171539	+
chr14	73318081	73320299	+
chr14	73320260	73322478	+
chr14	73844272	73846490	+
chr14	73844272	73846490	+
chr14	74790549	74792767	+
chr14	74790549	74792767	+
chr14	74794079	74796297	+
chr14	74794079	74796297	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr14	74795572	74797790	+
chr14	74795572	74797790	+
chr14	74810777	74812995	+
chr14	74818681	74820899	+
chr14	76646333	76648551	+
chr14	76679483	76681701	+
chr14	76679483	76681701	+
chr14	76680048	76682266	+
chr14	76680048	76682266	+
chr14	76935403	76937621	+
chr14	90811588	90813806	+
chr14	92509049	92511267	+
chr14	92510339	92512557	+
chr14	92546784	92549002	+
chr14	92552544	92554762	+
chr14	92567743	92569961	+
chr14	92568204	92570422	+
chr14	92574654	92576872	+
chr14	92580325	92582543	+
chr14	92581133	92583351	+
chr14	94976140	94978358	+
chr14	95051237	95053455	+
chr14	95065639	95067857	+
chr14	95065639	95067857	+
chr14	95151814	95154032	+
chr14	96110048	96112266	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr14	96169625	96171843	+
chr14	96169625	96171843	+
chr14	98426380	98428598	+
chr14	98438624	98440842	+
chr14	99594590	99596808	+
chr14	99672389	99674607	+
chr14	99677822	99680040	+
chr14	102008186	102010404	+
chr14	103228609	103230827	+
chr14	103272202	103274420	+
chr15	36672003	36674221	+
chr15	37612943	37615161	+
chr15	37747059	37749277	+
chr15	37862640	37864858	+
chr15	37863413	37865631	+
chr15	48933248	48935466	+
chr15	60910543	60912761	+
chr15	61458178	61460396	+
chr15	61458839	61461057	+
chr15	61466071	61468289	+
chr15	61466071	61468289	+
chr15	61473998	61476216	+
chr15	62600259	62602477	+
chr15	68094606	68096824	+
chr15	68399711	68401929	+
chr15	68399711	68401929	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr15	68456552	68458770	+
chr15	69157539	69159757	+
chr15	69157539	69159757	+
chr15	69162349	69164567	+
chr15	69175127	69177345	+
chr15	69175127	69177345	+
chr15	69183154	69185372	+
chr15	69183154	69185372	+
chr15	71788592	71790810	+
chr15	72103076	72105294	+
chr15	72895906	72898124	+
chr15	87463151	87465369	+
chr15	87464733	87466951	+
chr15	87482615	87484833	+
chr15	88036940	88039158	+
chr15	88107095	88109313	+
chr15	88171216	88173434	+
chr15	88171216	88173434	+
chr15	88180571	88182789	+
chr15	88180571	88182789	+
chr15	91089249	91091467	+
chr15	91091226	91093444	+
chr15	94237685	94239903	+
chr15	94254937	94257155	+
chr15	94254937	94257155	+
chr15	94682643	94684861	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr15	97109026	97111244	+
chr15	97375469	97377687	+
chr15	99524566	99526784	+
chr16	305824	308042	+
chr16	439604	441822	+
chr16	577188	579406	+
chr16	577188	579406	+
chr16	742583	744801	+
chr16	1276731	1278949	+
chr16	1276731	1278949	+
chr16	1283032	1285250	+
chr16	1283032	1285250	+
chr16	1285977	1288195	+
chr16	1285977	1288195	+
chr16	2239446	2241664	+
chr16	4131581	4133799	+
chr16	4141348	4143566	+
chr16	4141348	4143566	+
chr16	4165903	4168121	+
chr16	4358123	4360341	+
chr16	4358123	4360341	+
chr16	8679658	8681876	+
chr16	8786800	8789018	+
chr16	9048072	9050290	+
chr16	9048072	9050290	+
chr16	14370658	14372876	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr16	15489828	15492046	+
chr16	16974198	16976416	+
chr16	20357667	20359885	+
chr16	20637060	20639278	+
chr16	20637060	20639278	+
chr16	21306731	21308949	+
chr16	21429361	21431579	+
chr16	21430445	21432663	+
chr16	22997596	22999814	+
chr16	23055821	23058039	+
chr16	23070980	23073198	+
chr16	23070980	23073198	+
chr16	23292771	23294989	+
chr16	29228573	29230791	+
chr16	29522614	29524832	+
chr16	33862183	33864401	+
chr16	33862183	33864401	+
chr16	33862867	33865085	+
chr16	33862867	33865085	+
chr16	33869677	33871895	+
chr16	33869677	33871895	+
chr16	33870420	33872638	+
chr16	33870420	33872638	+
chr16	66918503	66920721	+
chr16	68371282	68373500	+
chr16	68727128	68729346	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr16	68981231	68983449	+
chr16	68981868	68984086	+
chr16	82537024	82539242	+
chr16	84043968	84046186	+
chr16	84047566	84049784	+
chr16	84049148	84051366	+
chr16	84052255	84054473	+
chr16	84052255	84054473	+
chr16	84056172	84058390	+
chr16	84056172	84058390	+
chr16	84062749	84064967	+
chr16	84064542	84066760	+
chr16	84065438	84067656	+
chr16	84082161	84084379	+
chr16	84082161	84084379	+
chr16	84111131	84113349	+
chr16	84164573	84166791	+
chr16	84343621	84345839	+
chr16	84783974	84786192	+
chr16	86380142	86382360	+
chr16	86380142	86382360	+
chr16	86397189	86399407	+
chr16	86397189	86399407	+
chr16	87516788	87519006	+
chr16	87520071	87522289	+
chr16	87521409	87523627	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr17	4097135	4099353	+
chr17	4382804	4385022	+
chr17	4383433	4385651	+
chr17	4424941	4427159	+
chr17	7329343	7331561	+
chr17	7794855	7797073	+
chr17	7794855	7797073	+
chr17	7800817	7803035	+
chr17	23848195	23850413	+
chr17	23875532	23877750	+
chr17	23883079	23885297	+
chr17	23887040	23889258	+
chr17	23984292	23986510	+
chr17	26916629	26918847	+
chr17	30501200	30503418	+
chr17	33674675	33676893	+
chr17	33675792	33678010	+
chr17	35589579	35591797	+
chr17	35720122	35722340	+
chr17	35731156	35733374	+
chr17	35857521	35859739	+
chr17	35958126	35960344	+
chr17	35958126	35960344	+
chr17	35961355	35963573	+
chr17	35964305	35966523	+
chr17	35985085	35987303	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr17	35985085	35987303	+
chr17	35991124	35993342	+
chr17	36095052	36097270	+
chr17	36095052	36097270	+
chr17	36096321	36098539	+
chr17	36096321	36098539	+
chr17	36824812	36827030	+
chr17	36824812	36827030	+
chr17	36915548	36917766	+
chr17	36920570	36922788	+
chr17	36944824	36947042	+
chr17	36944824	36947042	+
chr17	40343068	40345286	+
chr17	40475168	40477386	+
chr17	40561974	40564192	+
chr17	42952620	42954838	+
chr17	42952620	42954838	+
chr17	43105402	43107620	+
chr17	43124557	43126775	+
chr17	43124557	43126775	+
chr17	45061611	45063829	+
chr17	45319043	45321261	+
chr17	45748719	45750937	+
chr17	45748719	45750937	+
chr17	46084142	46086360	+
chr17	46216578	46218796	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr17	46216578	46218796	+
chr17	46277103	46279321	+
chr17	46290063	46292281	+
chr17	46293187	46295405	+
chr17	46375800	46378018	+
chr17	46379612	46381830	+
chr17	46379612	46381830	+
chr17	46381405	46383623	+
chr17	46381405	46383623	+
chr17	52772022	52774240	+
chr17	52774282	52776500	+
chr17	52788806	52791024	+
chr17	53265553	53267771	+
chr17	53265553	53267771	+
chr17	53331250	53333468	+
chr17	55554916	55557134	+
chr17	55554916	55557134	+
chr17	55776309	55778527	+
chr17	55778280	55780498	+
chr17	58883230	58885448	+
chr17	62359059	62361277	+
chr17	62359059	62361277	+
chr17	62377904	62380122	+
chr17	62377904	62380122	+
chr17	68124764	68126982	+
chr17	70251604	70253822	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr17	70265081	70267299	+
chr17	70267484	70269702	+
chr17	70276575	70278793	+
chr17	73927839	73930057	+
chr17	74512635	74514853	+
chr17	74518364	74520582	+
chr17	74524311	74526529	+
chr17	74524311	74526529	+
chr17	78164182	78166400	+
chr17	78617762	78619980	+
chr18	13383974	13386192	+
chr18	13414919	13417137	+
chr18	18704777	18706995	+
chr18	18706248	18708466	+
chr18	19555733	19557951	+
chr18	31930149	31932367	+
chr18	35356285	35358503	+
chr18	41626515	41628733	+
chr18	41626515	41628733	+
chr18	41635664	41637882	+
chr18	41635664	41637882	+
chr18	41946782	41949000	+
chr18	42106715	42108933	+
chr18	44232869	44235087	+
chr18	44232869	44235087	+
chr18	44233509	44235727	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr18	44233509	44235727	+
chr18	44754022	44756240	+
chr18	44754022	44756240	+
chr18	44762428	44764646	+
chr18	44762428	44764646	+
chr18	44763393	44765611	+
chr18	44763393	44765611	+
chr18	44770270	44772488	+
chr18	44770270	44772488	+
chr18	44776353	44778571	+
chr18	44776353	44778571	+
chr18	44801014	44803232	+
chr18	44801014	44803232	+
chr18	52958961	52961179	+
chr18	53059645	53061863	+
chr18	53254761	53256979	+
chr18	53610925	53613143	+
chr18	53611613	53613831	+
chr18	55771554	55773772	+
chr18	58949781	58951999	+
chr18	58972889	58975107	+
chr18	59043016	59045234	+
chr18	59058710	59060928	+
chr18	59076808	59079026	+
chr18	70059939	70062157	+
chr18	70059939	70062157	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr18	74752885	74755103	+
chr19	1131892	1134110	+
chr19	1135669	1137887	+
chr19	1135669	1137887	+
chr19	1158048	1160266	+
chr19	1821345	1823563	+
chr19	2475053	2477271	+
chr19	3658705	3660923	+
chr19	3658705	3660923	+
chr19	4920847	4923065	+
chr19	4921728	4923946	+
chr19	7605668	7607886	+
chr19	9817359	9819577	+
chr19	10457347	10459565	+
chr19	12520064	12522282	+
chr19	15459474	15461692	+
chr19	15459474	15461692	+
chr19	15946710	15948928	+
chr19	15946710	15948928	+
chr19	16434817	16437035	+
chr19	16795546	16797764	+
chr19	16795546	16797764	+
chr19	19520671	19522889	+
chr19	23374050	23376268	+
chr19	23374050	23376268	+
chr19	23375946	23378164	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr19	35880336	35882554	+
chr19	35880336	35882554	+
chr19	38960016	38962234	+
chr19	39061640	39063858	+
chr19	39267359	39269577	+
chr19	40757316	40759534	+
chr19	43480777	43482995	+
chr19	43480777	43482995	+
chr19	43503140	43505358	+
chr19	44311934	44314152	+
chr19	44332555	44334773	+
chr19	45677006	45679224	+
chr19	46094835	46097053	+
chr19	53256836	53259054	+
chr19	53748007	53750225	+
chr19	57001487	57003705	+
chr19	63000903	63003121	+
chr19	63001402	63003620	+
chr2	9864441	9866659	+
chr2	10078393	10080611	+
chr2	11555205	11557423	+
chr2	11555205	11557423	+
chr2	11560231	11562449	+
chr2	11569415	11571633	+
chr2	11569415	11571633	+
chr2	11581063	11583281	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr2	11587590	11589808	+
chr2	11587590	11589808	+
chr2	11589030	11591248	+
chr2	11589030	11591248	+
chr2	11615312	11617530	+
chr2	16550142	16552360	+
chr2	16550142	16552360	+
chr2	20133965	20136183	+
chr2	20133965	20136183	+
chr2	20443653	20445871	+
chr2	26847943	26850161	+
chr2	28452619	28454837	+
chr2	28452619	28454837	+
chr2	38985596	38987814	+
chr2	39259995	39262213	+
chr2	42422216	42424434	+
chr2	42422216	42424434	+
chr2	42445991	42448209	+
chr2	42445991	42448209	+
chr2	42864313	42866531	+
chr2	42874269	42876487	+
chr2	45832288	45834506	+
chr2	45833114	45835332	+
chr2	70695959	70698177	+
chr2	70695959	70698177	+
chr2	73065435	73067653	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr2	73334217	73336435	+
chr2	74977592	74979810	+
chr2	74977592	74979810	+
chr2	74979886	74982104	+
chr2	98215903	98218121	+
chr2	99556046	99558264	+
chr2	99591799	99594017	+
chr2	99599844	99602062	+
chr2	99618620	99620838	+
chr2	100274010	100276228	+
chr2	109454611	109456829	+
chr2	109462560	109464778	+
chr2	113649822	113652040	+
chr2	120877235	120879453	+
chr2	120877235	120879453	+
chr2	120879950	120882168	+
chr2	120879950	120882168	+
chr2	121037996	121040214	+
chr2	121607802	121610020	+
chr2	121793141	121795359	+
chr2	130712747	130714965	+
chr2	131907438	131909656	+
chr2	132741582	132743800	+
chr2	133055555	133057773	+
chr2	156963249	156965467	+
chr2	157467137	157469355	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr2	157467137	157469355	+
chr2	157497739	157499957	+
chr2	157499517	157501735	+
chr2	159071081	159073299	+
chr2	159291151	159293369	+
chr2	170870835	170873053	+
chr2	191358365	191360583	+
chr2	200860658	200862876	+
chr2	201318978	201321196	+
chr2	201320051	201322269	+
chr2	202557733	202559951	+
chr2	202557733	202559951	+
chr2	202558340	202560558	+
chr2	202558340	202560558	+
chr2	216551632	216553850	+
chr2	216552168	216554386	+
chr2	216554636	216556854	+
chr2	232278241	232280459	+
chr2	237184035	237186253	+
chr2	237184035	237186253	+
chr2	237240724	237242942	+
chr2	237304469	237306687	+
chr2	237613749	237615967	+
chr2	237613749	237615967	+
chr2	238060901	238063119	+
chr2	238131424	238133642	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr2	240657023	240659241	+
chr2	240657023	240659241	+
chr20	648544	650762	+
chr20	17822845	17825063	+
chr20	17823986	17826204	+
chr20	17823986	17826204	+
chr20	23285460	23287678	+
chr20	23288775	23290993	+
chr20	23288775	23290993	+
chr20	24577709	24579927	+
chr20	24983822	24986040	+
chr20	30091290	30093508	+
chr20	30091290	30093508	+
chr20	34077596	34079814	+
chr20	34113146	34115364	+
chr20	34115153	34117371	+
chr20	34115153	34117371	+
chr20	34648158	34650376	+
chr20	34648158	34650376	+
chr20	34648158	34650376	+
chr20	34648158	34650376	+
chr20	35325596	35327814	+
chr20	35330692	35332910	+
chr20	35340428	35342646	+
chr20	35340428	35342646	+
chr20	36170779	36172997	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr20	42243132	42245350	+
chr20	42826710	42828928	+
chr20	43382960	43385178	+
chr20	46747193	46749411	+
chr20	46913792	46916010	+
chr20	46913792	46916010	+
chr20	48125698	48127916	+
chr20	48125698	48127916	+
chr20	52172698	52174916	+
chr20	52172698	52174916	+
chr20	52176135	52178353	+
chr20	52176135	52178353	+
chr20	52198082	52200300	+
chr20	52198082	52200300	+
chr20	52244346	52246564	+
chr20	52244346	52246564	+
chr20	54766325	54768543	+
chr20	54766325	54768543	+
chr21	15296692	15298910	+
chr21	15298597	15300815	+
chr21	15493249	15495467	+
chr21	15493249	15495467	+
chr21	15503544	15505762	+
chr21	15503544	15505762	+
chr21	15737512	15739730	+
chr21	15880146	15882364	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr21	15880146	15882364	+
chr21	29802684	29804902	+
chr21	31822696	31824914	+
chr21	36245448	36247666	+
chr21	36246917	36249135	+
chr21	36246917	36249135	+
chr21	36528375	36530593	+
chr21	37857990	37860208	+
chr21	37857990	37860208	+
chr21	38061279	38063497	+
chr21	39207524	39209742	+
chr21	39207524	39209742	+
chr21	40631962	40634180	+
chr21	40633244	40635462	+
chr21	41823180	41825398	+
chr21	42352237	42354455	+
chr21	42378429	42380647	+
chr21	42667728	42669946	+
chr21	42668563	42670781	+
chr21	42682998	42685216	+
chr21	42684969	42687187	+
chr21	43298831	43301049	+
chr21	44393896	44396114	+
chr21	44397823	44400041	+
chr21	44439351	44441569	+
chr21	44439351	44441569	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr21	44439938	44442156	+
chr21	44439938	44442156	+
chr21	45570604	45572822	+
chr21	45630134	45632352	+
chr21	45630134	45632352	+
chr21	45646662	45648880	+
chr21	45646662	45648880	+
chr21	45665103	45667321	+
chr21	45665103	45667321	+
chr22	17539682	17541900	+
chr22	17539682	17541900	+
chr22	19568372	19570590	+
chr22	19570904	19573122	+
chr22	23077190	23079408	+
chr22	24125860	24128078	+
chr22	25409712	25411930	+
chr22	27515755	27517973	+
chr22	27538791	27541009	+
chr22	27538791	27541009	+
chr22	27544522	27546740	+
chr22	27548834	27551052	+
chr22	28110620	28112838	+
chr22	28111207	28113425	+
chr22	28206766	28208984	+
chr22	28242403	28244621	+
chr22	28779663	28781881	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr22	29016125	29018343	+
chr22	35903368	35905586	+
chr22	35914196	35916414	+
chr22	35914196	35916414	+
chr22	35921883	35924101	+
chr22	35921883	35924101	+
chr22	35922826	35925044	+
chr22	35922826	35925044	+
chr22	36288708	36290926	+
chr22	36295209	36297427	+
chr22	45151224	45153442	+
chr22	48634468	48636686	+
chr3	4729715	4731933	+
chr3	11784254	11786472	+
chr3	13491723	13493941	+
chr3	13491723	13493941	+
chr3	14272193	14274411	+
chr3	14272193	14274411	+
chr3	14280842	14283060	+
chr3	14280842	14283060	+
chr3	14428512	14430730	+
chr3	14686425	14688643	+
chr3	14687812	14690030	+
chr3	14692589	14694807	+
chr3	15308400	15310618	+
chr3	46681277	46683495	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr3	46682182	46684400	+
chr3	46682182	46684400	+
chr3	50276636	50278854	+
chr3	50282978	50285196	+
chr3	50616589	50618807	+
chr3	50616589	50618807	+
chr3	52440891	52443109	+
chr3	53527462	53529680	+
chr3	53559378	53561596	+
chr3	54349463	54351681	+
chr3	57037767	57039985	+
chr3	58061803	58064021	+
chr3	58427049	58429267	+
chr3	61767546	61769764	+
chr3	66607792	66610010	+
chr3	109025113	109027331	+
chr3	109025113	109027331	+
chr3	109026209	109028427	+
chr3	109026209	109028427	+
chr3	109187193	109189411	+
chr3	109187193	109189411	+
chr3	109200140	109202358	+
chr3	109206783	109209001	+
chr3	109206783	109209001	+
chr3	124822511	124824729	+
chr3	124822511	124824729	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr3	127467894	127470112	+
chr3	130767937	130770155	+
chr3	130894919	130897137	+
chr3	131079063	131081281	+
chr3	131172032	131174250	+
chr3	135514450	135516668	+
chr3	135528173	135530391	+
chr3	151922300	151924518	+
chr3	151936773	151938991	+
chr3	151936773	151938991	+
chr3	151956009	151958227	+
chr3	151957342	151959560	+
chr3	151973051	151975269	+
chr3	152044043	152046261	+
chr3	152044043	152046261	+
chr3	155418270	155420488	+
chr3	155423980	155426198	+
chr3	157837249	157839467	+
chr3	162300424	162302642	+
chr3	162472192	162474410	+
chr3	171520062	171522280	+
chr3	171520062	171522280	+
chr3	171534356	171536574	+
chr3	171543460	171545678	+
chr3	185026138	185028356	+
chr3	195027965	195030183	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr3	195036600	195038818	+
chr3	195036600	195038818	+
chr3	195051036	195053254	+
chr3	195069974	195072192	+
chr3	195069974	195072192	+
chr3	195071830	195074048	+
chr3	195071830	195074048	+
chr3	195078063	195080281	+
chr3	195078063	195080281	+
chr3	195079072	195081290	+
chr3	195079072	195081290	+
chr3	195121021	195123239	+
chr3	195121021	195123239	+
chr3	195121820	195124038	+
chr3	195337174	195339392	+
chr3	195338740	195340958	+
chr3	195350071	195352289	+
chr3	195350071	195352289	+
chr3	195358504	195360722	+
chr3	195358504	195360722	+
chr3	195513351	195515569	+
chr3	195515079	195517297	+
chr3	195515079	195517297	+
chr3	195695373	195697591	+
chr3	195696094	195698312	+
chr4	980677	982895	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr4	1228273	1230491	+
chr4	3431593	3433811	+
chr4	3431593	3433811	+
chr4	6364843	6367061	+
chr4	6455558	6457776	+
chr4	6467002	6469220	+
chr4	15612265	15614483	+
chr4	17310720	17312938	+
chr4	55401362	55403580	+
chr4	55402402	55404620	+
chr4	79449963	79452181	+
chr4	88260118	88262336	+
chr4	140816801	140819019	+
chr4	140988959	140991177	+
chr4	141015108	141017326	+
chr4	141389840	141392058	+
chr4	141389840	141392058	+
chr4	141392582	141394800	+
chr4	141392582	141394800	+
chr4	146875916	146878134	+
chr4	146875916	146878134	+
chr4	146879410	146881628	+
chr4	164525227	164527445	+
chr4	164525227	164527445	+
chr4	164587145	164589363	+
chr4	164588171	164590389	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr4	164588171	164590389	+
chr4	169848448	169850666	+
chr4	186671004	186673222	+
chr5	16640990	16643208	+
chr5	16642259	16644477	+
chr5	55360531	55362749	+
chr5	55784710	55786928	+
chr5	55790457	55792675	+
chr5	55791214	55793432	+
chr5	55810829	55813047	+
chr5	55810829	55813047	+
chr5	60509493	60511711	+
chr5	60664343	60666561	+
chr5	66545717	66547935	+
chr5	66545717	66547935	+
chr5	66546375	66548593	+
chr5	66546375	66548593	+
chr5	66565409	66567627	+
chr5	66565409	66567627	+
chr5	66566386	66568604	+
chr5	66566386	66568604	+
chr5	66567442	66569660	+
chr5	70961970	70964188	+
chr5	77192269	77194487	+
chr5	95033037	95035255	+
chr5	95058816	95061034	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr5	95088307	95090525	+
chr5	95088307	95090525	+
chr5	95089162	95091380	+
chr5	95089162	95091380	+
chr5	117372554	117374772	+
chr5	118927401	118929619	+
chr5	131733765	131735983	+
chr5	131741278	131743496	+
chr5	132412361	132414579	+
chr5	132412361	132414579	+
chr5	132611996	132614214	+
chr5	133399911	133402129	+
chr5	138995976	138998194	+
chr5	139067748	139069966	+
chr5	139067748	139069966	+
chr5	141639847	141642065	+
chr5	141639847	141642065	+
chr5	149216329	149218547	+
chr5	149236893	149239111	+
chr5	154427462	154429680	+
chr5	154427462	154429680	+
chr5	157103668	157105886	+
chr5	167655443	167657661	+
chr5	167675004	167677222	+
chr5	168458242	168460460	+
chr5	172811901	172814119	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr5	172811901	172814119	+
chr5	172831311	172833529	+
chr5	172894182	172896400	+
chr5	173032134	173034352	+
chr5	173819978	173822196	+
chr5	173842097	173844315	+
chr5	173842097	173844315	+
chr6	1585616	1587834	+
chr6	3662491	3664709	+
chr6	3692620	3694838	+
chr6	6648064	6650282	+
chr6	6648064	6650282	+
chr6	6703130	6705348	+
chr6	6703130	6705348	+
chr6	11154036	11156254	+
chr6	11154036	11156254	+
chr6	11160816	11163034	+
chr6	11160816	11163034	+
chr6	11174614	11176832	+
chr6	11174614	11176832	+
chr6	15143994	15146212	+
chr6	15143994	15146212	+
chr6	15197359	15199577	+
chr6	15197359	15199577	+
chr6	15329174	15331392	+
chr6	15329174	15331392	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr6	17496571	17498789	+
chr6	17496571	17498789	+
chr6	17580585	17582803	+
chr6	32629058	32631276	+
chr6	32641700	32643918	+
chr6	32641700	32643918	+
chr6	33792043	33794261	+
chr6	33807147	33809365	+
chr6	33891691	33893909	+
chr6	33971750	33973968	+
chr6	34224221	34226439	+
chr6	36193234	36195452	+
chr6	36197326	36199544	+
chr6	36197326	36199544	+
chr6	39341577	39343795	+
chr6	39341577	39343795	+
chr6	39344784	39347002	+
chr6	39344784	39347002	+
chr6	39349543	39351761	+
chr6	39349543	39351761	+
chr6	39350267	39352485	+
chr6	39350267	39352485	+
chr6	42150138	42152356	+
chr6	43409916	43412134	+
chr6	43789133	43791351	+
chr6	44185131	44187349	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr6	44312816	44315034	+
chr6	44312816	44315034	+
chr6	44736073	44738291	+
chr6	53436719	53438937	+
chr6	53436719	53438937	+
chr6	53437561	53439779	+
chr6	53437561	53439779	+
chr6	83335565	83337783	+
chr6	83335565	83337783	+
chr6	83336148	83338366	+
chr6	83336148	83338366	+
chr6	112019641	112021859	+
chr6	112019641	112021859	+
chr6	125559835	125562053	+
chr6	125564155	125566373	+
chr6	125566990	125569208	+
chr6	125946586	125948804	+
chr6	135544337	135546555	+
chr6	135573131	135575349	+
chr6	138917489	138919707	+
chr6	149442091	149444309	+
chr6	157189188	157191406	+
chr6	159047207	159049425	+
chr7	2544716	2546934	+
chr7	2692709	2694927	+
chr7	2694353	2696571	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr7	4688598	4690816	+
chr7	26558854	26561072	+
chr7	27742836	27745054	+
chr7	27742836	27745054	+
chr7	30165104	30167322	+
chr7	30175628	30177846	+
chr7	32685835	32688053	+
chr7	32727822	32730040	+
chr7	36278063	36280281	+
chr7	44158648	44160866	+
chr7	63696258	63698476	+
chr7	63696258	63698476	+
chr7	63736288	63738506	+
chr7	63736288	63738506	+
chr7	63764212	63766430	+
chr7	65115328	65117546	+
chr7	65117572	65119790	+
chr7	74045137	74047355	+
chr7	75773346	75775564	+
chr7	101374070	101376288	+
chr7	101402170	101404388	+
chr7	101433246	101435464	+
chr7	104743239	104745457	+
chr7	105066643	105068861	+
chr7	116550846	116553064	+
chr7	116550846	116553064	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr7	128478381	128480599	+
chr7	130960114	130962332	+
chr7	130961037	130963255	+
chr7	130961037	130963255	+
chr7	131450823	131453041	+
chr7	131451634	131453852	+
chr7	131548828	131551046	+
chr7	131559653	131561871	+
chr7	139087126	139089344	+
chr7	139655668	139657886	+
chr7	140337910	140340128	+
chr7	151021893	151024111	+
chr7	151183649	151185867	+
chr7	151202355	151204573	+
chr7	154854310	154856528	+
chr7	154854310	154856528	+
chr7	154864747	154866965	+
chr7	156000658	156002876	+
chr7	156739841	156742059	+
chr7	156818886	156821104	+
chr8	8130899	8133117	+
chr8	8183087	8185305	+
chr8	8183087	8185305	+
chr8	8184277	8186495	+
chr8	8184277	8186495	+
chr8	8208816	8211034	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr8	8402002	8404220	+
chr8	9013763	9015981	+
chr8	11624644	11626862	+
chr8	11683625	11685843	+
chr8	13057831	13060049	+
chr8	17230654	17232872	+
chr8	22025188	22027406	+
chr8	22025188	22027406	+
chr8	22651605	22653823	+
chr8	22656340	22658558	+
chr8	29195434	29197652	+
chr8	30507515	30509733	+
chr8	37037961	37040179	+
chr8	37120571	37122789	+
chr8	37566368	37568586	+
chr8	37566368	37568586	+
chr8	37569159	37571377	+
chr8	37569159	37571377	+
chr8	37574320	37576538	+
chr8	37574320	37576538	+
chr8	37670000	37672218	+
chr8	41535169	41537387	+
chr8	67586144	67588362	+
chr8	67586628	67588846	+
chr8	67596337	67598555	+
chr8	67596337	67598555	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr8	67787872	67790090	+
chr8	67787872	67790090	+
chr8	98346853	98349071	+
chr8	98346853	98349071	+
chr8	98844680	98846898	+
chr8	98844680	98846898	+
chr8	100985917	100988135	+
chr8	100985917	100988135	+
chr8	101775682	101777900	+
chr8	101800169	101802387	+
chr8	101988301	101990519	+
chr8	102201160	102203378	+
chr8	102201160	102203378	+
chr8	102206600	102208818	+
chr8	102543214	102545432	+
chr8	102543214	102545432	+
chr8	102547156	102549374	+
chr8	102547156	102549374	+
chr8	102554488	102556706	+
chr8	102554488	102556706	+
chr8	102557854	102560072	+
chr8	102570051	102572269	+
chr8	102584532	102586750	+
chr8	102594867	102597085	+
chr8	102595992	102598210	+
chr8	103681379	103683597	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr8	103681379	103683597	+
chr8	103682865	103685083	+
chr8	103682865	103685083	+
chr8	107704483	107706701	+
chr8	122231375	122233593	+
chr8	128869269	128871487	+
chr8	128880345	128882563	+
chr8	128881252	128883470	+
chr8	128940296	128942514	+
chr8	128948188	128950406	+
chr8	128951074	128953292	+
chr8	128991687	128993905	+
chr8	129188611	129190829	+
chr8	129222387	129224605	+
chr8	129222387	129224605	+
chr8	129223770	129225988	+
chr8	129223770	129225988	+
chr8	129233499	129235717	+
chr8	129233499	129235717	+
chr8	143624110	143626328	+
chr8	143624110	143626328	+
chr8	144167713	144169931	+
chr8	146253858	146256076	+
chr9	4745139	4747357	+
chr9	4745139	4747357	+
chr9	4855386	4857604	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr9	4855386	4857604	+
chr9	4857217	4859435	+
chr9	33095915	33098133	+
chr9	33152887	33155105	+
chr9	33210315	33212533	+
chr9	33210315	33212533	+
chr9	33218373	33220591	+
chr9	33218373	33220591	+
chr9	33223221	33225439	+
chr9	33223221	33225439	+
chr9	33709035	33711253	+
chr9	73587077	73589295	+
chr9	73587077	73589295	+
chr9	76904867	76907085	+
chr9	78375515	78377733	+
chr9	89732486	89734704	+
chr9	90148315	90150533	+
chr9	90333834	90336052	+
chr9	90333834	90336052	+
chr9	94467238	94469456	+
chr9	109045053	109047271	+
chr9	109342783	109345001	+
chr9	113813319	113815537	+
chr9	115885462	115887680	+
chr9	128444120	128446338	+
chr9	128445124	128447342	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chr9	129039051	129041269	+
chr9	129218368	129220586	+
chr9	129779773	129781991	+
chr9	129823391	129825609	+
chr9	129823391	129825609	+
chr9	129828056	129830274	+
chr9	129848410	129850628	+
chr9	129859181	129861399	+
chr9	129859181	129861399	+
chr9	130994852	130997070	+
chr9	131556816	131559034	+
chr9	131556816	131559034	+
chr9	135727272	135729490	+
chr9	137423301	137425519	+
chr9	137423301	137425519	+
chrX	40261357	40263575	+
chrX	40262328	40264546	+
chrX	40321311	40323529	+
chrX	100740264	100742482	+
chrX	100754589	100756807	+
chrX	108183263	108185481	+
chrX	109302128	109304346	+
chrX	109530919	109533137	+
chrX	136946384	136948602	+
chrX	136946384	136948602	+
chrX	139401140	139403358	+

Table 1 (Chapter 7): Active Enhancers Genomic Coordinates (Continued).

chrX	139416992	139419210	+
chrX	151814051	151816269	+
chrX	151833745	151835963	+
chrX	153300498	153302716	+
chrX	153530169	153532387	+
chrX	153530169	153532387	+
chrX	153868638	153870856	+

APPENDIX A

TCC technology is described in a 2011 Kalhor paper "Genome architectures revealed by tethered chromosome conformation capture and population-based modeling". For the genome-wide mapping of chromatin contacts, the authors have developed the TCC technology, a modified conformation capture method in which key reactions are carried out on solid phase instead of in solution. This tethering strategy leads to higher signal-to-noise ratios, enabling an in-depth analysis of interchromosomal interactions. They show that a specific group of functionally active loci are more likely to form interchromosomal contacts and that most of these contacts are a result of indiscriminate encounters between loci that are accessible to each other. They also introduce a structural modeling procedure that calculates a population of 3D genome structures from the TCC data. They show that the calculated population reproduces the hallmarks of chromosome territory positioning in agreement with independent fluorescence in situ hybridization (FISH) studies. This population-based approach allows for a probabilistic analysis of the spatial features of the genome, a capability that can accommodate the wide range of cell-to-cell structural variations that are observed in mammalian genomes.

To identify chromatin interactions using TCC, they preserved native chromatin contacts by chemically crosslinking DNA and proteins. The DNA was then digested with a restriction enzyme, and, after cysteine biotinylation of proteins, the protein-bound fragments were immobilized at a low surface density on streptavidin-coated beads. The immobilized DNA fragments were then ligated

while tethered to the surface of the beads. Finally, ligation junctions were purified, and ligation events were detected by massively parallel sequencing, a process that revealed the genomic locations of the pairs of loci that had formed the initial contacts. One of the main sources of noise in conformation capture experiments is random intermolecular ligations between DNA fragments that are not crosslinked to each other. Because randomly selected DNA fragments are more likely to originate from different chromosomes, these ligations tend to be overwhelmingly interchromosomal compared to intrachromosomal interactions.

Therefore, the authors measured the fraction of interchromosomal ligations in our tethered (TCC) and nontethered (Hi-C) HindIII libraries. This was done to compare their relative noise levels. In the tethered library, this fraction is almost half that of the nontethered library. They also compared the average difference between the observed interchromosomal contact frequencies in each library and those expected from completely random intermolecular ligations. This difference is twice as large in the tethered library compared to the nontethered library. Together, these observations indicate that the noise from random intermolecular ligations is considerably lower in the tethered library.

REFERENCES

1. Gosselin D, Glass CK. Epigenomics of macrophages. Immunol Rev. 2014;262(1):96–112.

2. Love PE, Warzecha C, Li L. Ldb1 complexes: The new master regulators of erythroid gene transcription.Trends Genet. 2014;30(1):1–9.

3. Jenks BG. Regulation of proopiomelanocortin gene expression: An overview of the signaling cascades, transcription factors, and responsive elements involved. Ann N Y Acad Sci. 2009;1163:17–30.

4. Lee SK, Pfaff SL. Synchronization of neurogenesis and motor neuron specification by direct coupling of bHLH and homeodomain transcription factors. Neuron. 2003;38(5):731–745.

5. Jones S. An overview of the basic helix-loop-helix proteins. Genome Biol. 2004;5(6):226.

6. Poulin G, Turgeon B, Drouin J. NeuroD1/beta2 contributes to cell-specific transcription of the proopiomelanocortin gene. Mol Cell Biol. 1997;17(11):6673–6682.

7. Verma-Kurvari S, Savage T, Gowan K, Johnson JE. Lineage-specific regulation of the neural differentiation gene MASH1. Dev Biol. 1996;180(2):605–617.

8. McNay DE, Pelling M, Claxton S, Guillemot F, Ang SL. Mash1 is required for generic and subtype differentiation of hypothalamic neuroendocrine cells. Mol Endocrinol. 2006;20(7):1623–1632.

9. Thaler JP, Lee SK, Jurata LW, Gill GN, Pfaff SL. LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. Cell. 2002;110(2):237–249.

10. Matthews JM, Visvader JE. LIM-domain-binding protein 1: A multifunctional cofactor that interacts with diverse proteins. EMBO Rep. 2003;4(12):1132–1137.

11. Krivega I, Dale RK, Dean A. Role of LDB1 in the transition from chromatin looping to transcription activation.Genes Dev. 2014;28(12):1278–1290.

12. Budry L, Couture C, Balsalobre A, Drouin J. The Ets factor Etv1 interacts with Tpit protein for pituitary pro-opiomelanocortin (POMC) gene transcription. J Biol Chem. 2011;286(28):25387–25396.

13. Maira M, Martens C, Philips A, Drouin J. Heterodimerization between

members of the Nur subfamily of orphan nuclear receptors as a novel mechanism for gene activation. Mol Cell Biol. 1999;19(11):7549–7557.

14. Cau E, Gradwohl G, Casarosa S, Kageyama R, Guillemot F. Hes genes regulate sequential stages of neurogenesis in the olfactory epithelium. Development. 2000;127(11):2323–2332.

15. Kageyama R, Ohtsuka T, Tomita K. The bHLH gene Hes1 regulates differentiation of multiple cell types. Mol Cells. 2000;10(1):1–7.

16. Langlais D, Couture C, Sylvain-Drolet G, Drouin J. A pituitary-specific enhancer of the POMC gene with preferential activity in corticotrope cells. Mol Endocrinol. 2011;25(2):348–359.

17. Morcillo P, Rosen C, Baylies MK, Dorsett D. Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. Genes Dev. 1997;11(20):2729–2740.

18. Morcillo P, Rosen C, Dorsett D. Genes regulating the remote wing margin enhancer in the Drosophila cut locus. Genetics. 1996;144(3):1143–1154.

19. Bach I, Carrière C, Ostendorff HP, Andersen B, Rosenfeld MG. A family of LIM domain-associated cofactors confer transcriptional synergism between LIM and Otx homeodomain proteins. Genes Dev.1997;11(11):1370–1380.

20. Heinlein CA, Chang C. Androgen receptor in prostate cancer. Endocr Rev. 2004;25:276–308.

21. Scher HI, Sawyers CL. Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. J Clin Oncol. 2005;23:8253–8261. doi:23/32/8253 [pii] 10.1200/JCO.2005.03.4777.

22. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011;44:667–678. doi:S1097-2765(11)00680-0 [pii] 10.1016/j.molcel.2011.08.027.

23. Kypta RM, Waxman J. Wnt/beta-catenin signalling in prostate cancer. Nature reviews. Urology. 2012doi:10.1038/nrurol.2012.116.

24. Wang Q, Carroll JS, Brown M. Spatial and Temporal Recruitment of Androgen Receptor and Its Coactivators Involves Chromosomal Looping and Polymerase Tracking. Molecular cell. 2005;19:631–642.

25. Gu B, Watanabe K, Dai X. Pygo2 regulates histone gene expression and H3 K56 acetylation in human mammary epithelial cells. Cell Cycle. 2012;11:79–87.

doi:10.4161/cc.11.1.18402.

26. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25. doi:10.1186/gb-2009-10-3-r25.

27. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics.2010;26:841–842. doi:10.1093/bioinformatics/btq033.

28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–140. doi:10.1093/bioinformatics/btp616.

29. Liao DF, Monia B, Dean N, Berk BC. Protein kinase C-zeta mediates angiotensin II activation of ERK1/2 in vascular smooth muscle cells. J Biol Chem. 1997;272:6146–6150.

30. Newman JJ, Young RA. Connecting transcriptional control to chromosome structure and human disease. Cold Spring Harb Symp Quant Biol. 2010;75:227–35.

31. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell.2011;144(3):327–39.

32. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet. 2011;12(4):283–93.

33. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012;482(7385):339–46.

34. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011;43(6):904–14.

35. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet.2009;10(3):155–9.

36. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465(7295):182–187.

37. Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell. 2011;145(4):622–34.

38. Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, Glass CK, Rosenfeld MG, Fu XD. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature. 2011;474(7351):390–4.

39. Welboren WJ, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN, Stunnenberg HG. ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J.2009;28(10):1418–28.

40. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, Fox EA, Silver PA, Gingeras TR, Liu XS, Brown M. Genome-wide analysis of estrogen receptor binding sites. Nat Genet. 2006;38(11):1289–97.

41. Kwon YS, Garcia-Bassets I, Hutt KR, Cheng CS, Jin M, Liu D, Benner C, Wang D, Ye Z, Bibikova M, Fan JB, Duan L, Glass CK, Rosenfeld MG, Fu XD. Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. Proc Natl Acad Sci U S A. 2007;104(12):4852–7.

42. Heintzman ND, Ren B. Finding distal regulatory elements in the human genome. Curr Opin Genet Dev.2009;19(6):541–9.

43. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009;459(7243):108–12.

44. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107(50):21931–6.

45. Ahlenstiel CL, Lim HG, Cooper DA, Ishida T, Kelleher AD, Suzuki K. Direct evidence of nuclear Argonaute distribution during transcriptional silencing links the actin cytoskeleton to nuclear RNAi machinery in human cells.Nucleic Acids Res. 2012;40(4):1579–95.

46. Mayer C, Schmitz KM, Li J, Grummt I, Santoro R. Intergenic Transcripts Regulate the Epigenetic State of rRNA Genes. Mol Cell. 2006;5(3):351–61. 22.

47. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY. A long noncoding RNA maintains active chromatin to

coordinate homeotic gene expression. Nature. 2011;472:120–4.

48. Melo CA,. eRNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. Mol Cell. 2013;49(3):524–35. 7.

49. Lai F, Orom UA, Matteo Cesaroni O, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature. 2013;494(7438):497–501.28.

50. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu X-D, Topol EJ, Rosenfeld MG, Frazer KA. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature. 2011;470(7333):264–8.

51. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature.2012;489:109–13.

52. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289–93.

53. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic Maps of LongNoncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Mol Cell. 2011;44:667–78.

54. Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature. 2009;460(7253):410–3.

55. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010;467(7314):430–5.

56. Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res. 2010;20(5):578–88.

57. Cai S, Kohwi-Shigematsu T. Intranuclear relocalization of matrix binding sites during T cell activation detected by amplified fluorescence in situ hybridization. Methods. 1999;19(3):394–402.

58. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science.

2008;322(5909):1845–8.

59. Heinlein CA, Chang C. Androgen receptor in prostate cancer. Endocr Rev. 2004;25:276–308.

60. Scher HI, Sawyers CL. Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. J Clin Oncol. 2005;23:8253–8261. doi:23/32/8253 [pii] 10.1200/JCO.2005.03.4777.

61. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011;44:667–678. doi:S1097-2765(11)00680-0 [pii] 10.1016/j.molcel.2011.08.027.

62. Kypta RM, Waxman J. Wnt/beta-catenin signalling in prostate cancer. Nature reviews. Urology. 2012doi:10.1038/nrurol.2012.116.

63. Wang Q, Carroll JS, Brown M. Spatial and Temporal Recruitment of Androgen Receptor and Its Coactivators Involves Chromosomal Looping and Polymerase Tracking. Molecular cell. 2005;19:631–642.

64. Gu B, Watanabe K, Dai X. Pygo2 regulates histone gene expression and H3 K56 acetylation in human mammary epithelial cells. Cell Cycle. 2012;11:79–87. doi:10.4161/cc.11.1.18402.

65. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25. doi:10.1186/gb-2009-10-3-r25.

66. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics.2010;26:841–842. doi:10.1093/bioinformatics/btq033.

67. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–140. doi:10.1093/bioinformatics/btp616.

68. Liao DF, Monia B, Dean N, Berk BC. Protein kinase C-zeta mediates angiotensin II activation of ERK1/2 in vascular smooth muscle cells. J Biol Chem. 1997;272:6146–6150.

69. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nature genetics 23, 185-188.

70. Baker, S.A., Chen, L., Wilkins, A.D., Yu, P., Lichtarge, O., and Zoghbi, H.Y.

(2013). An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. Cell 152, 984-996.

71. Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482-485.

72. Bell, A.C., West, A.G., and Felsenfeld, G. (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. Science (New York, NY 291, 447-450.

73. Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C.A., and Zoghbi, H.Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. Human molecular genetics 18, 2431-2442.

74. Berube, N.G., Smeenk, C.A., and Picketts, D.J. (2000). Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. Human molecular genetics 9, 539-547.

75. Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science (New York, NY 320, 1224-1229.

76. Chen, R.Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nature genetics 27, 327-331.

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

78. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-380.

79. Dunn, K.L., Zhao, H., and Davie, J.R. (2003). The insulator binding protein CTCF associates with the nuclear matrix. Experimental cell research 288, 218-223.

80. Garcia-Bassets, I., Kwon, Y.S., Telese, F., Prefontaine, G.G., Hutt, K.R., Cheng, C.S., Ju, B.G., Ohgi, K.A., Wang, J., Escoubet-Lozach, L.,. (2007). Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. Cell 128, 505-518.

81. Giacometti, E., Luikenhuis, S., Beard, C., and Jaenisch, R. (2007). Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. Proceedings of the National Academy of Sciences of the United States of America 104, 1931-1936.

82. Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell stem cell 14, 762-775.

83. Guy, J., Cheval, H., Selfridge, J., and Bird, A. (2011). The role of MeCP2 in the brain. Annual review of cell and developmental biology 27, 631-652.

84. Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science (New York, NY 315, 1143-1147.

85. Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nature genetics 27, 322-326.

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

87. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell reports 2, 345-357.

88. Horike, S., Cai, S., Miyano, M., Cheng, J.F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nature genetics 37, 31-40.

89. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nature biotechnology 30, 90-98.

90. Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R., and Berube, N.G. (2010). ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. Developmental cell 18, 191-202.

91. Kernohan, K.D., Vernimmen, D., Gloor, G.B., and Berube, N.G. (2014). Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. Nucleic acids research 42, 8356-8368.

92. Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I., and Bird, A.P. (2005). DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Molecular cell 19, 667-678.

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

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

95. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.,. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (New York, NY 326, 289-293.

96. Lioy, D.T., Garg, S.K., Monaghan, C.E., Raber, J., Foust, K.D., Kaspar, B.K., Hirrlinger, P.G., Kirchhoff, F., Bissonnette, J.M., Ballas, N.,. (2011). A role for glia in the progression of Rett's syndrome. Nature 475, 497-500.

97. Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A.K., and Rosenfeld, M.G. (2013). Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell 155, 1581-1595.

98. Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 151, 1417-1430.

99. Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A.,. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell reports 3, 1678-1689.

100. Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proceedings of the National Academy of Sciences of the United States of America 104, 2709-2714.

101. Nickerson, J. (2001). Experimental observations of a nuclear matrix. Journal of cell science 114, 463-474.

102. Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007). Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. Molecular and cellular biology 27, 864-877.

103. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.,. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381-385.

104. Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nature reviews Genetics 15, 234-246.

105. Pederson, T. (1998). Thinking about a nuclear matrix. Journal of molecular biology 277, 147-159.

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

107. Razin, S.V., Borunova, V.V., Iarovaia, O.V., and Vassetzky, Y.S. (2014). Nuclear matrix and structural and functional compartmentalization of the eucaryotic cell nucleus. Biochemistry Biokhimiia 79, 608-618.

108. Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., Sadreyev, R., and Lee, J.T. (2014). ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets. Cell 159, 869-883.

109. Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R., and Zoghbi, H. (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. Neuron 35, 243-254.

110. Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Molecular cell 37, 457-468.

111. Skowronska-Krawczyk, D., Ma, Q., Schwartz, M., Scully, K., Li, W., Liu, Z., Taylor, H., Tollkuhn, J., Ohgi, K.A., Notani, D.,. (2014). Required enhancer-matrin-3 network interactions for a homeodomain transcription program. Nature.

112. Smallwood, A., and Ren, B. (2013). Genome organization and long-range regulation of gene expression by enhancers. Current opinion in cell biology 25, 387-394.

113. Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proceedings of the National Academy of Sciences of the United States of America 99, 15536-15541.

114. Van Bortle, K., Nichols, M.H., Li, L., Ong, C.T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome biology 15, R82.

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

116. Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R.,. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome research 22, 1680-1688.

117. Weitzel, J.M., Buhrmester, H., and Stratling, W.H. (1997). Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. Molecular and cellular biology 17, 5656-5666.

118. Zhao, Y.T., Goffin, D., Johnson, B.S., and Zhou, Z. (2013). Loss of MeCP2 function is associated with distinct gene expression changes in the striatum. Neurobiology of disease 59, 257-266.

119. Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F.,. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proceedings of the National Academy of Sciences of the United States of America 111, 996-1001.

120. Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. (2004). Structure and evolution of transcriptional regulatory networks. Current Opinion in Structural Biology 14, 283-291.

121. Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

122. Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74.

123. Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. Trends Genet 28, 276-284.

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

125. Chepelev, I., Wei, G., Wangsa, D., Tang, Q., and Zhao, K. (2012). Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. Cell Res. 22, 490-503.

126. Dupe, V., Davenne, M., Brocard, J., Dolle, P., Mark, M., Dierich, A.,

Chambon, P., and Rijli, F.M. (1997). In vivo functional analysis of the Hoxa-1 3' retinoic acid response element (3'RARE). Development 124, 399-410.

127. Foulds, C.E., Feng, Q., Ding, C., Bailey, S., Hunsaker, T.L., Malovannaya, A., Hamilton, R.A., Gates, L.A., Zhang, Z., Li, C.,. (2013). Proteomic analysis of coregulators bound to ERalpha on DNA and nucleosomes reveals coregulator dynamics. Molecular Cell 51, 185-199.

128. Furlan-Magaril, M., Rincon-Arano, H., and Recillas-Targa, F. (2009). Sequential chromatin immunoprecipitation protocol: ChIP-reChIP. Methods in Molecular Biology 543, 253-266.

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

130. Harismendy, O., Notani, D., Song, X., Rahim, N.G., Tanasa, B., Heintzman, N., Ren, B., Fu, X.D., Topol, E.J., Rosenfeld, M.G.,. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature 470, 264-268.

131. Hartley, K.O., Gell, D., Smith, G.C., Zhang, H., Divecha, N., Connelly, M.A., Admon, A., Lees-Miller, S.P., Anderson, C.W., and Jackson, S.P. (1995). DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. Cell 82, 849-856.

132. 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.,. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459, 108-112.

133. 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.,. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics 39, 311-318.

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

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

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

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

138. Jin, S., and Weaver, D.T. (1997). Double-strand break repair by Ku70 requires heterodimerization with Ku80 and DNA binding functions. The EMBO Journal 16, 6874-6885.

139. Joseph, R., Orlov, Y.L., Huss, M., Sun, W., Kong, S.L., Ukil, L., Pan, Y.F., Li, G., Lim, M., Thomsen, J.S.,. (2010). Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. Molecular Systems Biology 6, 456.

140. Jozwik, K.M., and Carroll, J.S. (2012). Pioneer factors in hormone-dependent cancers. Nature Reviews Cancer 12, 381-385.

141. Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell 148, 473-486.

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

143. 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.,. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature 465, 182-187.

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

145. Langlais, D., Couture, C., Balsalobre, A., and Drouin, J. (2012). The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. Molecular Cell 47, 38-49.

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

147. Lonard, D.M., and O'Malley, B.W. (2012). Nuclear receptor coregulators:

modulators of pathology and therapeutic targets. Nature Reviews Endocrinology 8, 598-604.

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

149. Mohammed, H., D'Santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L.,. (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Reports 3, 342-349.

150. Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M.,. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature 466, 714-719.

151. Nesbit, M.A., Bowl, M.R., Harding, B., Ali, A., Ayala, A., Crowe, C., Dobbie, A., Hampson, G., Holdaway, I., Levine, M.A.,. (2004). Characterization of GATA3 mutations in the hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome. J. Biol. Chem. 279, 22624-22634.

152. Pascual, G., Fong, A.L., Ogawa, S., Gamliel, A., Li, A.C., Perissi, V., Rose, D.W., Willson, T.M., Rosenfeld, M.G., and Glass, C.K. (2005). A SUMOylation-dependent pathway mediates transrepression of inflammatory response genes by PPAR-gamma. Nature 437, 759-763.

153. Perissi, V., Scafoglio, C., Zhang, J., Ohgi, K.A., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2008). TBL1 and TBLR1 phosphorylation on regulated gene promoters overcomes dual CtBP and NCoR/SMRT transcriptional repression checkpoints. Molecular Cell 29, 755-766.

154. Reichardt, H.M., Kaestner, K.H., Tuckermann, J., Kretz, O., Wessely, O., Bock, R., Gass, P., Schmid, W., Herrlich, P., Angel, P.,. (1998). DNA binding of the glucocorticoid receptor is not essential for survival. Cell 93, 531-541.

155. Rosenfeld, M.G., Lunyak, V.V., and Glass, C.K. (2006). Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. Genes & Development 20, 1405-1428.

156. Ross-Innes, C.S., Stark, R., Holmes, K.A., Schmidt, D., Spyrou, C., Russell, R., Massie, C.E., Vowler, S.L., Eldridge, M., and Carroll, J.S. (2010). Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. Genes & Development 24, 171-182.

157. Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers:

from properties to genome-wide predictions. Nature Reviews Genetics 15, 272-286.

158. Siersbaek, R., Baek, S., Rabiee, A., Nielsen, R., Traynor, S., Clark, N., Sandelin, A., Jensen, O.N., Sung, M.H., Hager, G.L.,. (2014a). Molecular Architecture of Transcription Factor Hotspots in Early Adipogenesis. Cell Reports.

159. Siersbaek, R., Rabiee, A., Nielsen, R., Sidoli, S., Traynor, S., Loft, A., Poulsen, L.L., Rogowska-Wrzesinska, A., Jensen, O.N., and Mandrup, S. (2014b). Transcription Factor Cooperativity in Early Adipogenic Hotspots and Super-Enhancers. Cell Reports.

160. 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. Molecular and Cellular Biology 30, 3943-3955.

161. Taatjes, D.J., Marr, M.T., and Tjian, R. (2004). Regulatory diversity among metazoan co-activator complexes. Nature Reviews Molecular Cell Biology 5, 403-410.

162. Tee, W.W., Shen, S.S., Oksuz, O., Narendra, V., and Reinberg, D. (2014). Erk1/2 Activity Promotes Chromatin Features and RNAPII Phosphorylation at Developmental Promoters in Mouse ESCs. Cell 156, 678-690.

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

164. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B.,. (2012). The accessible chromatin landscape of the human genome. Nature 489, 75-82.

165. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nature Reviews Genetics 10, 252-263.

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

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

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

169. Xie, W., and Ren, B. (2013). Developmental biology. Enhancing pluripotency and lineage specification. Science 341, 245-247.

170. Yan, J., Enge, M., Whitington, T., Dave, K., Liu, J., Sur, I., Schmierer, B., Jolma, A., Kivioja, T., Taipale, M.,. (2013). Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. Cell 154, 801-813.

171. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., Chen, L.,. (2014). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat. Biotechnol. 2012;30:90-98.

172. Peng, C., Fu, L., Dong, P., Dong, Z., Li, J., Wang, X., Zhang, H. (2013). The sequencing bias relaxed characteristics of Hi-C derived data and implications for chromatin 3D modeling. Nucleic Acids Res. 41(19):e183.

173. Zhu, J., Adli, M., Zou JY., Verstappen G., Coyne, M., Zhang X., Durham T., Miri M., Deshpande V., De Jager PL.,. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell. 152(3):642-54.

174. Benjamini, Y., and Speed, T. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucl. Acids Res. doi: 10.1093/nar/gsk001.

175. Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nature genetics 23, 185-188.

176. Baker, S.A., Chen, L., Wilkins, A.D., Yu, P., Lichtarge, O., and Zoghbi, H.Y. (2013). An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. Cell 152, 984-996.

177. Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482-485.

178. Bell, A.C., West, A.G., and Felsenfeld, G. (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. Science (New York, NY 291, 447-450.

179. Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C.A., and Zoghbi, H.Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. Human molecular genetics 18, 2431-2442.

180. Berube, N.G., Smeenk, C.A., and Picketts, D.J. (2000). Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. Human molecular genetics 9, 539-547.

181. Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science (New York, NY 320, 1224-1229.

182 .Chen, R.Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nature genetics 27, 327-331.

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

184. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-380.

185. Dunn, K.L., Zhao, H., and Davie, J.R. (2003). The insulator binding protein CTCF associates with the nuclear matrix. Experimental cell research 288, 218-223.

186. Garcia-Bassets, I., Kwon, Y.S., Telese, F., Prefontaine, G.G., Hutt, K.R., Cheng, C.S., Ju, B.G., Ohgi, K.A., Wang, J., Escoubet-Lozach, L.,. (2007). Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. Cell 128, 505-518.

187. Giacometti, E., Luikenhuis, S., Beard, C., and Jaenisch, R. (2007). Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. Proceedings of the National Academy of Sciences of the United States of America 104, 1931-1936.

188. Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell stem cell 14, 762-775.

189. Guy, J., Cheval, H., Selfridge, J., and Bird, A. (2011). The role of MeCP2 in the brain. Annual review of cell and developmental biology 27, 631-652.

190. Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science (New York, NY 315, 1143-1147.

191. Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nature genetics 27, 322-326.

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

193. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell reports 2, 345-357.

194. Horike, S., Cai, S., Miyano, M., Cheng, J.F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nature genetics 37, 31-40.

195. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nature biotechnology 30, 90-98.

196. Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R., and Berube, N.G. (2010). ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. Developmental cell 18, 191-202.

197. Kernohan, K.D., Vernimmen, D., Gloor, G.B., and Berube, N.G. (2014). Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. Nucleic acids research 42, 8356-8368.

198. Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I., and Bird, A.P. (2005). DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Molecular cell 19, 667-678.

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

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

201. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.,. (2009). Comprehensive mapping of long-range interactions reveals folding

principles of the human genome. Science (New York, NY 326, 289-293.

202. Lioy, D.T., Garg, S.K., Monaghan, C.E., Raber, J., Foust, K.D., Kaspar, B.K., Hirrlinger, P.G., Kirchhoff, F., Bissonnette, J.M., Ballas, N.,. (2011). A role for glia in the progression of Rett's syndrome. Nature 475, 497-500.

203. Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A.K., and Rosenfeld, M.G. (2013). Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell 155, 1581-1595.

204. Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 151, 1417-1430.

205. Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A.,. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell reports 3, 1678-1689.

206. Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proceedings of the National Academy of Sciences of the United States of America 104, 2709-2714.

207. Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007). Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. Molecular and cellular biology 27, 864-877.

208. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.,. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381-385.

209. Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nature reviews Genetics 15, 234-246.

210. Razin, S.V., Borunova, V.V., Iarovaia, O.V., and Vassetzky, Y.S. (2014). Nuclear matrix and structural and functional compartmentalization of the eucaryotic cell nucleus. Biochemistry Biokhimiia 79, 608-618.

211. Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., Sadreyev, R., and Lee, J.T. (2014). ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets. Cell 159, 869-883.

212. Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R., and Zoghbi, H. (2002). Mice with

truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. Neuron 35, 243-254.

213. Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Molecular cell 37, 457-468.

214. Skowronska-Krawczyk, D., Ma, Q., Schwartz, M., Scully, K., Li, W., Liu, Z., Taylor, H., Tollkuhn, J., Ohgi, K.A., Notani, D.,. (2014). Required enhancer-matrin-3 network interactions for a homeodomain transcription program. Nature.

215. Smallwood, A., and Ren, B. (2013). Genome organization and long-range regulation of gene expression by enhancers. Current opinion in cell biology 25, 387-394.

216. Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proceedings of the National Academy of Sciences of the United States of America 99, 15536-15541.

217. Van Bortle, K., Nichols, M.H., Li, L., Ong, C.T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome biology 15, R82.

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

219. Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R.,. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome research 22, 1680-1688.

220. Weitzel, J.M., Buhrmester, H., and Stratling, W.H. (1997). Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. Molecular and cellular biology 17, 5656-5666.

221. Zhao, Y.T., Goffin, D., Johnson, B.S., and Zhou, Z. (2013). Loss of MeCP2 function is associated with distinct gene expression changes in the striatum. Neurobiology of disease 59, 257-266.

222. Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F.,. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proceedings of the National Academy of Sciences of the United States of America 111, 996-1001.

223. A.J., W., A.F., S., and B.J., M. (2010). Condensin and cohesin complexity: the expanding repertoire of functions. Nature reviews. Genetics 11, 391-404.

224. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.,. (2014). An atlas of active enhancers across human cell types and tissues. Nature 507, 455-461.

225. Catic, A., Suh, C.Y., Hill, C.T., Daheron, L., Henkel, T., Orford, K.W., Dombkowski, D.M., Liu, T., Liu, X.S., and Scadden, D.T. (2013). Genome-wide map of nuclear protein degradation shows NCoR1 turnover as a key to mitochondrial gene regulation. Cell 155, 1380-1395.

226. Cuylen, S., Metz, J., Hruby, A., and Haering, C.H. (2013). Entrapment of chromosomes by condensin rings prevents their breakage during cytokinesis. Dev Cell 27, 469-478.

227. D'Ambrosio, C., Schmidt, C.K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K., and Uhlmann, F. (2008). Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. Genes Dev 22, 2215-2227.

228. Dowen, J.M., Bilodeau, S., Orlando, D.A., Hubner, M.R., Abraham, B.J., Spector, D.L., and Young, R.A. (2013). Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements. Stem cell reports 1, 371-378.

229. Emmanuel, C., Gava, N., Kennedy, C., Balleine, R., Sharma, R., Wain, G., Brand, A., Hogg, R., Etemadmoghadam, D., George, J.,. (2011). Comparison of expression profiles in ovarian epithelium in vivo and ovarian cancer identifies novel candidate genes involved in disease pathogenesis. PloS one 6.

230. Foulds, C.E., Feng, Q., Ding, C., Bailey, S., Hunsaker, T.L., Malovannaya, A., Hamilton, R.A., Gates, L.A., Zhang, Z., Li, C.,. (2013). Proteomic analysis of coregulators bound to ERalpha on DNA and nucleosomes reveals coregulator dynamics. Mol Cell 51, 185-199.

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

232. Green, L., Kalitsis, P., Chang, T., Cipetic, M., Kim, J., Marshall, O., Turnbull, L., Whitchurch, C., Vagnarelli, P., Samejima, K.,. (2012). Contrasting roles of condensin I and condensin II in mitotic chromosome formation. Journal of cell science 125, 1591-1604.

233. Hah, N., Murakami, S., Nagari, A., Danko, C., and Kraus, W. (2013).

Enhancer transcripts mark active estrogen receptor binding sites. Genome research 23, 1210-1223.

234. Heale, J.T., Ball, A.R., Jr., Schmiesing, J.A., Kim, J.S., Kong, X., Zhou, S., Hudson, D.F., Earnshaw, W.C., and Yokomori, K. (2006). Condensin I interacts with the PARP-1-XRCC1 complex and functions in DNA single-strand break repair. Mol Cell 21, 837-848.

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

236. Hirano, T. (2012). Condensins: universal organizers of chromosomes with diverse functions. Genes Dev 26, 1659-1678.

237. Ho, P.C., Tsui, Y.C., Feng, X., Greaves, D.R., and Wei, L.N. (2012). NF-kappaB-mediated degradation of the coactivator RIP140 regulates inflammatory responses and contributes to endotoxin tolerance. Nat Immunol 13, 379-386.

238. Hsieh, C.-L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C., Lee, G.-S.M., Chen, S.,. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proceedings of the National Academy of Sciences of the United States of America.

239. Hudson, D., Ohta, S., Freisinger, T., Macisaac, F., Sennels, L., Alves, F., Lai, F., Kerr, A., Rappsilber, J., and Earnshaw, W. (2008). Molecular and genetic analysis of condensin function in vertebrate cells. Molecular biology of the cell 19, 3070-3079.

240. Jans, J., Gladden, J., Ralston, E., Pickle, C., Michel, A., Pferdehirt, R., Eisen, M., and Meyer, B. (2009). A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. Genes & development 23, 602-618.

241. Jeppsson, K., Kanno, T., Shirahige, K., and Sjogren, C. (2014). The maintenance of chromosome structure: positioning and functioning of SMC complexes. Nature reviews. Molecular cell biology 15, 601-614.

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

243. Kaikkonen, M., Spann, N., Heinz, S., Romanoski, C., Allison, K., Stender, J., Chun, H., Tough, D., Prinjha, R., Benner, C.,. (2013). Remodeling of the

enhancer landscape during macrophage activation is coupled to enhancer transcription. Molecular cell 51, 310-325.

244. Kim, J.H., Zhang, T., Wong, N.C., Davidson, N., Maksimovic, J., Oshlack, A., Earnshaw, W.C., Kalitsis, P., and Hudson, D.F. (2013). Condensin I associates with structural and gene regulatory regions in vertebrate chromosomes. Nat Commun 4, 2537.

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

246. Kranz, A.L., Jiao, C.Y., Winterkorn, L.H., Albritton, S.E., Kramer, M., and Ercan, S. (2013). Genome-wide analysis of condensin binding in Caenorhabditis elegans. Genome Biol 14, R112.

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

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

249. Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends in biochemical sciences 39, 170-182.

250. Leiserson, M.D., Vandin, F., Wu, H., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M.,. (2014). Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet.

251. Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature 430, 573-578.

252 .Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H.Q., Ong, C.T., Cubenas-Potts, C., Hu, M., Lei, E.P., Bosco, G.,. (2015). Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing. Mol Cell 58, 216-231.

253. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A., Merkurjev, D., Zhang, J., Ohgi, K., Song, X.,. (2013a). Functional roles of enhancer RNAs for

oestrogen-dependent transcriptional activation. Nature 498, 516-520.

254. Li, X., Zhou, Q., Sunkara, M., Kutys, M.L., Wu, Z., Rychahou, P., Morris, A.J., Zhu, H., Evers, B.M., and Huang, C. (2013b). Ubiquitylation of phosphatidylinositol 4-phosphate 5-kinase type I gamma by HECTD1 regulates focal adhesion dynamics and cell migration. J Cell Sci 126, 2617-2628.

255. Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M.J., Song, X., Zhang, F., Ma, Q., Ohgi, K.A.,. (2014). Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex. Cell 159, 358-373.

256. Melgar, M.F., Collins, F.S., and Sethupathy, P. (2010). Discovery of active enhancers through bidirectional expression of short transcripts. Genome biology 12.

257. Melo, C., Drost, J., Wijchers, P., van de Werken, H., de Wit, E., Oude Vrielink, J., Elkon, R., Melo, S., Léveillé, N., Kalluri, R.,. (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Molecular cell 49, 524-535.

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

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

260. Murakami-Tonami, Y., Kishida, S., Takeuchi, I., Katou, Y., Maris, J., Ichikawa, H., Kondo, Y., Sekido, Y., Shirahige, K., Murakami, H.,. (2014). Inactivation of SMC2 shows a synergistic lethal response in MYCN-amplified neuroblastoma cells. Cell cycle (Georgetown, Tex.) 13.

261. Nawaz, Z., Lonard, D., Smith, C., Lev-Lehman, E., Tsai, S., Tsai, M., and O'Malley, B. (1999). The Angelman syndrome-associated protein, E6-AP, is a coactivator for the nuclear hormone receptor superfamily. Molecular and cellular biology 19, 1182-1189.

262. Ono, T., Losada, A., Hirano, M., Myers, M., Neuwald, A., and Hirano, T. (2003). Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. Cell 115, 109-121.

263. Piazza, I., Rutkowska, A., Ori, A., Walczak, M., Metz, J., Pelechano, V., Beck, M., and Haering, C.H. (2014). Association of condensin with chromosomes

depends on DNA binding by its HEAT-repeat subunits. Nat Struct Mol Biol 21, 560-568.

264. Plank, J.L., and Dean, A. (2014). Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. Molecular cell 55, 5-14.

265. Pnueli, L., Rudnizky, S., Yosefzon, Y., and Melamed, P. (2015). RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin alpha-subunit gene. Proc Natl Acad Sci U S A 112, 4369-4374.

266. Ross-Innes, C., Stark, R., Holmes, K., Schmidt, D., Spyrou, C., Russell, R., Massie, C., Vowler, S., Eldridge, M., and Carroll, J. (2010). Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. Genes & development 24, 171-182.

267. Ruff, M., Gangloff, M., Wurtz, J.M., and Moras, D. (2000). Estrogen receptor transcription and transactivation: Structure-function relationship in DNA- and ligand-binding domains of estrogen receptors. Breast cancer research : BCR 2, 353-359.

268. Ryu, B., Kim, D., Deluca, A., and Alani, R. (2007). Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression. PloS one 2.

269. Sarkar, A., and Zohn, I. (2012). Hectd1 regulates intracellular localization and secretion of Hsp90 to control cellular behavior of the cranial mesenchyme. The Journal of cell biology 196, 789-800.

270. Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA Facilitates NELF Release from Immediate Early Genes. Mol Cell 56, 29-42.

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

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

273. St-Pierre, J., Douziech, M., Bazile, F., Pascariu, M., Bonneil, E., Sauvé, V., Ratsima, H., and D'Amours, D. (2009). Polo kinase regulates mitotic chromosome condensation by hyperactivation of condensin DNA supercoiling activity. Molecular cell 34, 416-426.

274. Sun, T., Wang, X., He, H.H., Sweeney, C.J., Liu, S.X., Brown, M., Balk, S., Lee, G.S., and Kantoff, P.W. (2014). MiR-221 promotes the development of androgen independence in prostate cancer cells via downregulation of HECTD2 and RAB1A. Oncogene 33, 2790-2800.

275. Van Bortle, K., Nichols, M.H., Li, L., Ong, C.-T.T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome biology 15.

276. Villalobos, M., Olea, N., Brotons, J.A., Olea-Serrano, M.F., Almodovar, J.M.R.d., and Pedraza, V. (1995). The E-screen assay: a comparison of different MCF7 cell stocks. Environmental Health Perspectives 103.

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

278. Watson, P.J., Fairall, L., and Schwabe, J.W. (2012). Nuclear hormone receptor co-repressors: structure and function. Molecular and cellular endocrinology 348, 440-449.

279. White, K.A., Yore, M.M., Deng, D., and Spinella, M.J. (2005). Limiting effects of RIP140 in estrogen signaling: potential mediation of anti-estrogenic effects of retinoic acid. The Journal of biological chemistry 280, 7829-7835.

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

281. Zhou, J., Bi, D., Lin, Y., Chen, P., Wang, X., and Liang, S. (2012). Shotgun proteomics and network analysis of ubiquitin-related proteins from human breast carcinoma epithelial cells. Molecular and cellular biochemistry 359, 375-384.